



Cairo University

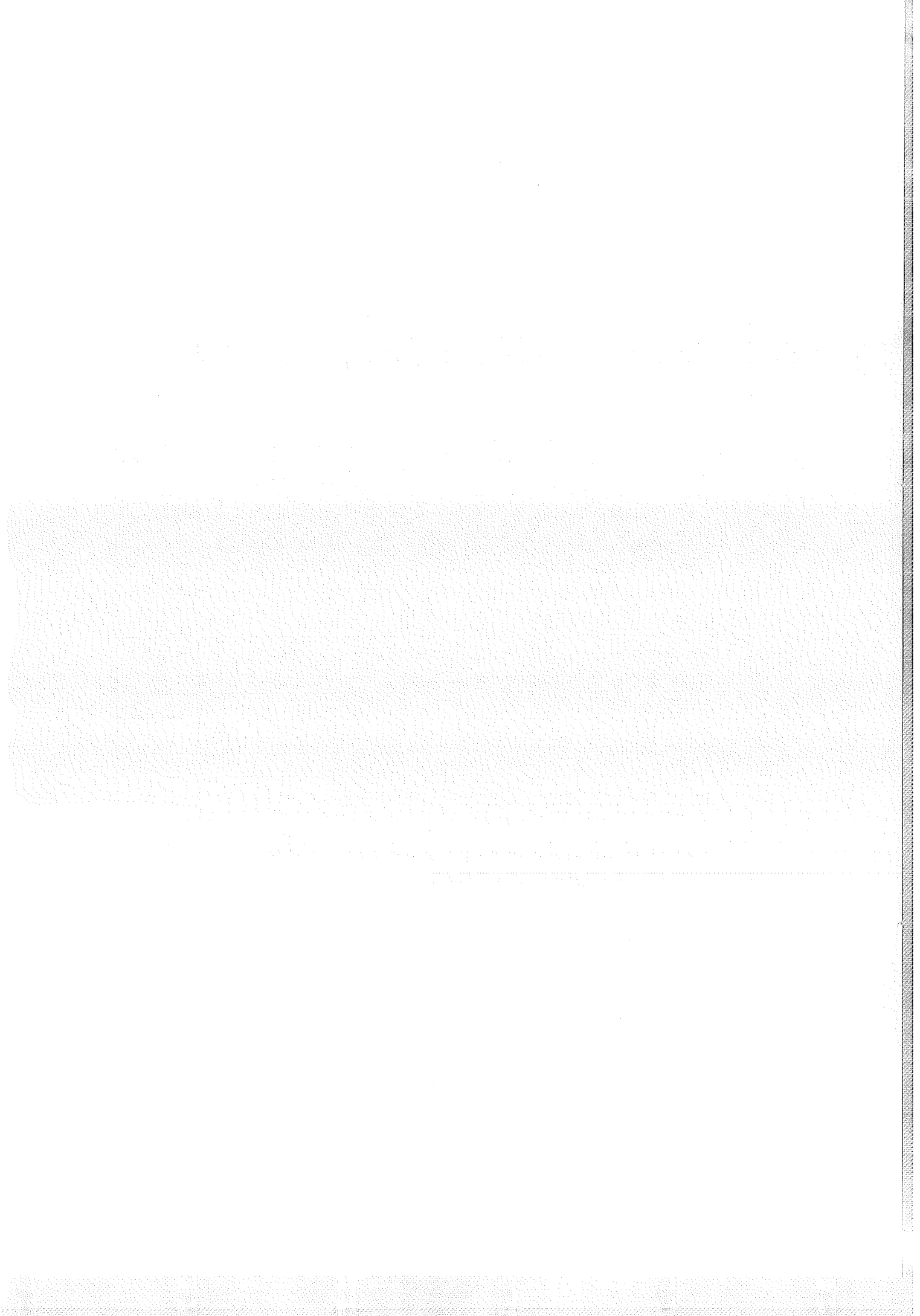
Institute of Statistical Studies and Research



**The 48th Annual Conference on Statistics , Computer
Sciences and Operation Research**

APPLIED STATISTICS

23-26 Dec ,2013



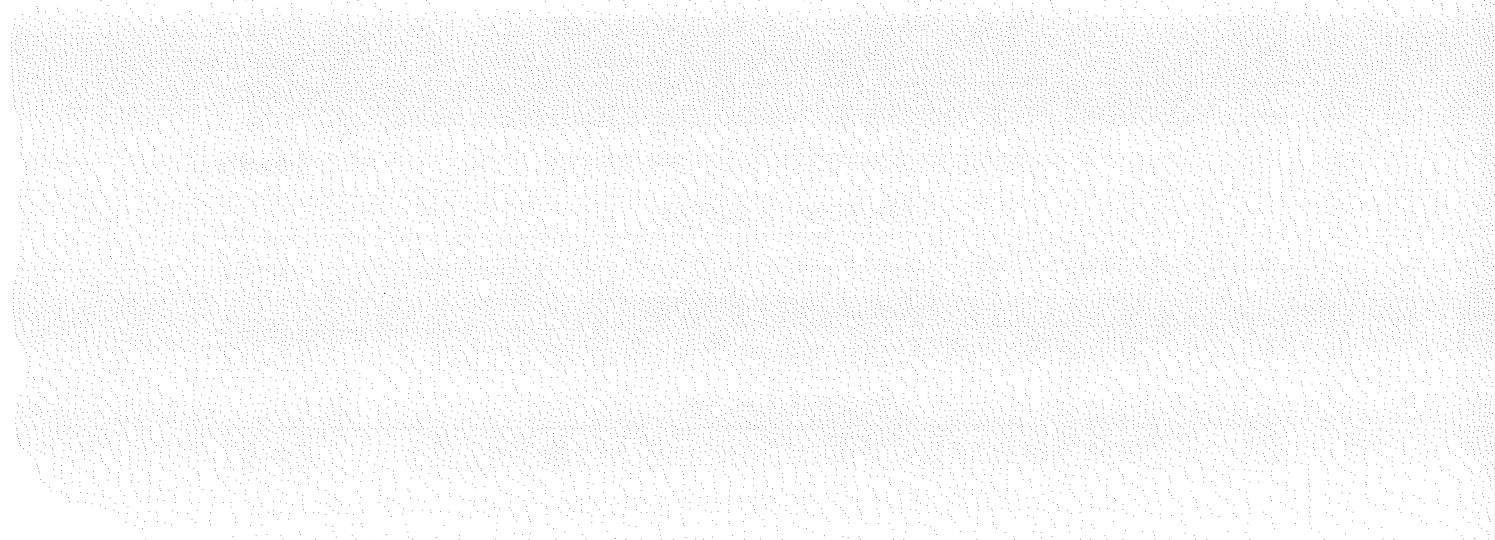
Index

APPLIED STATISTICS

1	Competing Risks Survival Analysis Applied to Oro-Dental Cancer Patients Abd El Massieh P.M, Aboul Ezz E.H.A, Mousa A.M	1-16
2	Using Rough Set Approach for Discovering Systemic Lupus Erythematosus in Pregnant Females and its Relation to Correlation Analysis N.R.Hassan , E.A.Hussein, Y.M.Ibrahim	17-25
3	Robust Regression for handling Outliers Using Mathematical Programming Hazem Refaat	26-45
4	Proposed SUR-ridge Estimators With Application M.A.Abdel-Fattah	46-61
5	Time Series Analysis of Infation in Egypt Kandil,S.M	62-82
6	Time Series: Identification and Modeling for Discrete Dynamic System Zidan,A.I	83-97
7	Calibration Estimation in Stratified Sampling El-Sheikh A.A, Abdel-Rahman H. Mohamed	98-111

THE UNIVERSITY OF CHICAGO
LIBRARY

1000 S. MICHIGAN AVE.
CHICAGO, ILL. 60607



1000 S. MICHIGAN AVE.
CHICAGO, ILL. 60607

1000 S. MICHIGAN AVE.
CHICAGO, ILL. 60607

1000 S. MICHIGAN AVE.
CHICAGO, ILL. 60607

1000 S. MICHIGAN AVE.
CHICAGO, ILL. 60607

Competing Risks Survival Analysis Applied to Oro-Dental Cancer Patients

Abd El Massieh P.M.^a Aboul Ezz E.H.A.^b Mousa A.M.^c

Abstract

Background: A topic that has received attention in both the statistical and medical literature is the estimation of the probability of failure for endpoints that are subject to competing risks. Methods of standard survival analysis such as the complement of Kaplan-Meier (KM) for estimation of cumulative incidence curves (1-KM), the log-rank test for comparison of these curves, and the standard Cox model for the assessment of covariates, are still used by medical researchers, however they lead to biased and uninterpretable results when an analysis focuses on time-to-first event in scenarios where there are competing events. **Methods:** in this article we compared these standard survival methods with other alternative methods as the cumulative incidence competing risks (CICR) curves, the Gray's test for comparing these curves, and the competing risks regression model (CRR). The data used included 348 patients with primary cancer of the oral cavity diagnosed at Tanta cancer center (TCC) during the period 2003-2007, data were analyzed using R program. **Objective:** our main aim is to attract the attention of medical researchers to use the appropriate statistical technique according to the nature of the available data taking into consideration the presence of the competing risks. **Results:** The incidence of each main outcome was overestimated by the standard methods. The magnitude of overestimation varied with baseline patient characteristics and follow-up duration, with most of this variation being related to the rate of the competing event. **Conclusion:** Using specific techniques for the analysis of competing risks will ensure that the results are unbiased and can be correctly interpreted.

Keywords: Competing risk; Kaplan-Meier curve; Cumulative incidence; Survival data.

1- Introduction

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is **time** until an **event** occurs, in other words it is the analysis of data measured from a specific time of origin until an event of interest or a specific endpoint [11].

The analysis of many time-to-event endpoints in various scientific disciplines, particularly in the medical sciences, is complicated by the presence of competing risks. The competing risk is defined as an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event [6].

Most survival analysis must consider a key analytical problem called censoring, censoring occurs when some information about individual's survival time is available, but the survival time exactly is not known. There are generally three reasons why censoring may occur, a person doesn't experience the event before the study ends, a

^a Research assistant, Oro-Dental Genetics Department., National Research Center, Cairo, Egypt

^b Professor of Craniofacial and Oral Genetics, Basic Dental Science Department., National Research Centre, Cairo, Egypt

^c Professor of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo, Egypt

person is lost to follow-up during the study period, or a person withdraws from the study because of death (if death is not the event of interest) or some other reason which might be a competing risk [10].

When the data consist of patients who experience an event and censored individuals, a nonparametric estimate of the cumulative incidence can be obtained using the Kaplan–Meier method. Under this approach, the censoring mechanism is assumed to be noninformative. In other words, the survival time of an individual (or the time at which a subject experiences an event) is assumed to be independent of a mechanism that would cause the patient to be censored. Often times, a patient may experience an event (competing risk event) other than the one of interest which alters the probability of experiencing the event of interest. In this setting, it would often be of interest to calculate the cumulative incidence of a specific event of interest. Any subject who does not experience the event of interest can be treated as censored. However, a patient experiencing a competing risk event is censored in an informative manner. Hence, the Kaplan–Meier estimation procedure may not be directly applicable. The cumulative incidence function for an event of interest must be calculated by appropriately accounting for the presence of competing risk events [16].

Methods for estimating the probability of failure for events that are subject to competing risks are not new, and some work has been done in this area. Gooley T.A. et al. (1999) estimated the probability of squamous cell carcinoma progression, in the head and neck, by calculating 1-KM and CIC. The hazard of failure estimated from 1-KM at that time was 0.58 while the hazard of failure from CIC was 0.46. The difference being due to the patients who failed from the competing risk. Yan Y. et al (2000) compared two methods of incidence estimates for AIDS related illnesses: one from unadjusted Kaplan-Meier and Cox proportional hazards models which did not adjust for death from competing risks (Model 1), and another from corresponding multiple decrement models that adjusted for competing risks (Model 2). Model 1 incidence estimates can substantially overestimate the occurrence of the diseases of interest; and the magnitude of overestimation depends on the follow-up time and the characteristics of study subjects. Tai B.C. et al. (2001) compared four methods for estimating survival probabilities, namely Kaplan Meier Ignore-all (KM-I), Kaplan-Meier Censor-all (KM-C), Cumulative Incidence Curve Risk (CICR), and LunnMc-Neil Unstratified approaches (LM-U). The estimates of the KM-I and KM-C methods, were consistently larger as compared with the CICR rates; the LM-U method consistently produced cumulative incidence rates that were comparable to the CICR. Alberti C. et al. (2003) compared 3 methods, 1- event free survival (1-EFS), 1-KM and CIC for estimating cumulative incidence for kidney graft loss with death before graft loss as competing event. They indicated overestimation of graft loss by the 1 – KM method, and the difference increased slightly with time. Satagopan JM et al. (2004) used two published data sets to compare the two methods 1-KM versus the CIC in cumulative incidence estimate. They concluded that the estimated cumulative incidence of an event of interest derived using the Kaplan-Meier approach was, in general, larger than estimates obtained when accounting for competing risks. Southern D.A. et al. (2006) estimated time to coronary artery bypass grafting (CABG), with death as competing event. They used “K-M censor all method,” “K-M censor death only method,” “K-M ignore all method,” and the CICR method. They concluded that, the CICR method is most appropriate, as KM methods will tend to overestimate event rates. Abdollah F. et al (2011) estimated cancer specific mortality (CSM) and other-cause mortality (OCM) rates in prostate cancer (PCa) patients. Cumulative incidence smoothed plots were generated for CSM and OCM according to the competing-risks method. All statistical

analyses were performed using the R statistical package. Their hypothesis states that significant CSM differences may exist according to treatment type. Because an important proportion of PCa patients succumb to other causes, they relied on competing-risks analyses to estimate CSM and to compare it among different treatment types. Campigotto F. et al (2012) studied the risk factors and the role of anticoagulants in the management of venous thromboembolic events (VTE) in cancer patients. A significant percentage of patients enrolled in these cancer-anticoagulant studies ultimately die from their underlying illness. They concluded that analyzing these cohorts based on Kaplan–Meier analysis is biased because study subjects who die prior to the development of a VTE are censored.

Despite this, it is still quite common to see inappropriate methods used to estimate such probabilities for endpoints that suffer from competing risks [3]. In particular, it is not unusual to see researchers utilize the complement of a Kaplan–Meier estimate (1-KM) to represent the probability of occurrence of a specified endpoint, even if the appropriate data contain competing-risk events. The common misuse of this methodology, (1-KM), for estimation purposes stems from a lack of thorough understanding among clinical researchers of the assumptions required to obtain interpretable Kaplan–Meier estimates. A lack of knowledge concerning the mechanics of calculating the cumulative incidence curves (CIC) estimates also likely contributes to the misuse. Moreover, the computation of a CICR is currently not included in main stream commercially available ready made statistical packages [6].

2- Methods for estimation of the survival probabilities

In this section, we describe estimating the survival probabilities or the cumulative incidence for an event of interest in the presence of competing risk events. We first review methods that have been suggested but which are not recommended such as **survival at a fixed time point**. At their worst, these methods can give a very misleading impression of the true situation. One method is to report the proportion of patients alive at a fixed time-point. Another possibility is only to use information on those patients that have reached a fixed point in the observation time or that have died before this time [12].

The Kaplan–Meier estimate of survival, also known as the product-limit estimate, provides a nonparametric estimate of the overall survival probability of an event of interest. The cumulative incidence is then calculated as one minus this survival probability. Every patient in the data set has a follow-up time and status (event or censored). The follow-up times where an event has occurred are ordered from the smallest to the largest (noting that there can be ties since more than one patient may have the event at the same follow-up time). Consider consecutive event times t_{j-1} and t_j . The Kaplan–Meier estimate of the overall survival probability up to event time t_j proceeds as follows. Let n_j be the number of event-free individuals up to time t_j . Suppose d_j events have occurred at time t_j . The estimated survival probability at time t_j is given by the ratio $(n_j - d_j)/n_j$. The overall survival probability up to time t_j , denoted $S(t_j)$, is the probability of surviving up to and including time t_j . Therefore, the overall survival probability up to t_j is estimated as the product of the probabilities of survival in all the previous times: $S(t_j) = \prod (n_i - d_i)/n_i = S(t_{j-1}) \times (n_j - d_j)/n_j$; where the product is over $i = 1, \dots, j$. The overall survival probability for any time between t_{j-1} and t_j is the same as $S(t_{j-1})$, the survival probability up to, but not including, t_j [4, 8, 11, 16].

The cumulative incidence curve (CIC) is a nonparametric estimation of cumulative incidence of the event of interest taking into account the informative nature

of censoring due to competing risks [11]. If competing risks are present, then hazards for each of two types of failure exist. Therefore, the number of failures from the competing risk will influence the number of failures from the cause of interest, and consequently the estimate of the probability of failure from this cause. Failures from the competing risk reduce the number of patients at risk of failure from the cause of interest. The step-by-step calculations to construct CIC are detailed below.

- 1- We first estimate the hazard at ordered time point t_j when the event of interest (c) occurs. This hazard estimate is simply the number of events that occur at t_j divided by the number at risk at t_j (analogous to the KM estimate). We can write this as

$$\widehat{h}_c(t_j) = \frac{m_{cj}}{n_j}$$

Where the m_{cj} denotes the number of events for risk c at time t_j and n_j is the number of subjects at that time. Thus, at any particular time, m_{cj}/n_j is the estimated proportion of subjects failing from risk c.

- 2- To be able to fail at time t_j , the subject needs to be "around to fail"; that is, he must have survived the previous time when a failure occurred. The probability of surviving the previous time t_{j-1} is denoted $\widehat{S}(t_{j-1})$, where $\widehat{S}(t)$ denotes the overall survival curve rather than the cause-specific survival curve $\widehat{S}_c(t)$. We must consider "overall" survival here, because the subject must have survived all other competing events.

- 3- The probability (i.e., incidence) of failing from event-type c at time t_j is then simply the probability of surviving the previous time period multiplied by $\widehat{h}_c(t_j)$.

$$\widehat{I}_c(t_j) = \widehat{S}(t_{j-1}) \times \widehat{h}_c(t_j)$$

- 4- The cumulative incidence at time t_j is then the cumulative sum up to time t_j , (i.e., from $j = 1$ to $j = j$) of these incidence values over all event-type c failure times.

$$\begin{aligned} CIC_c(t_j) &= \sum_{j=1}^j \widehat{I}_c(t_{j\Box}) \\ &= \sum_{j=1}^j \widehat{S}(t_{j\Box-1}) \times \widehat{h}_c(t_{j\Box})(1) \end{aligned}$$

Although, as previously mentioned, the CIC is equal to $1 - KM$ when there are no competing risks, formula (1) for $CIC_c(t_j)$ differs from $1 - KM_c$ when there are competing risks. In particular, the CIC formula (1) uses the overall survival function $\widehat{S}(t)$ that counts events from competing risks in addition to the event-type of interest as failures. In contrast, the formula for $1 - KM_c$ uses the event-type-specific survival function $\widehat{S}_c(t)$, which treats failures from competing risks as censored observations; this formula has been called the censor method [11, 16].

3- Application to Egyptian data

Our study comprised a retrospective cohort of 348 patients with primary cancer of the oral cavity diagnosed at Tanta cancer center (TCC) during the period 2003-2007. For survival analysis, the patients were followed from their date of diagnosis to early 2011. Survival is defined as the time from diagnosis until 31 May 2011 (end of follow up period) or until death from cancer, whichever occurred first. Survival was censored if

a patient was alive at the end of follow-up, lost to follow-up, or withdrawn from the study. Ethical approval for the study was given by the Research Ethical Committee (REC) of Ministry of Health and Population – Central Directorate for Research and Health Development– and also from the Medical Research Ethics Committee (MREC) of the National Research Center (NRC) Cairo-Egypt.

We first aimed the medical records of 479 patients diagnosed with primary tumors of the oral cavity - coded according to the International Classification of Diseases for Oncology, third edition (ICD-O-3 codes C00-C10) - who were admitted to TCC between January 2003 and December 2007. Patients where there were no data available on survival in their medical records, and those diagnosed with benign tumors were excluded. Of the 479 patients 131 were excluded and the remaining 348 were included in the study.

Information abstracted from patient records included age, sex, smoking status, the date of diagnosis, basis on which diagnosis was made, primary tumor site, and histopathological types were retrieved from histopathological reports (tumor type), and laterality. National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results Program (SEER) Summary Staging of cancer was recorded, tumor grade according to American Joint Committee on Cancer (AJCC). Different treatment modalities (surgical, radiotherapy, chemotherapy, immune therapy, hormone therapy or any combinations) and the final dates of admission to the center were also recorded, as were the cause of death, if it happened in hospital.

Table (1) shows the base line characteristics of the 348 cases included in the study, the different variables used in the analysis with their categories and the reference category for each variable is specified. The analyses were performed using the R statistical software version 3.0.2 which is free software available at <http://www.R-project.org>.

4- Analysis of results and conclusive remarks

In this survival analysis, the cases were followed for a maximum period of 84 months with mean and standard deviation (SD) of 13 (15.3) months, and with range of (0 – 84) month. First the KM survival estimates at 1, 2 and 5 years for each variable were calculated as shown in table (2), and the log rank test was used to test the significance among the different levels of each variable, with the p-values as shown in the last column of table (2).

The overall 12 months KM survival was (68.1%), 24 months survival was (51.7%), and the 60 months survival was (28.2%), with the median over all survival of 31 months. Recall that in the KM method those who died of other cause were censored as those who were alive, lost to follow-up, withdrawn at their censored times or by the end of follow-up period which ever occurred first.

From the results recorded in table (2), it was shown that there was a high statistical significance among the different treatment modalities with p-value = 0.000. The different tumor sites showed statistical significance with p-value = 0.001, also different tumor stages revealed statistical significance with p-value = 0.002. The different age categories were statistically significant with p-value = 0.005. Different smoking statuses showed also statistical significance with p-value = 0.015. The subgroups having different tumor grades were only marginally significant with p-value = 0.0549. There was no statistical significance between males and females on survival with p-value = 0.66. Concerning the laterality of the tumor, and the

histopathological types, different subgroups were found to be highly non significant with p-values = 0.716 and 0.959 respectively.

The cumulative incidence competing risk (CICR) estimates of the event of interest (cancer death) were calculated in the presence of a competing risk (other cause death) and comparing these CICR estimates among different groups in the presence of competing risk using Grey's test [7]. Then these CICR estimates were compared with their counterparts of the standard survival analysis using the 1- KM approach as shown in table (3).

The overall 12 months (CICR) estimates – calculated while considering those who died of other causes as competing risk – was (30.7%), 24 months CI was (45.2%), and the 60 months CI was (62.2%). While the overall 12 months cumulative event incidence using the (1-KM) method was (31.9%), 24 months was (48.3%), and the 60 months was (71.8%). **It is clear at this point that the 1- KM overestimates the cumulative incidence probabilities and the magnitude of overestimation increases with time.**

From the results recorded in table (3), it was shown that there was a high statistical significance among the different treatment modalities with p-value = 0.000. The least incidence estimate for cancer death of 40.5% at 2 years was recorded for those who received combination of different treatment modalities, this estimate was smaller than its counterpart using the 1-KM approach of 43.6%. CICR estimate at 5 years for those who received combination was 53.5% which was smaller than those who took single treatment option (70.3%), and also smaller than its counterpart using the 1-KM approach of 61.9%.

Different tumor stages revealed statistical significance (p-value = 0.005), with the least incidence rate of 32.7% at 2 years was recorded for those with localized tumor stages, which was also smaller than its counterpart using the 1-KM approach of 33.7%.

The different age categories were statistically significant (p-value = 0.016), the least CICR estimate of 34.1% at 2 years was recorded for patients less than 40 years of age, this incidence estimate was smaller than that obtained using the 1-KM method (34.6%).

Different smoking statuses showed also statistical significance (p-value = 0.015), the least CICR estimate of 39.2% at 2 years was for the non smoker group, which was smaller than its counterpart using the 1-KM approach of 43.1%.

The subgroups having different tumor grades were only marginally significant with p-value = 0.049. There was no statistical significance between males and females on survival with p-value = 0.96. Also different subgroups concerning tumor site, the laterality of the tumor, and the histopathological type were highly non significant with p-values = 0.117, 0.659 and 0.929 respectively.

From Table (3), on observation, even for predictors where there were no significant effect between its levels on survival, it was noticed that the 1-KM method always overestimate the cumulative incidence of an event of interest, the magnitude of overestimation varied with baseline patient characteristics and follow-up duration.

On comparing the relative efficiency of CICR estimator relative to 1-KM estimator for cumulative incidence of the event of interest (died from cancer causes) in presense of a competing risk (died from other causes), the CICR estimators was found to be more efficient as shown in table (4).

After initial descriptive statistics, simple analyses of the relations between survival and age, sex, smoking, primary site, grade, laterality, tumor stage, treatment modalities, and histopathological type were evaluated using simple unadjusted Cox

model. All variables are factors with a baseline (reference category) specified for each as shown in table (1), so each p-value provided in the 6th column of table (5) represented a test for the difference of each level with respect to the baseline. An overall p-value for each variable is required when modeling a factor with more than two levels as shown in the 7th column of table (5), it can be obtained through the Wald test.

The results of the simple analysis were shown in table (5). Sex, laterality and type were not associated with survival and were highly non significant. Age, smoke, site, treatment, and stage, however, related to survival. Tumor grade was only marginally significant ($p = 0.095$).

From table (5) it can be seen that: younger patients survived longer (p -value = 0.006), with the hazard of those who were 60 years of age or more were twice the hazard of those with less than 40 years ($HR = 2.134$, 95% CI = 1.1516 to 3.953).

Non smokers also enjoyed longer survival periods (p -value = 0.016), with the hazard of smokers was 1.76 times that of non smokers ($HR = 1.76$, 95% CI = 1.1608 to 2.675).

Tumor site has statistically significant effect on survival (p -value = 0.005), with the hazards of malignant neoplasms affecting the base of the tongue and the oropharynx were much higher than the hazard of those affecting the lips, ($HR = 8.7$, 95% CI = 2.5804 to 29.373) and ($HR = 5.79$, 95% CI = 1.8718 to 17.958) respectively.

Patients received no treatment were more likely to die sooner (p -value = 0.000), with those who received combination had better prognosis than those who received no treatment ($HR = 0.3239$, 95% CI = 0.193 to 0.5438). Also those who received single treatment had better prognosis than those who received no treatment ($HR = 0.4366$, 95% CI = 0.248 to 0.7689).

Tumor stage was a statistically significant predictor on survival (p -value = 0.002), with the hazard of those who had distant metastasis was 2.78 times the hazard for localized tumors ($HR = 2.788$, 95% CI = 1.528 to 5.088).

The results of fitting simple unadjusted competing risk regression CRR model to show the effect of the independent variables on survival were shown in table (6) using unadjusted Fine and Grey subdistribution hazard model [5]. Sex, site, laterality and type were not associated with survival. Age, smoke, grade, treatment, and stage, however, had statistical significant effect on survival with p -values equal to 0.014, 0.023, 0.0, 0.000 and 0.006 respectively.

Older people faced worse prognosis than younger ones. The hazard of those who were 60 years of age or more were almost twice the hazard of those with less than 40 years ($HR = 1.85$, 95% CI = 1.028 to 3.34).

Also smokers had worse prognosis than non smokers, with the hazard of smokers was 1.71 times that of non smokers ($HR = 1.71$, 95% CI = 1.13 to 2.58).

Patients who developed tumors of grade IV and V had higher hazards compared to those developed grade I ($HR = 3.908$, 95% CI = 1.585 to 9.64) and ($HR = 9.537$, 95% CI = 5.706 to 15.94) respectively.

Patients who received combination of different treatment options had better prognosis than those who received no treatment ($HR = 0.339$, 95% CI = 0.202 to 0.566). Also those who received single treatment had better prognosis than those who received no treatment ($HR = 0.445$, 95% CI = 0.253 to 0.785).

Patients who developed distant metastasis had worse prognosis than those with localized tumors ($HR = 2.62$, 95% CI = 1.439 to 4.78).

At this stage when comparing the relative efficiency of the estimators from simple unadjusted CRR models relative to those estimators obtained from simple

unadjusted Cox, it was found that the CRR estimators were more efficient as shown in table (7).

In clinical studies, we often seek to estimate the incidence of an event of interest. However, in addition to this event of interest, the patients are usually exposed to other events that can preclude occurrence of the event of interest, for instance by causing death. This competing-risk setting is extremely common in the medical literature. The KM method is still widely used to estimate the incidence of an event of interest in competing-risk settings. It provides an estimate of failure over time when patients who could not have developed the event of interest at the time of analysis are censored. In many settings where patients are followed for the occurrence of the disease of interest, there are two types of censoring, informative and noninformative. Censoring is noninformative for those who are alive without disease at the end of the study if enrollment date is independent of the disease progression, and for those who are lost during follow up if the underlying reason is not related to the disease of interest. However, censoring is informative for those who die previous to developing the disease because we know those patients will never develop the disease. Censoring those who die without the development of the disease does not make any biological sense, and violates the underlying assumption for valid Kaplan-Meier incidence estimates and the analogous Cox proportional hazards model incidence estimates [6].

Although the KM method was first developed to analyze a single time-related event such as death, two KM-based approaches were subsequently considered to hold appeal for analyzing competing-risk data. One of these approaches equates all competing events to the event of interest, thus yielding the complement of event-free survival (1-EFS). The other most widely used KM-based approach censors patients in whom competing events occur, together with those who remain free of the event of interest at the end of follow-up (i.e., who are withdrawn prematurely or lost to follow-up). However, the resulting 1-KM curves are misleading, because they rest on the assumption that time to censoring and time to failure are independent [2,17].

In real-life settings, where patients are exposed to many competing risks, the CI estimate has been proposed as the appropriate tool for estimating the incidence of any event [1,13,15]. Indeed, such curves provide straightforward probabilities that are easily understood by practitioners. The CI estimate corrects the overestimation observed with the 1-KM method. This overestimation can be large when many patients experience competing events before the event of interest, also the magnitude of overestimation depends on the follow-up time and the characteristics of study subjects. Nevertheless, overestimation of the incidence of an outcome of interest can have detrimental consequences. When comparing the incidence of events among patient subgroups, if the overestimation with the 1-KM method is of greater magnitude in one subgroup, the comparison could be biased [19, 20]. When follow-up time is short, or if the competing risk is low, the difference between traditional survival analysis and competing risk approach may not be substantial.

In summary, when estimating the incidence of an event of interest over time based on censored data in a competing risk setting, one should keep in mind that 1-KM, 1-EFS, and CI curves address different questions. The 1-KM curve tells us the probability of the event of interest assuming that there is no risk of death before that event (as patients who die before the event are censored), and the 1-EFS curve gives us the probability of survival free of event (as the event and death unrelated to the event are lumped together). For estimating the actual probability of an event in a competing-risk setting, the CI method that estimates the actual proportion of patients experiencing the event of interest, is the best tool [2].

On conclusion: failures from the competing risk reduce the number of patients at risk of failure from the cause of interest. The Kaplan–Meier approach overestimates the probability of failure and therefore overestimates the cumulative incidence function. On the other hand, the competing risk approach treats a competing event as an actual event rather than a censored observation. We conclude that the Kaplan–Meier analysis is an inappropriate statistical methodology when patients are subjected to competing risks.

Table (1):Base line data.

Variable	No. (%)	Variable	No.(%)
sex		Basis of Diagnosis	
male ^a	204 (58.6)	Clinical only	3 (0.9)
female	144 (41.4)	Cytology / Hematology	9 (2.6)
Age (years)		Histopathology of Metastasis	4 (1.1)
< 40 ^a	54 (15.5)	Histopathology of Primary	332 (95.4)
40 – 59	126 (36.2)	Tumor Grade	
≥ 60	168 (48.3)	Grade I ^a	60 (17.2)
Smoking Status		Grade II	138 (39.7)
non smoker ^a	137 (39.4)	Grade III	26 (7.5)
smoker	99 (28.4)	Grade IV	5 (1.4)
Unknown	112 (32.2)	Grade V, T-cell	1 (0.3)
Primary Tumor Site		Grade VI, B-cell	6 (1.7)
MN* of Lip ^a	52 (14.9)	Unspecified	112 (32.2)
MN* of base of Tongue	6 (1.7)	Laterality	
MN* of other & unspecified parts of tongue	67 (19.3)	Not paired ^a	158 (45.4)
MN* of Gum	52 (14.9)	Unilateral	179 (51.4)
MN* of Floor of the Mouth	8 (2.3)	Bilateral	2 (0.6)
MN* of Palate	15 (4.3)	Unknown / Midline	9 (2.6)
MN* of other & unspecified parts of Mouth	38 (10.9)	Treatment Modalities	
MN* of Parotid Gland	44 (12.6)	None given ^a	41 (11.8)
MN* of other & unspecified parts of MSG**	16 (4.6)	Single treatment	137 (39.4)
MN* of Tonsils	42 (12.1)	Combination	170 (48.8)
MN* of Oropharynx	8 (2.3)	Patient Status	
SEER Stage		Dead	150 (43.1)
Localized ^a (Stage I Lymphoma)	143 (41.1)	Alive	140 (40.2)
Regional	138 (39.7)	Lost to follow	58 (16.7)
Distant (St.III, IV Lymphoma)	31 (8.9)	Cause of Death	
Unknown	36 (10.3)	Cancer	116 (33.3)
Histopathological type		Other	34 (9.8)
SCC***	193 (55.5)		
Non- SCC	155 (44.5)		

* MN: Malignant neoplasm, **MSG: major salivary glands, *** SCC: Squamous cell carcinoma. ^a Reference category.

Table (2): The unadjusted KM survival estimates and comparing them across strata using log rank test.

Variable	N	Survival Rates %			P-value
		12 months	24 months	60 months	
Sex					
Male	204	65.4	50.9	33.1	0.66
Female	144	71.3	52.7	25.4	
Age					
< 40	54	81.9	65.4	NA	0.005
40 – 59	126	76.0	57.4	33.0	
>= 60	168	58.0	43.2	20.9	
Smoke					
Non smoker	137	73.4	56.9	33.6	.015
Smoker	99	54.6	43.1	19.3	
Unknown	112	76.2	54.1	33.0	
Tumor Grade					
I	60	83.0	70.9	NA	0.0549
II	138	63.9	41.9	26.3	
III	26	68.2	54.6	NA	
IV	5	60.0	NA	NA	
V	1	NA	NA	NA	
VI	6	75.0	75.0	NA	
Unspecified	112	67.7	56.5	39.1	
Tumor Site					
MN ^a of Lip	52	74.0	74.0	49.3	.001
MN ^a of base of Tongue	6	33.3	NA	NA	
MN ^a of other&unspec parts of tongue	67	67.2	52.5	NA	
MN ^a of Gum	52	59.0	42.8	20.3	
MN ^a of Floor of the Mouth	8	83.3	41.7	NA	
MN ^a of Palate	15	50.5	40.4	NA	
MN ^a of other&unspec. Parts of Mouth	38	90.0	45.8	NA	
MN ^a of Parotid Gland	44	83.2	65.1	43.4	
MN ^a of other&unspec. Parts of MSG*	16	93.8	72.9	NA	
MN ^a of Tonsils	42	53.9	49.4	49.4	
MN ^a of Oropharynx	8	23.8	NA	NA	
Treatment					
None given	41	47.5	28.5	NA	0.000
Single treatment	137	64.3	49.7	19.8	
Combination	170	72.9	56.4	38.1	
Laterality					
Not paired	158	68.4	55.6	23.7	0.716
Unilateral	179	67.2	48.3	31.3	
Bilateral	2	1.0	1.0	NA	
Unknown	9	88.9	NA	NA	
SEER stage					
Localized	143	77.0	66.3	25.1	0.002
Regional	138	66.5	47.2	42.6	
Distant	31	49.1	34.4	NA	
Unknown	36	62.2	31.1	NA	
Histological type					
SCC	193	69.0	48.7	27.7	0.959
Non-SCC	155	66.9	55.5	31.2	

^a MN: Malignant Neoplasm.

Table(3):The Unadjusted cumulative incidence estimates of death due to cancer for patients with cancer of the oral and oropharyngeal cancers by sex, age, smoking status, grade, site, treatment, laterality, stage, and type of tumor, comparing them with their 1-KM counterparts.

Variable	N	1-KM I				CICR I			
		Cumulative incidence Rates %			P-value	Cumulative incidence Rates%			P-value
		12 months	24 months	60 months		12 months	24 months	60 months	
Sex									
Male	204	34.6	49.1	66.9	0.66	33.0	45.4	57.0	0.96
Female	144	28.7	47.3	74.6		27.9	45.2	67.8	
Age									
< 40	54	18.1	34.6	NA	0.005	18.1	34.1	NA	0.016
40 – 59	126	24.0	42.6	67		23.0	39.4	56.2	
>= 60	168	42.0	56.8	79.1		39.8	52.5	67.4	
Smoke									
Non smoker	137	26.6	43.1	66.4	0.015	25.1	39.2	57.5	0.015
Smoker	99	45.4	56.9	80.7		44.5	54.8	69.5	
Unknown	112	23.8	45.9	67.0		22.9	43.2	58.0	
Tumor Grade									
I	60	17.0	29.1	NA	0.061	16.7	27.6	NA	0.049
II	138	36.1	58.1	73.7		35.0	54.1	64.4	
III	26	31.8	45.4	NA		30.3	42.0	NA	
IV	5	40.0	NA	NA		40.0	NA	NA	
V	1	NA	NA	NA		NA	NA	NA	
VI	6	25.0	25.0	NA		20.8	20.8	NA	
Unspecified	112	32.3	43.5	60.9		30.6	40.6	53.6	
Tumor Site									
MN ^a of Lip	52	26.0	26.0	50.7	0.001	25.2	25.2	43.6	0.117
MN ^a of Tongue base	6	66.7	NA	NA		66.7	NA	NA	
MN ^a of other&unspec parts of tongue	67	32.8	47.5	NA		32.8	45.3	NA	
MN ^a of Gum	52	41.0	57.2	79.7		39.0	52.1	70.0	
MN ^a of Floor of Mouth	8	16.7	58.3	NA		14.3	42.9	NA	
MN ^a of Palate	15	49.5	59.6	NA		41.9	50.4	NA	
MN ^a of other&unspec. Parts of Mouth	38	10.0	54.2	NA		9.3	48.2	NA	
MN ^a of Parotid Gland	44	16.8	34.9	56.6		16.8	34.0	51.4	
MN ^a of other&unspec. Parts of MSG*	16	6.2	27.1	NA		6.3	27.1	NA	
MN ^a of Tonsils	42	46.1	50.6	50.6		44.2	48.5	48.5	
MN ^a of Oropharynx	8	76.2	NA	NA		64.3	NA	NA	
Treatment									
None given	41	52.5	71.5	NA	0.000	51.2	69.4	NA	0.000
Single treatment	137	35.7	50.3	80.2		33.9	47.4	70.3	
Combination	170	27.1	43.6	61.9		26.0	40.5	53.5	
Laterality									
Not paired	158	31.6	44.4	76.3	0.716	30.0	41.5	61.0	0.659
Unilateral	179	32.8	51.7	68.7		31.7	48.1	61.4	
Bilateral	2	0.0	0.0	NA		0.0	0.0	NA	
Unknown	9	11.1	NA	NA		11.1	NA	NA	
SEER stage									
Localized	143	23	33.7	74.9	0.002	22.6	32.7	60.2	0.005
Regional	138	33.5	52.8	57.4		31.6	48.2	51.7	
Distant	31	50.9	65.6	NA		50.9	62.7	NA	
Unknown	36	37.8	68.9	NA		33.9	58.9	NA	
Histological type									
SCC	193	31.0	51.3	72.3	0.957	29.8	47.5	61.5	0.929
Non-SCC	155	33.1	44.5	68.8		31.6	42.0	61.2	

^a MN: Malignant Neoplasm.

Table (4): The relative efficiency of CICR estimator relative to KM estimator.

Variables	KM estimate			CICR estimate			RE	RE	RE
	SE 12 mon.	SE 24 mon.	SE 60 mon.	SE 12 mon.	SE 24 mon.	SE 60 mon.	12 mon.	24 mon.	60 mon.
Sex (male)	0.0654	.509	.331	0.04175	0.04619	0.06116	1.56647	11.02	5.4120
Sex (female)	0.0456	.0560	0.0934	0.04442	0.05433	0.08929	1.02656	1.03	1.0460
Age (< 40)	0.0675	0.0918	-	0.06862	0.09244	-	.98368	.99	
Age (40-59)	0.0480	0.0612	0.1059	0.04638	0.05760	0.08595	1.03493	1.06	1.2321
Age (>=60)	0.0480	0.0531	0.0806	0.04605	0.04997	0.06798	1.04235	1.06	1.1856
Smoke (none)	0.0451	0.0565	0.0971	0.04282	0.05212	0.08631	1.05325	1.08	1.1250
Smoke (smoker)	0.0594	0.0633	0.0964	0.05903	0.06183	0.08222	1.00627	1.02	1.1725
Smoke(Unknown)	0.0592	0.0757	0.1171	0.05770	0.07284	0.10181	1.02600	1.04	1.1502
Grade (I)	0.0696	0.0884	-	0.06979	0.08529	-	.99728	1.04	
Grade (II)	0.0506	0.0610	0.0864	0.04955	0.05784	0.07286	1.02119	1.05	1.1858
Grade (III)	0.109	0.123		0.10881	0.12050	-	1.00175	1.02	-
Grade (IV)	0.219	-	-	0.25000	-	-	.87600	-	-
Grade (V)	-	-	-	-	-	-	-	-	-
Grade (VI)	0.217	0.217		0.21246	0.21246	-	1.02137	1.02	-
Grade (unspecified)	0.0556	0.0625	0.0938	0.05294	0.05925	0.08079	1.05025	1.05	1.1610
Site (lip)	0.096	0.096	0.211	0.09576	0.09576	0.20116	1.00251	1.00	1.0489
Site (tongue base)	0.192	-	-	0.22696	-	-	.84596	-	-
Site (tongue other)	0.0703	0.0863	-	0.07135	0.08283	-	.98528	1.04	-
Site (gum)	0.0841	0.0924	0.1068	0.08177	0.08775	0.11068	1.02849	1.05	.9649
Site (mouth floor)	0.152	0.222	-	0.14483	0.21686	-	1.04951	1.02	-
Site (palate)	0.158	0.155	-	0.15311	0.15640	-	1.03194	.99	-
Site (mouth other)	0.0671	0.1248	-	0.06393	0.12103	-	1.04959	1.03	-
Site (parotid gland)	0.0631	0.0941	0.1880	0.06408	0.09405	0.18843	.98471	1.00	.9977
Site (MSG other)	0.0605	0.1382	-	0.06250	0.14670	-	.96800	.94	
Site (tonsils)	0.0926	0.0951	0.0951	0.09123	0.09415	0.09415	1.01502	1.01	1.0101
Site (oropharynx)	0.2026	-	-	0.25035	-	-	.80927	-	-
Treatment (none)	0.107	0.107	-	0.10783	0.11314	-	.99230	.95	-
Treatment (single)	0.0634	0.0720	0.1464	0.06102	0.06977	0.16060	1.03900	1.03	.9116
Treatment(combination)	0.0380	0.0463	0.0739	0.03677	0.04338	0.06029	1.03345	1.07	1.2257
Laterality (notpaired)	0.0489	0.0558	0.1006	0.04683	0.05297	0.07962	1.04420	1.05	1.2635
Laterality (unilateral)	0.0423	0.0506	0.0773	0.04115	0.04759	0.06940	1.02795	1.06	1.1138
Laterality (bilateral)	0	0	-	0	0	-	-	-	-
Laterality(unknown)	0.105	-	-	0.11111	-	-	.94501	-	-
SEER (Local)	0.0468	0.0575	0.1237	0.04636	0.05632	0.10478	1.00949	1.02	1.1806
SEER (regional)	0.0464	0.0539	0.0586	0.04427	0.05032	0.05363	1.04811	1.07	1.0927
SEER (distant)	0.106	0.116	-	0.11011	0.11618	-	.96267	1.00	-
SEER (unknown)	0.131	0.143	-	0.12249	0.14285	-	1.06948	1.00	-
Type (SCC)	0.0417	0.0507	0.0781	0.04046	0.04742	0.06365	1.03065	1.07	1.2270
Type (non SCC)	0.0480	0.0543	0.0925	0.04624	0.05201	0.08284	1.03806	1.04	1.1166

$$R. E. = \frac{S. E. \text{ of KM}}{S. E. \text{ of CICR}}$$

Table (5): Results of fitting simple unadjusted Cox's model.

Variable	Coef.(β)	HR.	SE for (β)	95% CI of HR	P-value	Wald
Sex						
Male ^b						
Female	-0.08239	0.9202	0.18732	0.6379-1.329	0.66	0.66
Age						
< 40 ^b						
40 – 59	0.223	1.249	0.334	0.649-2.407	0.505	0.006
>= 60	0.758	2.134	0.315	1.1516-3.953	0.016	
Smoke						
Non smoker ^b						
Smoker	0.56648	1.76205	0.21296	1.1608-2.675	0.00781	0.016
Unknown	0.04071	1.04155	0.24711	0.6417-1.691	0.86914	
Tumor Grade						
I						
II	0.6175	1.8543	0.3163	0.99755-3.447	0.0509	
III	0.5318	1.7020	0.4364	0.72367-4.003	0.2229	
IV	1.3772	3.9640	0.5770	1.27947-12.281	0.0170	0.095
V	2.3389	10.3701	1.0478	1.33006-80.852	0.0256	
VI	-0.4068	0.6658	1.0386	0.08695-5.097	0.6953	
Unspecified	0.4769	1.6110	0.3247	0.85260-3.044	0.1419	
Tumor Site						
MN ^a of Lip ^b						
MN ^a of base of Tongue	2.16402	8.70610	0.62045	2.5804-29.373	0.000487	
MN ^a of other&unspec parts of tongue	0.37692	1.45778	0.41701	0.6438-3.301	0.366071	
MN ^a of Gum	0.70928	2.03252	0.41462	0.9018-4.581	0.087138	
MN ^a of Floor of the Mouth	0.98842	2.68698	0.61467	0.8055-8.963	0.107822	
MN ^a of Palate	0.75341	2.12424	0.54226	0.7339-6.149	0.164711	0.005
MN ^a of other&unspec. Parts of Mouth	0.50635	1.65922	0.45832	0.6757-4.074	0.269251	
MN ^a of Parotid Gland	0.02217	1.02242	0.45852	0.4162-2.511	0.961428	
MN ^a of other&unspec. Parts of MSG*	0.15295	1.16526	0.57226	0.3796-3.577	0.789262	
MN ^a of Tonsils	0.65230	1.91994	0.43077	0.8253-4.466	0.129960	
MN ^a of Oropharynx	1.75749	5.79786	0.57683	1.8718-17.958	0.002313	
Treatment						
None given						
Single treatment	-0.8287	0.4366	0.2887	0.248-0.7689	0.0041	0.000
Combination	-1.1272	0.3239	0.2643	0.193-0.5438	0.000	
Laterality						
Not paired ^b						
Unilateral	0.1021	1.1075	0.1935	0.7579-1.618	0.598	
Bilateral	-0.4276	0.6521	1.0126	0.0896-4.745	0.673	0.729
Unknown	0.6913	1.9963	0.7249	0.4822-8.265	0.340	
SEER stage						
Localized ^b						
Regional	0.3210	1.3785	0.2209	0.894-2.126	0.146	0.002
Distant	1.0254	2.7881	0.3069	1.528-5.088	0.000	
Unknown	0.9152	2.4973	0.3496	1.258-4.956	0.009	
Histological type						
SCC ^b						
Non-SCC	-0.009487	0.990558	0.188341	0.6848-1.433	0.96	0.96

^aMalignant Neoplasm. ^b Reference category.

Table (6):Results of fitting simple unadjusted CR model (Fine & Grey subdistribution hazards model).

Variable	Coef. (β)	HR.	SE for (β)	95 % CI of HR	P-value	Wald
Sex						
Male						
Female	0.0144	1.01	0.181	0.712 to 1.45	0.94	0.94
Age						
< 40						
40 – 59	0.120	1.13	0.319	0.603 to 2.11	0.71	0.014
>= 60	0.617	1.85	0.301	1.028 to 3.34	0.04	
Smoke						
Non smoker						
Smoker	0.5366	1.71	0.209	1.13 to 2.58	0.01	0.023
Unknown	0.0565	1.06	0.233	0.67 to 1.67	0.81	
Tumor Grade						
I						
II	0.548	1.729	0.280	0.999 to 2.99	0.050	
III	0.432	1.540	0.419	0.677 to 3.50	0.300	
IV	1.363	3.908	0.460	1.585 to 9.64	0.003	0.0
V	2.255	9.537	0.262	5.706 to 15.94	0.000	
VI	-0.549	0.578	1.096	0.067 to 4.95	0.620	
Unspecified	0.436	1.547	0.294	0.869 to 2.75	0.140	
Tumor Site						
MN ^a of Lip						
MN ^a of base of Tongue	1.563	4.77	0.768	1.059 to 21.49	0.042	
MN ^a of other&unspec parts of tongue	0.337	1.40	0.399	0.641 to 3.06	0.400	
MN ^a of Gum	0.666	1.95	0.400	0.888 to 4.27	0.096	
MN ^a of Floor of the Mouth	0.626	1.87	0.547	0.640 to 5.47	0.250	
MN ^a of Palate	0.511	1.67	0.531	0.589 to 4.72	0.340	0.24
MN ^a of other&unspec. Parts of Mouth	0.432	1.54	0.407	0.694 to 3.42	0.290	
MN ^a of Parotid Gland	0.124	1.13	0.444	0.474 to 2.70	0.780	
MN ^a of other&unspec. Parts of MSG ^b	0.188	1.21	0.514	0.440 to 3.31	0.710	
MN ^a of Tonsils	0.621	1.86	0.430	0.801 to 4.32	0.150	
MN ^a of Oropharynx	1.382	3.98	0.543	1.374 to 11.54	0.011	
Treatment						
None given						
Single treatment	-0.809	0.445	0.289	0.253 to 0.785	0.005	0.000
Combination	-1.083	0.339	0.262	0.202 to 0.566	0.000	
Laterality						
Not paired						
Unilateral	0.156	1.169	0.185	0.813 to 1.68	0.40	
Bilateral	-0.246	0.782	0.732	0.186 to 3.28	0.74	0.58
Unknown	0.781	2.184	0.688	0.567 to 8.41	0.26	
SEER stage						
Localized						
Regional	0.291	1.34	0.213	0.881 to 2.03	0.170	
Distant	0.964	2.62	0.306	1.439 to 4.78	0.001	0.006
Unknown	0.717	2.05	0.311	1.114 to 3.77	0.021	
Histological type						
SCC						
Non-SCC	0.0243	1.02	0.183	0.716 to 1.47	0.89	0.89

^aMalignant Neoplasm. ^b Major Salivary Gland.

Table (7): Relative efficiency of CRR estimator relative to the Cox regression estimators.					
Variables	Simple unadjusted Cox regression models		Simple unadjusted Fine and Gray model		RE
	β	SE for (β)	β	SE for (β)	
Sex (female)	-0.08239	0.18732	0.0144	0.181	1.03
Age (40-59)	0.223	0.334	0.120	0.319	1.05
Age (≥ 60)	0.758	0.315	0.617	0.301	1.05
Smoke (smoker)	0.56648	0.21296	0.5366	0.209	1.02
Smoke (Unknown)	0.04071	0.24711	0.0565	0.233	1.06
Grade (II)	0.6175	0.3163	0.548	0.280	1.13
Grade (III)	0.5318	0.4364	0.432	0.419	1.04
Grade (IV)	1.3772	0.5770	1.363	0.460	1.25
Grade (V)	2.3389	1.0478	2.255	0.262	4.00
Grade (VI)	-0.4068	1.0386	-0.549	1.096	.95
Grade (unspecified)	0.4769	0.3247	0.436	0.294	1.10
Site (tongue base)	2.16402	0.62045	1.563	0.768	.81
Site (tongue other)	0.37692	0.41701	0.337	0.399	1.05
Site (gum)	0.70928	0.41462	0.666	0.400	1.04
Site (mouth floor)	0.98842	0.61467	0.626	0.547	1.12
Site (palate)	0.75341	0.54226	0.511	0.531	1.02
Site (mouth other)	0.50635	0.45832	0.432	0.407	1.13
Site (parotid gland)	0.02217	0.45852	0.124	0.444	1.03
Site (MSG other)	0.15295	0.57226	0.188	0.514	1.11
Site (tonsils)	0.65230	0.43077	0.621	0.430	1.00
Site (oropharynx)	1.75749	0.57683	1.382	0.543	1.06
Treatment (single)	-0.8287	0.2887	-0.809	0.289	1.00
Treatment (combination)	-1.1272	0.2643	-1.083	0.262	1.01
Laterality (unilateral)	0.1021	0.1935	0.156	0.185	1.05
Laterality (bilateral)	-0.4276	1.0126	-0.246	0.732	1.38
Laterality (unknown)	0.6913	0.7249	0.781	0.688	1.05
SEER stage (regional)	0.3210	0.2209	0.291	0.213	1.04
SEER stage (distant)	1.0254	0.3069	0.964	0.306	1.00
SEER stage (unknown)	0.9152	0.3496	0.717	0.311	1.12
Type (non SCC)	-0.009487	0.188341	0.0243	0.183	1.03

$$R. E. = \frac{\text{S. E. of } \beta \text{ in Cox model}}{\text{S. E. of } \beta \text{ in CRR model}}$$

References

- 1- Abdollah F., Sun M., Thuret R., Jeldres C., Tian Z., Briganti A., Shariat F.S., Perrotte P., Rigatti P., Montorsi F., and Karakiewicz P.I. (2011), "A Competing-Risks Analysis of Survival After Alternative Treatment Modalities for Prostate Cancer Patients: 1988–2006," *Journal of European Urology*, 59, 88-95.
- 2- Alberti C., Metivier F., Landais P., Thervet E., Legendre C., and Chevret S. (2003), "Improving Estimates Of Event Incidence Over Time In Populations Exposed To Other Events: Applications To Three Large Databases," *Journal of Clinical Epidemiology*, 56, 536-545.
- 3- Campigotto F., Neuberg D., and Zwicker JI. (2012), "Biased Estimation of Thrombosis Rates in Cancer Studies using the Method of Kaplan and Meier," *Journal of Thrombosis and Haemostasis*, 10, 1449–14451.
- 4- Costella J.P. (2010), "A simple alternative to Kaplan–Meier for survival curves," *Peter MacCallum Cancer Centre Working Paper No. sep/2010*.

- 5- Fine JP, and Gray RJ. (1999), "A Proportional Hazards Model for the Subdistribution Of A Competing Risk," Journal of the American Statistical Association, 94, 496-509.
- 6- Gooley T.A., Leisenring W., Crowley J., and Storer B.E. (1999), "Estimation of Failure Probabilities In The Presence Of Competing Risks: New Presentation Of Old Estimators," Statistics in Medicine, 18, 695-706.
- 7- Gray RJ. (1988), "A Class of K-Sample Tests for Comparing The Cumulative Incidence Of A Competing Risk," Annals of Statistics, 16, 1141-1154.
- 8- Kaplan E.L., and Meier P. (1958), "Non Parametric Estimation from incomplete observations," Journal of American Statistical Association, 457-481.
- 9- Katsahian S., Porcher R., Mary JY., and Chevert S. (2004), "The Graft-Versus-Leukaemia Effect after Allogeneic Bone- Marrow Transplantation: Assessment Through Competing Risks Approaches," Statistics in Medicine, 23, 3851-3863.
- 10- Klein J.P., and Moeschberger M.L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data* (2nd edition), New York: Springer-Verlag.
- 11- Kleinbaum D.G., and Klein M. (2012), *Survival Analysis: a Self Learning Text* (3rd edition), New York: Springer.
- 12- Machin D., Cheung Y.B., and Parmar M.K.B. (2006), *Survival Analysis: A Practical Approach* (2nd Edition), England: John Wiley & Sons.
- 13- Pepe MS., and Mori M. (1993), "Kaplan-Meier Marginal or Conditional Probability Curves In Summarizing Competing Risks Failure Time Data?," Statistics in Medicine, 12, 737-751.
- 14- Phoebe M. Abdel Massieh (expected 2014). Evaluation of Statistical Methods Used in The Analysis of Competing Risks in the Oro-Dental Field. Ph.D. Thesis. Institute of Statistical Studies and Research. Cairo. Egypt.
- 15- Pintilie M. (2011), "An Introduction to Competing Risks Analysis. Revista Española de Cardiología," 64(7), 599 – 605.
- 16- Satagopan JM., Ben-Porat L., Berwick M., Robson M., Kutler D., and Auerbach AD. (2004), "A Note on Competing Risks In Survival Data Analysis," British Journal of Cancer, 91, 1229-1235.
- 17- Southern D.A., Faris P.D., Brant R., Galbraith P.D., Norris C.M., Knudtson M.L., and Ghali W.A. (2006), "Kaplan Meier Methods Yielded Misleading Results in Competing Risk Scenarios," Journal of Clinical Epidemiology, 59, 1110-1114.
- 18- Tai BC., Machin D., White I., and Gebiski V. (2001), "Competing Risks Analysis Of Patients With Osteosarcoma: A Comparison Of Four Different Approaches," Statistics in Medicine, 20(5), 661-684.
- 19- Yan Y., Moore RD., and Hoover DR. (2000), "Competing Risk Adjustment Reduces Overestimation of Opportunistic Infection Rates In AIDS," Journal of Clinical Epidemiology, 53, 817-822.
- 20- Yan Y., Hoover DR., Moore RD., and Xiong C. (2003), "Multivariate Estimation Of Cumulative Incidence, Prevalence, And Morbidity Time Of A Disease When Death Is Likely," Journal of Clinical Epidemiology, 56, 546-552.

Using Rough Set Approach for Discovering Systemic Lupus Erythematosus in Pregnant Females and its Relation to Correlation Analysis

N. R.Hassan ^(a)

E.A.Hussein ^(b)

Y.M.Ibrahim^(c)

Abstract

Systemic Lupus Erythematosus is an autoimmune disease that mostly affects females in child bearing age. The process of discovering natural phenomena or complex system was limited, until recently, to finding formulas that fit empirical data. Artificial intelligence methods encompass various paradigms dedicated to solving real-world problems in decision making, pattern classification, and learning. Prominent among these paradigms are fuzzy sets, neural networks, genetic algorithms and rough sets which are used to model data in the form of decision tables or rules. The aim of this paper is to use one of the artificial intelligent methods such as rough sets to discover pregnant females with Systemic Lupus Erythematosus. Thus, it can help to improve medical care offered to pregnant females with better evaluation trying to improve maternal and fetal outcome. A predictive rough set model was built using real data, and the model was validated and its results were also homogenous to results of correlation analysis between Systemic Lupus Erthematosus and input variables.

Key words: Correlation analysis;Data mining; Machine learning; Rough set; Systemic Lupus Erythematosus

1. Introduction

Systemic lupus Erythematosus (SLE) is a multisystem autoimmune disease with a broad spectrum of clinical presentations, affecting mostly women in childbearing age. SLE is a chronic disease of variable severity with significant morbidity that can be fatal if not treated early. The fertility rate in patients with SLE is usually not affected. The impact of SLE on pregnancy and of pregnancy on maternal lupus has been reported by several studies but with variable results. Some reports indicate that there is an increase in SLE flares during pregnancy [8] and others found no difference in flares between pregnant and non-pregnant females with SLE [16].

Pregnancies in females with SLE have been reported to have higher rates of fetal loss, preterm delivery, preeclampsia (pregnancy induced hypertension) and intrauterine growth retardation.

Historically, maternal well being of SLE patients was compromised to the extent that the medical community recommended against pregnancy in SLE patients .Nowadays with better improvement in disease management activity, pregnancy in SLE patients is no longer an absolute contraindication. However, fetal and maternal complications still exist.

(a) Lecturer Of Operations Research, Institute of Statistical Studies and Research, Cairo University. (naglaa777subkiii@yahoo.com)

(b) Lecturer of Obstetrics and Gynecology, Kasr AL Aini Hospital, Cairo University. (emanalyhussein@yahoo.com)

(c) Lecturer Of Applied Statistics, Institute of Statistical Studies and Research,Cairo University. (yasminority@yahoo.com)

Careful planning of pregnancy coupled with multidisciplinary monitoring and treatment usually decreases the risks for the mother and the infant [2].

All patients who were diagnosed as SLE patients according to American College of Rheumatology ACR [15] fulfilled at least 4 of the 11 criteria which include:

1. Malar Rash
2. Discoid Rash
3. Photosensitivity
4. Oral Ulcers
5. Arthritis
6. Serositis (pleuritis or pericarditis)
7. Renal Disorders
8. Neurological disorder (seizures or psychosis)
9. Hematologic disorder (hemolytic anemia, Leukopenias, thrombocytopenia)
10. Immunological disorders, the presence of:
 - An abnormal titre of Antinuclear antibodies (ANA)
 - Anti -double stranded (ds) DNA antibodies
 - Positive finding of antiphospholipid antibodies: Anticardiolipin antibodies and / or lupus anticoagulant. In this study we used four laboratory investigations done to our pregnant patients.

Recently, most research efforts in machine learning have been directed towards the invention of new algorithms and much less into gaining experience in applying those to important practical applications. Therefore, this paper focuses on developing a system for doctors and clinicians to provide an insight into understanding the relations between complex medical data.

Various published algorithms have been applied to build a computer-aided analysis system in the medical field [4], [10]. The most commonly used algorithms are neural networks, Bayesian classifiers, genetic algorithms, decision trees, and fuzzy theory [11]–[14] and [17]. Unfortunately, the techniques developed have not been sufficient to introduce an effective computer aided analysis in clinical use.

Rough sets theory is another effective approach for managing uncertainties and it can be employed for tasks such as feature identification, dimensionality reduction, pattern classification and image segmentation.

In 2004, Bernsel A., presented new methods for association analysis based on Rough Set theory that were developed and successfully applied to both simulated and biological genotype data [1]. An estimation of the significance level of the results was assessed through a randomization procedure. The results here confirm earlier findings of two Single Nucleotide Polymorphisms (SNPs) possibly associated to Systemic Lupus Erythematosus (SLE). Additionally, one new SNP was found which might be significantly associated to SLE.

In 2005, Ilczuk et al., presented important criteria in the medical domain. They showed that attribute reduction performed as a pre-processing step before the generation of decision rules increases not only future classification accuracy but also dramatically decreases the time needed for this computation. Their experiments indicate that using the Wrapper attribute selection method together with different learning algorithms can find smaller subsets of strong attributes which performed with a similar or better accuracy during the classification [1].

Swiniarski and Hargis described an application of rough set methods to feature selection and reduction as a front-end to a neural-network-based texture image recognition system. Their application included singular value decomposition (SVD) for feature extraction, and rough sets methods for feature selection and reduction. For texture classification, a feed-forward back-propagation neural network was employed.

2. Material and Method

2.1 Rough Set Theory

Rough sets theory was introduced by Pawlak in 1982. It was developed based on mathematical tools to deal with vagueness and uncertainty in the classification of objects in a set [2]. The rough set philosophy is founded on the assumption that there is some information regarding features which can be associated with every object of the universe. In rough sets, the data is organized in a table called decision table, which are flat tables containing attributes as columns and data elements as rows. The class label is the decision attribute, the rest of the attributes are the condition attributes. For rows, rough set theory employs the notion of indiscernible class to group similar objects together; while for columns, it employs the notion of indiscernible attribute to identify significant attributes [6].

We call the class label a decision feature, the rest of the features are conditional. Let O, F denote a set of sample objects and a set of functions representing object features, respectively. Assume that $B \subseteq F, x \in O$. Further, let $[x]_B$ denote:

$$[x]_B = \{y : x \sim_B y\}$$

Rough set theory define three regions based on the equivalent classes induced by the feature values: lower approximation $\underline{B}X$, upper approximation $\overline{B}X$, and boundary $BND_B(X)$. A lower approximation of a set X contains all equivalence classes $[x]_B$ that are subsets of X , and upper approximation contains all equivalence classes $[x]_B$ that have objects in common with X , while the boundary is the set of all objects in $\overline{B}X$ that are not contained in $\underline{B}X$. So, we can define a rough set as any set with a non-empty boundary. The indiscernibility relation $\sim_B$ (or Ind_B) is a fundamental principle of rough set theory. Informally, $\sim_B$ is a set of all objects that have matching descriptions. Based on the selection of B , $\sim_B$ is an equivalence relation which partitions a set of objects O into equivalence classes. The set of all classes in a partition is denoted by $O/\sim_B$ (also by O/Ind_B). The set O/Ind_B is called the quotient set. Affinities between objects of interest in the set $X \subseteq O$ and classes in a partition can be discovered by identifying those classes that have objects in common with X . Approximation of the set X begins by determining which elementary sets $[X]_B \in O/\sim_B$ are subsets of X . For a detailed review of the basic material, reader may consult sources such as [5, 9].

2.2 Prediction model using rough sets:

Functionally, Rough Set Prediction Model (RSPM) can be partitioned into four distinct phases:

- **Data collection phase**
- **Pre-processing phase:** This phase includes tasks such as extra variables addition and computation, decision classes, assignments, data cleansing, completeness, correctness, attribute creation, attribute selection and discretization.

- **Applying rough set model for Analysis and Rule Generating Phase:** This phase includes the generation of preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, rule evaluation and prediction processes.
- **Classification and Prediction phase and validation tests:** This phase utilize the rules generated from the previous phase to predict the disease. Figure 1 illustrates a summary of the overall steps in the proposed Rough Set Prediction Model.

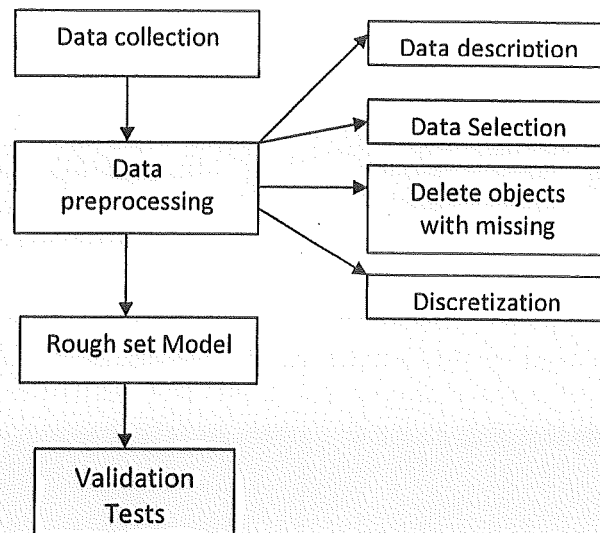


Figure 1: The process of prediction and analysis using Rough Set Model

3. Case Study

3.1 Data collection:

The data used in our research was obtained from the clinic of High Risk Unit at Kasr El-Aini Hospital, Cairo University from October 2011 to January 2012. So, the data covers an approximate three months period. It contains data for twenty six SLE pregnant females with ages between 18 to 38 years (average 28). Thus, the universe U includes 26 patients. We transformed original clinical free-text reports from clinical dataset into mainly four conditional attributes (ANA, ANT-DNA, ACL and LAC). All the patients were followed to delivery.

3.2 Preprocessing phase

Data description:

The data represents a multi valued information system referred to as data set1 or DS1 which contains a set of patients ($P_1, P_2, \dots, P_{26}$) besides, a set of attributes (ANA, ANT-DNA, ACL, LAC and SLE). The first four attributes are called "conditional attributes" and the last one (SLE) is called "decision attribute". The values of all these attributes are (0, 1) where (0) represents that this feature does not exist and (1) represents that this feature or attribute exists. Table 1 represents the frequency of different Immunological disorders in the patients with SLE in the study.

Table1: Immunological disorders in SLE pregnant patients

	Yes	No
ANA	23	0
Anti-DNA	16	7
ACL	8	15
LAC	20	3

Analysis of SLE patients: Autoantibodies were positive as follows: ANA in all 23 patients, Anti- DNA antibodies was positive in 16 out of 23 cases [69.6%]. Anticardiolipin antibodies (ACL) in 8 cases only [34.7%] and lupus anticoagulant (LAC) positive in 20 cases [86.9%].

Data splitting is also performed which created two randomly generated subsets, one subset for analysis containing 75% of those patients in data set and one validation containing the remainder 25% of those patients.

Serology in SLE

Since SLE is associated with a number of autoantibodies, it is important to understand their relevance in clinical practice. Some of these are useful as diagnostic markers, others help in quantifying disease activity and still others are primarily of research interest, making no contribution.

1. Antinuclear antibody (ANA):

ANA is a good screening test for SLE because 95% of cases show a high titre (1:80 or more) of this autoantibody. A negative test result makes the diagnosis highly improbable however ANA may be positive in other rheumatic disorders or chronic infections, and sometimes in normal individuals. Thus, the specificity of ANA for diagnosis of SLE is quite low (40% only).

2. Anti double stranded antibody (Anti DNA)

The positivity of anti-dsDNA in SLE at the time of presentation is in the range of 60% (although the cumulative positivity during the course of disease may approach 90%). Hence, anti-dsDNA can not be a good screening test for SLE. When positive, the test establishes the diagnosis of SLE. The anti-dsDNA titres most often correlate with disease activity.

3. Anticardiolipin antibody (ACL)

As many as 12% of patients may have only ACL antibody. It is recommended that both IgG and IgM ACL should be tested in a suspected case of APS.

4- Lupus anticoagulant (LAC)

A less sensitive test but it is assed with the clinical syndrome of thrombosis more often than is ACL. A patient with APS may test positive for either or both LAC and ACL. Both tests should be carried out when investigating APS[3] .

3.3 Applying rough sets model

After the preprocessing step, the rough sets model is applied as analysis tool to extract rules from decision tables to predict SLE as the autoimmune disease. In validation step, the rules are tested for their validity. The tested sample is randomly selected and classified only once, with its size being set arbitrarily to a constant percentage (25%) of all patients (objects) as we mentioned above.

This phase includes the generation of preliminary knowledge, such as computation of object reducts from data, rule evaluation and prediction processes. Thus, applying rough set software, the rule derived from re-ducs can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data.

3.4 Experimental Results:

After applying the reduction action, it generated 5 rules as shown in Table 2

Table 2: The rules generated from reduct

	Rules
1	ANA(1) AND ANT-DNA(1)AND ACL(1) AND LAC(1) →SLE(1)
2	ANA(1) AND ANT-DNA(1)AND ACL(0) AND LAC(1) →SLE(1)
3	ANA(1) AND ANT-DNA(0)AND ACL(0) AND LAC(1) →SLE(1)
4	ANA(1) AND ANT-DNA(0)AND ACL(0) AND LAC(0) →SLE(1) OR SLE(0)
5	ANA(0) AND ANT-DNA(0)AND ACL(0) AND LAC(0) →SLE(0)

The rules which are generated from re-duct using ROSETTA software contain the following knowledge:

Rule 1: ANA(1) AND ANT-DNA(1)AND ACL(1) AND LAC(1) →SLE(1)

Which means "IF ANA, ANT-DNA, ACL and LAC exist in analysis Then SLE exists".

Rule 2 ANA(1) AND ANT-DNA(1)AND ACL(0) AND LAC(1) →SLE(1)

Which means "IF ANA, ANT-DNA and LAC exist in analysis, Then SLE exists". So, it doesn't matter ACL exists or not to recognize SLE.

Rule 3 IF ANA(1) AND ANT-DNA(0)AND ACL(0) AND LAC(1) →SLE(1)

Which means "IF ANA and LAC exist in analysis, Then SLE exists". So, it doesn't matter ANT-DNA or ACL exist or not to recognize SLE.

Rule 4 IF ANA(1) AND ANT-DNA(0)AND ACL(0) AND LAC(0) →SLE(1) OR SLE(0)

Which means "IF ANA exists in analysis without any evidence of existing ANT-DNA, ACL and LAC, then we can say that SLE exists or not". So, finding ANA alone is not a proof that SLE exists.

Rule 5 ANA(0) AND ANT-DNA(0)AND ACL(0) AND LAC(0) →SLE(0)

Which means "IF ANA, ANT-DNA,ACL and LAC don't exist in analysis Then SLE does not exist".

Classification and Prediction phase and validation tests:

The results from Table 1 are compared with the actual observations, and it is seen that the rough set model has predicted the SLE accurately. Whereas, if the accuracy is higher, the classification is less ambiguous and the quality of rule is better, therefore, the classification is unambiguous and acceptable.

Validation set: When tested on validation set the accuracy was 0.91905

The experiment shows the ability of rough sets to select a reduced set of pattern features, providing better generalization.

3.5 Determining the relative importance of the variables and relation to correlation analysis

The importance of variables is a relative measure of how significant each of the input variables is in the predictive model.

The important variables are ANA, ANT-DNA, ACL and LAC as input variables related to SLE as output.

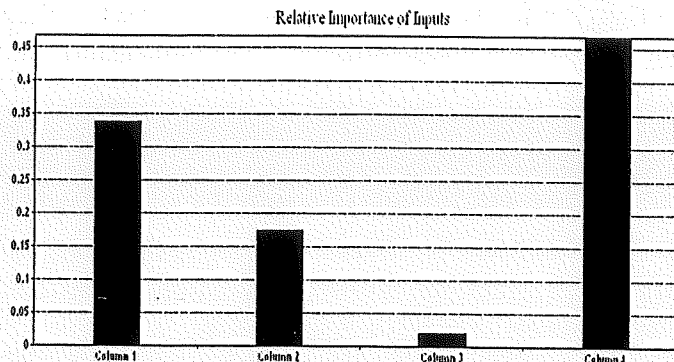


Figure 2: The relative importance of input variables

There are four variables as inputs and NeuroShell software is used to determine which variables are more important or have relative importance using very large series of experimental data that contains all parameters such as ANA, ANT-DNA, ACL and LAC. The results, as shown in figure (2) were as following:

ANA represents column 4 which is the most important variable to detect SLE with ratio =.468

LAC represents column 1 which has second level of importance to SLE with ratio = .338

ANT-DNA represents column 2 which has importance to SLE with ratio=.174

ACL represents column 3 which has importance level =.020

It is worth comparing these results with the correlation between SLE and the other variables as calculated by the contingency coefficient. Table 3 shows these correlations. It is evident from the table that the previous results from rough set analysis are homogenous with the correlations between SLE and other variables. The variable ACL which had the least importance has no significant correlation with SLE while the other 3 more important variables have significant correlation with SLE with the same order for importance and correlation magnitude.

Table 3: Contingency coefficients between SLE and other variables

Variable	Contingency Coefficient	p-value
ANA	0.554	<0.001
LAC	0.55	0.001
ANT-DNA	0.445	0.011
ACL	0.254	0.18

4. Conclusions

Predicting disease complications and estimation of its prognosis are two important constituents of medical management. In this paper, we used rough sets as a method to examine whether some patient groups have the SLE or not and which features or attributes could be used as an evidence on having SLE.

The suggested model proves the ability of the rough set approach to discover facts hidden in medical patients record, besides, the efficiency of the suggested rules which represent the general knowledge of any specialist doctor. The results are also homogenous with correlation analysis between SLE and the other study variables.

References

- [1] Bernsel, A. "Analysis of SNP microarray data using the rough set approach", master thesis, Department of Medical Sciences, Uppsala University, 2004, ISSN 1401-2138.
- [2] Cervera R, Font J, Carmona F, Balasch J. Pregnancy outcome in systemic lupus erythematosus: good news for the new millennium. *Autoimmun Rev.* 2002; 1:354-9.
- [3] Kumar, A. *J Indian Rheumatol Assoc.*, 10, 2002 ,80 – 96.
- [4] Ilczuk, G., Mlynarski, R., Wakulicz-Deja, A., Drzewiecka, A. and Kargul, W. "Rough Set Techniques for Medical Diagnosis Systems", *Computers in Cardiology, IEEE*, 2005;32, 837-840.
- [5] N. Karssemeijer, M. Thijssen, J. Hendriks, and E. L. Van Erning, *Digital Mammography: Computational Imaging and Vision*. The Netherlands: Kluwer, 1998.
- [6] Pawlak Z. *Rough Sets*, *Int. J. of information and computer science.*, V. 11, 1982, p. 341-356.
- [7] Pawlak, Z. *Rough Sets. Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers, 1991, The Netherlands.
- [8] Pawlak Z., Grzymala, B. J., Slowinski R. and Ziarko, W. *Rough sets. Communications of the ACM*, V. 38, No. 11, 1995, pp. 89-95
- [9] Petri M. Hopkins lupus pregnancy center: 1987-1996. *Rheum Dis Clin North Am* 1997; 23: 1-13.
- [10] Polkowski, L. *Rough Sets: Mathematical Foundations*. Physica-Verlag, 2003.
- [11] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis", *Artif. Intell. Med.*, vol. 18, no. 3, pp. 205-219, 2000.

- [12] R. Swiniarski, "Rough sets and bayesian methods applied to cancer detection," in Proc. 1st Int. Conf. Rough Sets Current Trends Comput., 1998, pp. 609–616
- [13] S. Das, A. Abraham, and S. Sarkar, "A hybrid rough set–particle swarm algorithm for image pixel classification," in Proc. 6th Int. Conf. Hybrid Intell. Syst., 2006, pp. 26–32.
- [14] S. Tsumoto, "Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model," Inform. Sci.: Int. J., vol. 162, no. 2, pp. 65–80, 2004.
- [15] S. Widz, D. 'Sle,zak, and K. Revett, "Application of rough set based dynamic parameter optimization to mri segmentation," in Proc. IEEE Annu. Meet. Inform. Process. NAFIPS, 2004, vol. 1, pp. 440–445.
- [16] Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982; 25: 1271–1277.
- [17] Urowitz MB, Gladman DD, Farewell VT, Stewart J, McDonald J. Lupus and pregnancy studies. Arthritis Rheum 1993; 36:1392–1397.
- [18] Y. Yao, "A comparative study of fuzzy sets and rough sets", J. Inform. Sci., V. 109, pp. 227–242, 1998.

THE UNIVERSITY OF CHICAGO

THE UNIVERSITY OF CHICAGO

THE UNIVERSITY OF CHICAGO

Robust Regression for handling Outliers using Mathematical Programming

***Hazem Refaat Ahmed**

Abstract:

We consider the problem of handling outliers in regression models. Many real data sets contain outliers. These outliers may constitute a threat for a number of classical models such as: Ordinary Least-Squares(OLS) model. Outliers are handled through a broad class estimation techniques, known in literature as: robust regression. The current research reviews some attempts for treating outliers in regression models , starting from the bounded-influence methods , high-break down methods and ending with methods of Quadratic Programming .Special interest is given to the latest methods with applications to simulated data.

Keywords: Bounded-Influence; High-breakdown ; Leverages; Outliers ; Penalty ; Robust Regression ; Weight functions; Penalty ; Quadratic Mixed Integer Programming.

Introduction:

In linear regression models data often contains outlying observations .The performance of linear regression model may be undermined by the existence of such outliers because they pull the regression line away from the line depicted by majority of the data , hence affecting the efficiency of estimates obtained from the regression analysis. Therefore, it is important to identify these observations in order to bound their influence or entirely delete them from the data set. Unfortunately, it is much more difficult to diagnose outliers when there are several of them and diagnostics for such multiple outliers may give rise to extensive computations. Also if these outliers are in groups or have a particular pattern, then we may face the problem of masking and swamping effects.

Outliers often rise from non-normal, heavy tailed distributions or contaminated distributions. In context of regression models, outliers can be classified as outliers in

*Department of Applied Statistic and Econometrics, Institute of Statistical Studies and Research,
Cairo University, Egypt.
E-Mail:hazemrefaatissr103@gmail.com

response variable (residual outliers) or outliers in predictor variables (leverage outliers). Both types of outliers pose a serious threat to the standard least-squares analysis and dealing with this problem is through robust regression methods.

This paper is organised as follows:

In section one; we review most of classical robust regression approaches. The mathematical programming and its role in robust regression is given in section 2. Section 3 will give simulation and Monte-Carlo experiments for the most recent approaches. Conclusion and recommendations are addressed in section 4.

1-Literature on Robust Regression:

Huber (1964) introduced the method of M-estimation as a generalization to maximum likelihood estimation which depends on minimizing the sum of some function of residuals. Huber M-estimator bounds the effect of outliers in the Y-space but it does not bound the effect of outliers in the X-space. To overcome the X-outliers deficiency of Huber M-estimator, other bounded-influence estimators, known as, Generalized M-estimators (GM) were introduced. Generalized M-estimators depend on using some weight function that bound the effect of X-outliers. Various estimators have been proposed such as : the method of resistant line(Tukey,1971), R-estimators(Jurecova,1971), L-estimators (Bickel, 1973), Mallows estimator (1975), Handschin et al. estimator (1975), Hill estimator (1977) and Welsch estimator (1980). One important disadvantage of these estimators is that they have low breakdown point that reaches at most $\frac{1}{p}$, where p is the number of parameters in the model. So, another broad class of estimators known as, the high-breakdown estimators, were introduced. These estimators achieve high-breakdown point up to 50%. The breakdown point of an estimator is the smallest fraction of anomalous or bad data that can render the estimator useless. For example, the sample mean has breakdown point of $\frac{1}{n}$ since a very extreme outlier can cause the sample mean to assume an implausible value.

Rousseeuw(1984) introduced the least median of squares(LMS) method which depends on minimizing the h^{th} ordered squared residual with n, p denoting the sample size and the number of parameters with $h = [n/2] + [(p+1)/2]$. Rousseeuw(1984) introduced The least trimmed squares(LTS) estimator which depends on minimizing the sum of the h smallest squared residuals with

$h = [n(1 - \alpha)] + 1$, for some selected trimming percentage α . Bassett (1991) and Hossjer(1991) introduced The least trimmed sum of absolute deviations (LTA) estimator which depends on minimizing the sum of the h^{th} smallest absolute residuals with h as defined in the LTS method. LTS is a famous and popular high-breakdown estimator that gives better efficiency than LMS and have less computational problems than LMS with large sample sizes.

One disadvantage of LTS, is that it can loose efficiency, since $(n - h)$ observations have to be removed from the sample even they are not outliers. Another problem with LTS is that the size $(n - h)$ of the outlier subset is rarely known.

Simpson et al.(1992) suggested a procedure that combines the high-breakdown and bounded-influence properties in the two-stage estimation method.

2- Mathematical Programming in Robust Regression:

First, we present the estimator that is proposed by Comariopoulous and Zioutas (2002) based on concepts of mathematical programming.

Consider an M-estimator of Huber-type (Huber and Dutter, 1974) obtained by:

$$\underset{B}{MIN} \sum_{i=1}^n \sigma^2 \rho\left(\frac{u_i}{\sigma}\right) \quad (2.1)$$

where the function ρ is given by:

$$\rho(u_i/\sigma) = \begin{cases} \frac{1}{2}(u_i/\sigma)^2 & \text{for } |u_i/\sigma| \leq k \\ k|u_i/\sigma| - k^2/2 & \text{for } |u_i/\sigma| > k \end{cases} \quad (2.2)$$

and σ is a scale parameter.

Different values of the tuning parameter k , $k = k_i$, leads to different bounded-influence estimators .For example, Huber M-estimator could be defined by choosing $k = 1.5$ or $k = 1.345$. Handschin et al . estimator (1975) could be defined by using

$$k_i = k \sqrt{1 - h_{ii}} .$$

Now, define ε_i as follows:

$$\varepsilon_i = \begin{cases} |u_i| - k_i\sigma & |u_i| > k_i\sigma \\ 0 & |u_i| < k_i\sigma \end{cases}$$

ε_i is interpreted as the distance of pulling y_i towards its fitted value, so, it is called the pulling distance.

The robust estimator could be defined by solving the equivalent problem:

$$\min \sum_{i=1}^n (u_i^{*2} + 2k_i\sigma\epsilon_i) \quad (2.3)$$

Where u_i^* are the modified residuals and also, they are called metrically winsorized residuals, defined by:

$$u_i^* = \begin{cases} u_i & |u_i| \leq k_i\sigma \\ k_i\sigma = u_i - \epsilon_i & |u_i| > k_i\sigma \\ k_i\sigma = u_i + \epsilon_i & |u_i| < -k_i\sigma \end{cases} \quad (2.4)$$

and $2k_i\sigma\epsilon_i$ can be considered as the penalty cost of pulling y_i towards its fitted value for a distance ϵ_i with $k_i = k w(x_i)$. An important interpretation of the problem (2.3) is that the estimator is defined by modifying the observations y_i with the smallest possible amount of overall pulling distances $E = \sum_i \epsilon_i$.

In general, in robust estimation, the bad influence of a point (x_i, y_i) is bounded by shortening the residuals, proportionally to the corresponding value of the tuning parameter k_i and the robust estimates will be the solution of the resulting system of equations:

$$\sum_i \psi(u_i) x_i = 0, \quad i = 1, \dots, n \quad (2.5)$$

The solution of (2.5) can be interpreted as a procedure where large residuals are modified by pulling the y -values towards the regression line for a distance ϵ_i . Then, we apply least-squares to the modified data obtaining a robust estimation of regression parameters.

Camarinopoulos and Zioutas (2002) considered the Mallows weights (Mallows, 1975) given by :

$$w(x_i) = \min \left\{ \left(\frac{b}{(x_i - m_x)' c_x^{-1} (x_i - m_x)} \right)^{\frac{\alpha'}{2}}, 1 \right\}, \quad (2.6)$$

$$0 \leq \alpha' \leq \infty$$

Where:

m_x and c_x are robust estimators of the multivariate mean and variance-covariance matrix, respectively, b is a cut-off robustness constant chosen from the Chi-square distribution and α' is another robustness constant to control and limit the bounds of the weight function. Simpson et.al(1992) used $\alpha' = 2$ and b as 0.90 percentile of the Chi-square distribution on $p-1$ degrees of freedom.

The robust estimation, in this case, can be obtained by solving the system of equations:

$$\sum_i \psi(\tilde{u}_i) \sqrt{w_i} x_i = 0 \quad (2.7)$$

where $\tilde{u}_i$ is the residual resulting from the weighted data $(\tilde{x}_i, \tilde{y}_i)$,

$$\tilde{u}_i = \tilde{y}_i - \tilde{x}_i B = u_i \sqrt{w_i}.$$

It should be noted that the pulling distances ε_i modify only the residuals by pulling y_i towards its fitted value and it is not associated with modifying the x -values.

Now, consider the robust estimation problem as an allocation problem using mathematical programming. The idea was based on constraining the total pulling distance. Then, the estimator selectively downweights bad leverage points by optimally allocating the pulling resource. As optimality criteria, the leverage measure h_{ii} and the residuals u_i were considered, where h_{ii} is diagonal element of the hat matrix $H = X(X^T X)^{-1} X^T$ (i.e: $h_{ii} = x_i^T (X^T X)^{-1} x_i$).

The estimator is defined by minimizing a Huber function and optimally allocating the total distance as follows:

$$\min_B \sum_{i=1}^n (u_i^2 + 2 k w_i \sigma \varepsilon_i)$$

Subject to:

$$\sum_{i=1}^n \varepsilon_i \leq E \quad (2.8)$$

In the above objective function, the penalty cost, $2 k w_i \sigma \varepsilon_i$, of pulling y_i towards its fitted value depends on weight $w_i = w(x_i)$, which is based on Mallows weight function.

For an outlying observation in the X -space, the corresponding penalty cost is very small and thus the y_i could be easily pulled towards the regression line, resulting in

excessive downweighting , so , this estimator can be regarded as a bounded–influence estimator.

Inferentially, the problem (2.8) is an optimization problem where the pulling resource E has to be allocated to the points, which correspond to large residuals u_i and small penalty cost $k w_i \sigma$. Usually, good leverage points do not correspond to large residuals and, so, in searching for an optimum solution of (2.8), these points are not being pulled.

The minimization problem (2.8) can be described in linear regression model as follows:

$$\underset{B_1, B_2, \varepsilon_i, u_i^*}{\text{MIN}} \quad \sum_{i=1}^n (u_i^{*2} + 2 k w_i \sigma \varepsilon_i)$$

Subject to:

$$x_i^T B_1 - x_i^T B_2 + u_i^* + \varepsilon_i \geq y_i$$

$$x_i^T B_1 - x_i^T B_2 - u_i^* - \varepsilon_i \leq y_i$$

$$\sum_{i=1}^n \varepsilon_i \leq E$$

(2.9)

Where the decision variables are:

$$B_1^T = (B_{11}, \dots, B_{1p}), B_2^T = (B_{21}, \dots, B_{2p}), \varepsilon_i \text{ and the winsorized residual } u_i^* .$$

The constraint $\sum \varepsilon_i \leq E$ enters the model in order to limit the pulling resource.

The quadratic programming problem (2.9) has a convex objective function and the constraints form a closed convex set (Zioutas et al., 1997). Therefore, during the minimization process is implemented, it searches the feasible solutions on the extreme points or comers of the convex region. The iterative method stops at the extreme point (B_1, B_2, u^*, E) which is the required global minimum .

Zioutas and Avramidis (2005) , proposed an estimation method , using the previous formulation of quadratic mixed integer programming , in order to delete, not downweighting , the bad leverage points .

. In this approach, a penalty cost for deleting or discarding outlier is introduced.

The estimation problem is illustrated as follows:

Consider the estimator that is defined as the solution to the problem:

$$\text{Min } \sum_{i=1}^n \rho(u_i^2) ,$$

Where ρ is a function of residuals, defined as follows :

$$\rho(u_i^2) = \begin{cases} u_i^2 & |u_i| \leq k_i \sigma \\ (k_i \sigma)^2 & |u_i| > k_i \sigma \end{cases} \quad (2.10)$$

where σ is a robust residual scale , and $k_i = kw'(x_i)$ is a well known cutoff for the outliers that depends on the design weights . Zioutas and Avramidis (2005) used Mallows weights $w'(x_i) = w(x_i)$. The interpretation for the proposed function of residuals in (2.10) is simple for large residuals u_i (i.e : $|u_i| \geq k_i \sigma$) the squared residuals are less rapidly increasing. Since the squared residual is reduced to the value $(k_i \sigma)^2$, which depends only on the design points and the initial scale estimate σ and this value, $(k_i \sigma)^2$, can be interpreted as a penalty cost for deleting an outlier . The computation of this estimator is implemented through quadratic mixed integer programming as follows:

$$\text{MIN}_{B_1, B_2, \varepsilon_i, u_i, \delta_i} \sum_{i=1}^n (u_i^2 + \delta_i (k_i \sigma)^2)$$

Subject to:

$$\begin{aligned} x_i^T B_1 - x_i^T B_2 + u_i &\geq y_i - \varepsilon_i \\ x_i^T B_1 - x_i^T B_2 - u_i &\leq y_i + \varepsilon_i \\ \varepsilon_i &\leq \delta_i M^* , \end{aligned} \quad (2.11)$$

$$\sum_{i=1}^n \delta_i \leq n - (n + p + 1) / 2 ,$$

δ_i : zero-one variable

$$B_1, B_2, u_i, \varepsilon_i \geq 0$$

Where M^* is an upper limit of the residuals u_i . The decision variables are:

Unknown regression coefficients, $B_1^T = (B_{11}, \dots, B_{1p})$, $B_2^T = (B_{21}, \dots, B_{2p})$,

the residuals u_i , the pulling distances ε_i and δ which is a zero-one decision vector , to indicate which observations must be removed .

-The first two constraints describe geometrically the regression line for the cases of positive or negative residuals, since in the mathematical programming formula, the decision variables are constrained to positive values.

-The third constraint put a restriction on the pulling distance ε_i according to the value of δ_i such that :

* For $\delta_i = 1$:

The third constraint allows the pulling distance to take any upper limit M on the maximum absolute residual of the regression.

* For $\delta_i = 0$:

The third constraint restricts the pulling distance ε_i to zero. So, there is no pulling of y_i towards its fitted value.

-The fourth constraint indicates that the number of deleted points cannot exceed $(n - (n + p + 1) / 2)$ and this is the number of discarded points in the LTS estimator .

The previous mixed quadratic programming formula is convex and therefore a global optimum solution can be obtained.

A more recent method in robust estimation using mathematical programming is given by Zioutas et al.(2007) which is basically based on the LTS estimator as the pervious estimator but a slight modification was made to penalty cost for deleting outliers which is based on a robust leverage measure and robust measure of scale. This new estimation method is based on the idea of the LTS estimator, proposed by Rousseeuw and Leroy (1987). The basic idea, here, is to define a subset of observations that produces significant or large reduction in the residual sum of squares and this set is most likely to be outliers. Penalty costs are inserted into the objective function in order to delete only those observations, which cause the significant reduction, where the reduction of residual sum of squares (SSE) due to the deletion of an observation i is given by:

$$SSE - SSE_{(i)} = \frac{u_i^2}{1 - h_{ii}} \quad (2.12)$$

where : SSE is residual sum of squares , $SSE_{(i)}$ is the residual sum of squares computed after the deletion of an observation i , h_{ii} is the leverage of the corresponding observation which lies on the diagonal elements of the hat matrix H

and u_i is the corresponding residual. We compare the reduction that happens after deleting observation i to its penalty cost such that if:

$$\frac{u_i^2}{1 - h_{ii}} > (k\sigma)^2 \quad (2.13)$$

then, we delete this observation. So, we seek to find a subset of size q (possibly outliers) which its deletion gives the largest reduction in residual sum of squares, hence, the remaining set will be the clean one. But the problem is that the subset q of outliers in most data sets, are rarely known. This estimator handle this problem by not requiring the knowledge of the q subset of outliers, instead, some quantity (penalty cost) is inserted into the function of the residuals. So, observations that produces reduction larger than their penalty costs are removed. The estimation problem is formulated as follows:

$$\text{MIN}_B \left[S_n(B) = S_{h^*}(B) + S_{n-h^*}(B) = \sum_{i=1}^{h^*} (u_i^2) + \sum_{i=h^*+1}^n ((k_i\sigma)^2) \right] \quad (2.14)$$

So, the estimator minimizes the total sum, the sum, the sum of the h^* squared residuals (where h^* is not known priori) in the clean set and the sum of penalties for deleting the rest ($q = n - h^*$) observations.

Here, $(k_i\sigma)^2$ is defined a priory, where σ is a robust measure of scale, taken from the LTS method, $k_i = k w_i$ where k is the tuning parameter, w_i is a suitable robust weight function to be inserted into the penalty cost. Now the Penalized Trimmed Sum of Squares(PTS) estimator can be defined as:

$$\text{Min} \sum_{i=1}^n \rho(u_i) \quad (2.15)$$

where $\rho(u_i)$ is the function of the residuals, given by

$$\rho(u_i) = \begin{cases} u_i^2 & |u_i| < k_i\sigma\sqrt{1-h_{ii}} \\ (k_i\sigma)^2 & |u_i| > k_i\sigma\sqrt{1-h_{ii}} \end{cases}$$

Choice of the penalty cost:

If it is assumed the penalty cost in the residual function is constant for each point (i.e:

$(k_i\sigma)^2 = (k\sigma)^2$ then, when solving (2.14), we have four cases:

- Points that are low -leverage outliers (i.e: that correspond to small h_{ii} and large u_i) will be deleted due to their large residual .

- Points that are bad high –leverage outliers (i.e: that correspond to large value of h_{ii} and large u_i) will also be deleted.
- Points that are high good leverage (i.e: that correspond to small residuals, high value of h_{ii}) will also, be deleted.
- Points that have small leverage, small u_i represents the clean data set.

The choice of the penalty cost is very important in controlling the robustness and efficiency of the estimator, because case 1 and case 2, represents the case of good data points, however, this points are deleted. The third case represents the case of bad data points which must be deleted. Using fixed penalty, both good data points and bad data points are handled by deletion. Moreover, The problem in using the diagnostic

measure $\frac{u_i^2}{1 - h_{ii}}$. h_{ii} itself is not a robust measure of leverage and its value is

distorted by clustering. So, it is required to find a robust measure of h_{ii} that is not affected by clustering and takes into account the masking effect.

Note: There are two effects of outliers that may happen in simulation and affect efficiency and mean-square error as

-Masking effect happens when the observation is really outlier but it is falsely identified as non-outlier.

-Swamping effect happens when the observation is really not outlier but it is falsely identified as outlier.

Zioutas et al. (2007) suggested using the following weight function in the penalty cost of the residual function:

$$w_i^{**} = \frac{\sqrt{1 - h_{ii}^*}}{\sqrt{1 - h_{ii}}} \quad (2.16)$$

Where: h_{ii} is the leverage of a point based on the whole data set. h_{ii}^* is the leverage of the same observation as it joins a subset of h^* clean observations: $h_{ii}^* = x_i^T (X^T X)^{-1} x_i$, where X is the $(h^* + 1) \times p$ matrix that contains the h^* clean data and the joint potential outlier i .

The clean subset is found through Minimum Covariance Determinant (MCD) method (Rousseeuw and Van Drissen (1999)). The procedure is explained as follows:

Step1: Draw a random subset J of size p , p is the number of parameters in the model.

Step2: Compute $T_0 = \text{average}(J)$, $S_0 = \text{Cov}(J)$.

Step3: Compute the Mahahlanbois distances:

$$d_i^2 = (x_i - T_0)^T S_0^{-1} (x_i - T_0) \quad , \quad i = 1, 2, \dots, n$$

for all observations, and then sort these distances and take the h^{th} smallest distances where $h = [n/2] + [(p+1)/2]$.

Step4: Repeat the last three steps for the h^* new subset and iterate until one of the following conditions is satisfied.

$\text{Det}(S_L) = 0$ or $\text{det}(S_L) = \text{det}(S_{L-1})$, where S denotes covariance matrix and L denotes the number of iterations and the final subset that gives the lowest value of $\text{det}(S)$ among all possible sub samples of size p , would be the subset of the clean data.

Weights are applied by following the previous MCD procedure.

The problem is formulated using quadratic mixed integer programming as follows:

$$\begin{aligned} & \text{Min}_{B, \varepsilon_i, u_i, \delta_i} \sum_{i=1}^n (u_i^2 + \delta_i (k_i \sigma)^2) \\ & \text{Subject to:} \\ & x_i^T B_1 - x_i^T B_2 + u_i \geq y_i - \varepsilon_i \\ & x_i^T B_1 - x_i^T B_2 - u_i \leq y_i + \varepsilon_i \\ & \varepsilon_i \leq \delta_i M \\ & \delta_i \in \{0, 1\} \\ & B, u_i, \varepsilon_i \geq 0 \quad i = 1, 2, \dots, n \end{aligned} \quad (2.17)$$

The interpretation for this formula is as discussed before in the previous estimation method (Zioutas and Avramidis , 2005)

This problem has linear constraints and a convex quadratic objective function. Therefore, we have a unique global optimum solution.

It should be noted that, the only difference in the formulation introduced by Zioutas et al.(2007) , is in bounds in the PTS function which becomes

$$k_i \sigma \sqrt{1 - h_{ii}} = k w_i^{**} \sigma \sqrt{1 - h_{ii}} = k \sigma \frac{\sqrt{1 - h_{ii}^*}}{\sqrt{1 - h_{ii}}} \sqrt{1 - h_{ii}} = k \sigma \sqrt{1 - h_{ii}^*} \quad (2.18)$$

Where: $k_i \sigma$ is the penalty cost for deleting an outlier, σ is a robust measure of scale, $k_i = k w_i^{**}$ with w_i^{**} as the robust weight function and $k = 3$. It should be noted also, that the PTS estimator converges by the LTS, by decreasing the penalty cost.

Properties of the PTS estimator:

1-bounded-influence :

Because the penalty cost is downweighted for high-leverage points, their influence is bounded, and the masking problem is eliminated.

2- Efficiency:

Because this estimator discards only catastrophic observations, since there is penalty $(k, \sigma)^2$ for each deletion. High – leverage points is not deleted if its residual is smaller than (k, σ) . Besides, it gives the best model fit to the clean data points .

3- High- breakdown point:

The PTS estimator achieves this property because:

- It takes high-breakdown estimate for residual scale σ taken from LTS.
- It takes design weights from the high-breakdown MCD and robust measure of leverage.

Combinations of bounded-influence and high-breakdown point has been sought through two-stage estimation method(Refaat,2010)

Of these two-stage combinations, it has been shown –through a simulation study– that the two-stage estimation method which is based on PTS estimator in the first method and a bounded-influence estimator in the second estimator is the best and most efficient robust estimator against outliers or contamination in data(Refaat, 2010).

Ziotas et al.(2007) proposed a modification of the PTS estimator and introduced γ – Insistive PTS(IPTS) estimator. The idea is based on defining γ – Insistive loss function from Support Vector Machine (Vapnick,1998) . The support vectors machine is a mathematical Programming-based technique, which has the advantage of reducing the computational difficulty by ignoring small residuals. Specifically residuals within an interval $(-\gamma, \gamma)$ are not included in the loss function and those points outside the so-called γ -tube which are called support vectors, define the regression line. The new PTS estimator has the ability to handle group of outliers and suffers a little from masking effects. .

In the γ – Insistive loss function approach based on support vector machine, small errors are not penalized and it is attempted to fit a tube with radius γ to the data, by ignoring (tolerating) small errors. So , we have the following loss function:

$$|y - f(x)|_{\varepsilon} = |y - X^T B|_{\varepsilon} = \max(0, |y - X^T B| - \gamma)$$

Where The accuracy parameter γ controls the number of points outside the tube with radius γ .

Zioutas et al.(2007) adapted the technique by modifying the γ -insensitive loss function in a squared form, and all the errors smaller than γ are penalized with a constant value γ^2 . Thus, the proposed γ -insensitive loss function becomes:

$$((y - f(x))^2_\epsilon = (y - X^T B)^2_\epsilon = \max(\gamma^2, (y - X^T B)^2)$$

where, the accuracy parameter γ controls the number of points outside the tube. The new estimator (IPTS) is defined by solving:

$$\begin{aligned} & \underset{B}{\text{MIN}} \sum_{i=1}^n \rho_{\gamma, k, \sigma}(u_i) , \\ & \rho_{\epsilon, k, \sigma}(u_i) = \begin{cases} \gamma^2 & |u_i| \leq \gamma \\ u_i^2 & \gamma < |u_i| < k_i \sigma \sqrt{1 - h_{ii}} \\ (k_i \sigma)^2 & |u_i| \geq k_i \sigma \sqrt{1 - h_{ii}} \end{cases} \end{aligned} \quad (2.19)$$

Where : $k_i \sigma = \max(\gamma, k_i \sigma)$

The minimization of the loss function (2.19) is equivalent to the following constraint optimization problem:

$$\begin{aligned} & \underset{B, s_i, u_i, \delta_i}{\text{Min}} \sum_{i=1}^n (u_i^2 + \delta_i (k_i \sigma)^2) \\ & \text{Subject to:} \\ & x_i^T B_1 - x_i^T B_2 + u_i \geq y_i - s_i \\ & x_i^T B_1 - x_i^T B_2 - u_i \leq y_i + s_i \\ & u_i \geq \gamma \\ & s_i \leq \delta_i M \\ & \delta_i \in \{0, 1\} \\ & u_i, s_i \geq 0 \quad i = 1, 2, \dots, n \end{aligned} \quad (2.20)$$

Where:

$k_i \sigma = \max\{\gamma, k_i \sigma\}$, δ is the zero-one decision vector and M is an upper limit of the residuals u_i .

The IPTS formula is convex, therefore a unique optimum solution can be found and the IPTS is estimated. A final note remarked regarding the above formula is that all

points inside the γ -tube do not contribute to the solution. So, we could remove any one of them, and still obtain the same solution.

Avramidis and Zioutas(2011) presented new class of regression estimates called generalized PTS (GPTS). The new GPTS estimator is defined as the PTS but with penalties suitable for bounding the influence function of all observations.

Consider The PTS estimator defined in (2.15) which is given by solving:

$$\sum \psi(u_i) = 0, i=1,2,\dots,n$$

Where:

$$\psi(u_i) = \begin{cases} u_i & |u_i| < k_i \sigma \sqrt{1-h_{ii}^*} \\ 0 & \text{otherwise} \end{cases}$$

Based on the formula of the PTS estimator, Avramidis and Zioutas(2011) defined a new efficient PTS estimator, called GPTS estimator, which is bounded influence and is given by the solution of the system:

$$\sum \psi(u_i) = 0, \text{ where:}$$

$$\psi(u_i) = \begin{cases} u_i & \frac{|u_i|}{\sqrt{1-h_{ii}}} \leq \frac{k\sqrt{p}\sqrt{1-h_{ii}^*}\sigma}{\sqrt{nh_{ii}}} \\ 0 & \text{otherwise} \end{cases}$$

Also, they found that $2.5 < k < 3$ works well for the PTS or GPTS estimator. Based on robust scale $\hat{\sigma}$ and leverage h_{ii}^* , the GPTS estimator achieves simultaneously the three goals of robust regression:

(a) high breakdown point, (b) bounded influence, (c) high efficiency (say 0.95) vs. OLS when the data follow Gaussian distribution.

In order to minimize the penalty loss function in a robust regression, Zioutas and Avramidis used the same QMIP formula in (2.20).

3- Monte-Carlo Simulations:

First: Monte-Carlo experiments were performed to compare the performance of LTS, PTS and IPTS:

The observations y_i were obtained following the regression model of two variables:

$y_i = B_0 + B_1 x_{1i} + B_2 x_{2i} + e_i$, where the coefficient values are:

$\beta_1 = 1.20, \beta_2 = -0.80$ and a zero constant term $\beta_0 = 0.0$. The error term has Gaussian distribution $N(0,16)$, x_{1i} and x_{2i} are values drawn also from normal distributions $N(20,36)$ and $N(30,64)$ respectively. Outliers are obtained by drawing extra values from the uniform distribution $U(80,220)$ and added to x_{1i} , x_{2i} or y_i .

The results of the OLS estimator and the three robust estimators :

LTS($k = 0.6n$) , PTS($k = 3$) and (IPTS($k = 3$ and various values of ε)) are reported . The performance criteria used for comparing the estimators are mean-square error fitting and relative efficiency with respect to OLS which is computed as

follows: Relative efficiency= $\frac{\text{mean square error fitting (OLS)}}{\text{mean square error fitting (Robust)}}$

The results were as follows :

Table(1)
(50 points in R^2 including 12, outliers (24% contamination))
Results of 500 simulation runs

Estimation procedure	Mean Square Error Fitting	Relative efficiency
True function	253.86	8.362
OLS	2122.79	
LTS	317.15	6.693
PTS	271.23	7.827
$IPTS_{\varepsilon=0.8\sigma}$	263.41	8.059

Table(2)
(50 points in R^2 including 16, outliers (32% contamination))
Results of 500 simulation runs

Estimation procedure	Mean Square Error Fitting	Relative efficiency
True function	253.86	10.036
OLS	2547.86	
LTS	322.87	7.891
PTS	282.23	9.028
$IPTS_{\varepsilon=0.8\sigma}$	274.16	9.293

Tables 1–2 present the measures of the performance criteria for the four estimators in the presence of average and high contamination respectively. Taking into account all performance criteria we see that the IPTS estimator outperforms the other estimators.

In Table 1, the LTS estimator has a reasonable performance but the PTS and IPTS estimators improve significantly the fitting error and the relative efficiency. The IPTS estimator is the best among all others concerning all criteria in Table 2, where the data is heavily contaminated with 16 outliers, the IPTS still the best among all other robust estimators.

Table(3)
(75 points in R^2 including 10, outliers (13% contamination))
Results of 500 simulation runs

Estimation procedure	Mean Square Error Fitting
LTS	0.47
PTS	0.30
$IPTS_{\epsilon=0.8\sigma}$	0.27
$IPTS_{\epsilon=1.1\sigma}$	0.24
$IPTS_{\epsilon=1.3\sigma}$	0.23

For large data sets we could increase the radius γ , in order to delete the outliers from the original data set. We can empirically conclude from the results that the γ -insensitive PTS procedure provides a significant improvement with respect to the PTS, in terms of mean-square error fitting.

Second: Monte-Carlo experiments were performed to compare the performance of LTS, PTS and GPTS:

The model is : $y = XB + e$ (3.1)

where y is the $n \times 1$ vector of the response variable $y = (y_1, y_2, \dots, y_n)^T$, X is a full rank $n \times p$ matrix of explanatory variables, $x_i = (x_{i,1}, \dots, x_{i,p})$, for $i = 1, 2, \dots, n$, b is a $p \times 1$ vector of unknown parameters $B = (b_1, b_2, \dots, b_p)^T$ and u is a $n \times 1$ vector $e = (e_1, \dots, e_n)^T$ of iid random errors with expectation zero and variance σ^2 . a proportion $(1 - \lambda)$ in the data samples $(x_1, x_2, \dots, x_p, y)$ follow a multivariate normal distribution, where λ is the contamination fraction. Without loss of generality the mean vector is taken equal to 0 and the identity as the covariance matrix. These observations follow the model in (3.1) with $b = 0$. The contaminated observations are high leverage outliers with $x = (1, 100, 0, \dots, 0)$ and $y = \text{slope} \times 100$, where the slope of contamination takes values 1 or 2.2. The mean square error fitting(MSE) is used as a measure of performance.

Table(4)

Samples with 10% outliers located at $(x_1, x_2, \dots, x_p, y) = (1, 100, 0, \dots, 0, 100)$

n	P	LTS (MSE)	PTS (MSE)	GPTS (MSE)
100	2	0.692	0.030	0.025
	3	0.890	0.052	0.052
	5	1.193	0.099	0.099
500	5	0.803	0.014	0.041
	10	0.935	0.029	0.039
	20	1.207	0.069	0.055
1000	5	0.715	0.007	0.012
	10	0.859	0.014	0.014
	20	1.158	0.028	0.028

For heavier contamination $\lambda = 0.10$ and slope 2.2, Table 4 shows the results of these simulations. We observe that the GPTS and PTS have the best overall performance especially for the small data sets $n = 100$. However, for larger data sets ($n \geq 500$), the GPTS, PTS perform comparable, since the outliers are easier to detect by both robust estimators due to the large slope value.

To continue the implementation of the GPTS algorithm and compare it with, LTS and PTS, one extra Monte Carlo experiments is performed to evaluate the performance of our robust procedure.

The observations y_i , were obtained following the regression model as in (3.1) with $p = 3$ and $n = 50$, where the coefficient values are $b_1 = 1.20$, $b_2 = 0.80$ and a zero constant term $b_0 = 0.0$. The error term e is $N(0, 162)$, while x_1 and x_2 are iid values drawn also from normal distributions $N(20, 62)$ and $N(30, 82)$ respectively. Different kinds of outliers are generated as follows:

An extra value is drawn from the uniform distribution $U(80, 220)$ and for the regression outlier is added to x_1 or x_2 , for the "good" leverage point is added to x_1 or x_2 but the value of the dependent variable y follows their contamination, according to the above regression model, for the response outlier is added to y . All simulation results are based on 1500 replications enough to obtain a relative error $< 10\%$.

Mean absolute error of estimates and mean-square error are measures of performance.

Table(5): Samples with 32% outliers
(X-outliers=10, Y-outliers=6,n=50,p=3)
 $B_0 = 0.0, B_1 = 1.2, B_2 = -0.8$

Estimation procedure	LTS	PTS	GPTS
Mean Absolute Error of $\hat{B}_0$	11.17	7.78	7.77
Mean Absolute Error of $\hat{B}_1$	0.789	0.256	0.212
Mean Absolute Error of $\hat{B}_2$	0.227	0.208	0.202
Mean Square Error	476.33	309.32	296.06

In case of only bad high leverage contamination, shown in Table 5, the Penalized Trimmed Squares approach has shown remarkable improvement in both robustness and efficiency, with GPTS the best.

4- Conclusions and recommendations:

Two new methodologies for deleting outliers in robust regression (IPTS and GPTS) are presented which reduces the problem to a Quadratic Mixed Integer Programming. The two approaches are shown to perform well when compared to the original LTS and its modified version PTS. Further research may be needed to handle the problem of outliers for large size sample data ($n > 1000$) possibly by solving the dual problem of (2.24). Moreover, modifications of (2.24) should be investigated that will provide the optimum value of γ . Also, a combination of bounded-influence estimator and IPTS and GPTS in a two-stage estimation procedure should be considered.

ACKNOWLEDGEMENTS:

The author is indebted to Professor Dr.Amany Mousa and Dr.Ahmed El-Sheikh , Department of Applied Statistic and Econometrics, Institute of Statistical Studies and Research, Cairo University, whose helpful comments and suggestions led to a much improved presentation.

References:

- 1-Avramidis,A. and Zioutas,G.(2011)," Locally and globally robust Penalized Trimmed Squares regression", *Simulation Modelling Practice and Theory* ,19,148–160.
- 2-Bassett,G.W.(1991), "Equivariant,Monotonic, 50%break-down estimators," *The American Statistician*, 45,135-137.
- 3- Bickel, P.J.(1973) , "on some analogues to linear combination of order statistics in the linear model ", *Ann. Stat.*,1,597-616.
- 4-Brikes,D. and Dodge ,Y. (1993), " *Alternative Methods of Regression*", Wiley Series in Probabilities and Statistics.
- 5- Camarinopoulos,L. and Zioutas, G. and,(2002), " Formulating robust regression estimation as an optimum allocation problem" *Journal of Statistical Computation and Simulation*,9,687-705.
- 6- Giltinan ,R.J , Carroll, R.J and Ruppert, D.(1986), "Some new estimation methods for weighted regression when there are possible outliers", *Technometrics*, 28, 219-230.
- 7- Handschin,E., Kohlos,J.,Fiechter, A., and Schweppe, F.(1975), "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*,2,329-337.
- 8- Hill, R.W.(1977), "*Robust Regression when there are outliers in the carriers*", PhD Dissertation, Harvard University, Boston, MA.
- 9-Hossjer,O.(1991), "*Rank-based estimates in the linear model with high-breakdown point*", PH.D. Dissertation, Uppsala University, Uppsala, Sweden.
- 10-Huber,P.J.(1964),"Robust estimation of a location parameter,"*Ann. Math. Stat.*,35,73-101.
- 11-Huber,P.J. and Dutter,R.(1974), "*Numeical. Solutions of Robust Regression Problem, in: COMPSTAT 1974*", Physika Verlag, Vienna.
- 12-Jurecova,J.(1971),"Nonparametric estimates of regression coefficients", *Ann. Math. Stat.*,42,1328-1338.
- 13- Mallows, C.L.(1975), "On some topics in robustness", *Memorandum Bell Telephone Laboratories*, Murray Hill, NJ.
- 14- Refaat,H.(2010)," *Handling Outliers using High-Breakdown and Bounded-Influence Robust Estimation in Linear Regression Models*",Msc. Thesis,Cairo university,Egypt.

- 15-Rousseeuw,P.J.(1984), "Least median of squares," *Journal of the American Statistical Association*,79,871-880.
- 16-Rousseeuw,P.J. and Leroy, A. (1987), "*Robust Regression and Outlier Detection*", Wiley, New York.
- 17- Rousseeuw,P.J. and Van Drissen ,K. (1999), " A fast algorithm for the minimum covariance determinant estimator ," *Technometrics* ,41,212-223.
- 18-Simpson, D.G., Ruppert,D. and Carroll, R.J.(1992), "On one-step GM-estimates and stability of inference in linear regression,"*Journal of the American Statistical Association*, 87,439-450.
- 19-Tukey,J.W.(1971),"*Explanatory Data Analysis*",Addison-Wesely,MA.
- 20-Vapnick, V. N. (1998), "*Statistical Learning Theory*", Wiley, New York.
- 21-Welsch,R.E,(1980), "Regression sensitivity analysis and bounded-influence estimation" *paper presented at NBER Conference on Criteria for Evaluation of Econometric Models*, Academic press, University of Michigan, New York.
- 22--Zioutas ,G. and Avramidis,A.(2005), "Deleting Outliers in Robust Regression with Mixed integer programming ", *Acta Mathematicae applicataesinica,English Series* ,2, 323-334.
- 23--Zioutas,G.,Camarinopoulos,L. and Bora,E.(1997), "Robust autoregressive estimates using quadratic programming ",*European Journal of Operational Research* , 101, 486-498.
- 24-Zioutas ,G. and Avramidis,A. and L.Pitsoulis(2007), "A Penalized trimmed squares and a modification of support vectors for unmasking outliers in linear regression", *REVSTAT-Statistical Journal*,5,115-136.

Proposed SUR-ridge Estimators with Application

M. A. Abdel-Fattah¹

Abstract The ridge regression (RR) as a shrinkage technique handles the multicollinearity problem in regression analysis. This article reviewed and proposed some seemingly unrelated regression ridge (SUR-ridge) estimators based on Zeebari et al. (2012) and Abdel-Fattah (2013). Following Abdel-Fattah (2013), the current article (i) estimates the SUR error covariance matrix using the ordinary ridge (OR) estimator instead of the ordinary least squares (OLS) and (ii) estimates the error covariance matrix using the restricted residuals instead of the unrestricted. A system of two regression equations based on the data due to Theil (1971) has been considered. Based on the results, more precise estimates can be obtained.

Keywords Multicollinearity, Ridge regression, SUR-ridge parameters, Restricted residuals.

1. Introduction

Over five decades have passed since Zellner (1962) introduced the SUR model for the first time as one of his successful scientific contributions. Zellner (1962) has found out that the yielded generalized least squares (GLS) estimator is at least asymptotically more efficient than the OLS estimator in the following cases; (i) when error terms in different equations are highly correlated and (ii) when exogenous variables in different equations are not highly correlated.

The highly intercorrelated exogenous variables (multicollinearity) are a data problem that remains and cannot be removed, but we only can handle its misleading results. RR introduces bias by adding a quantity to the diagonal of $X'X$ or $X'\hat{V}^{-1}X$ but reduces the variances of the estimates. Hoerl and Kennard (1970a,b) contribution to select this quantity was followed by numerous research studies both in the univariate case (multiple regression) and in the multivariate case (SUR model). To handle the problem, Alkhamisi and Shukur (2008) and Zeebari et al. (2012) are among the researchers who depended on the RR technique in the context of SUR models. Alkhamisi and Shukur (2008) proposed SUR-ridge parameters and compared them through a simulation study with other existent parameters which are generalizations for the multiple regression model. Zeebari et al. (2012) modified the parameters which were presented in Alkhamisi and Shukur (2008) and compared them through a simulation study. Alkhamisi and Shukur (2008) and Zeebari et al. (2012) approach includes the covariance structure among equations, but they estimated the error covariance matrix, Σ_M , using the unrestricted residuals obtained from the OLS method and not the obtained from a method that handles the problem (e.g. RR, principal components regression, partial least squares regression or latent roots regression).

¹ Department of Applied Statistics and Econometrics,
Institute of Statistical Studies and Research, Cairo University, Egypt
e-mail: mahmoud.ahmed@rocketmail.com, mahmoud.ahmed@cu.edu.eg

2. Model and Estimators

Consider the usual system of M seemingly unrelated regressions as;

$$Y = X\beta + u, \quad (1)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix}_{MT \times 1} = \begin{bmatrix} X_1 & 0_{T \times P_2} & \cdots & \cdots & 0_{T \times P_M} \\ 0_{T \times P_1} & X_2 & & & 0_{T \times P_M} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ 0_{T \times P_1} & 0_{T \times P_2} & \cdots & \cdots & X_M \end{bmatrix}_{MT \times P} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix}_{P \times 1} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}_{MT \times 1}, \quad (2)$$

The i^{th} equation of the M equation regression system is;

$$Y_i = X_i \beta_i + u_i, \quad i = 1, 2, \dots, M \quad (3)$$

where; Y_i is a $T \times 1$ vector of observations on the i^{th} endogenous variable, X_i is a $T \times P_i$ matrix of observations on P_i exogenous variables, β_i is a $P_i \times 1$ vector of unknown regression coefficients and u_i is a $T \times 1$ vector of contemporaneous correlated random errors.

The commonly used SUR model assumes the following;

- X_i is fixed with full column rank P_i .
- $\lim_{T \rightarrow \infty} \frac{1}{T} (X_i' X_i) = Q_{ii}$ is non-singular with finite and fixed elements.
- $\lim_{T \rightarrow \infty} \frac{1}{T} (X_i' X_j) = Q_{ij}$ is non-singular with finite and fixed elements.
- $E(u_i' X_j) = 0$, $u_i \sim \text{IID} (0, \sigma_{ii} I_T)$, $i = 1, 2, \dots, M$ and $\text{COV}(u_i, u_{i'}) = 0$.
- The $MT \times 1$ vector of random error terms in Eq. 2 is assumed to have the following variance-covariance matrix as;

$$E(uu')_{MT \times MT} = V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & & & \sigma_{2M} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \cdots & \sigma_{MM} \end{bmatrix} \otimes I_T = \Sigma_M \otimes I_T, \quad (4)$$

where; $\sigma_{ii'} = E(u_{it} u_{it'})$ for $t = 1, 2, \dots, T$ and $i, i' = 1, 2, \dots, M$, Σ_M is an $M \times M$ positive definite symmetric matrix and $\otimes$ represents the kronecker product operator.

The OLS estimator, $\hat{\beta}_{OLS}$, of the parameters is given by;

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y, \quad (5)$$

But from assumption v, it is clear that the errors across equations are correlated, taking this correlation into account leads to efficient estimates, namely GLS which is called SUR (Zellner, 1962) and is given by;

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}Y, \quad (6)$$

$$\text{COV}(\hat{\beta}_{GLS}) = (X'V^{-1}X)^{-1}, V^{-1} = \Sigma_M^{-1} \otimes I_T, \quad (7)$$

In many applications, the Σ_M matrix may be unknown. Thus, Σ_M elements must be estimated by replacing the unknowns by their estimators, then the feasible GLS estimator can be obtained by;

$$\hat{\beta}_{FGLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y, \quad (8)$$

$$\text{COV}(\hat{\beta}_{FGLS}) = (X'\hat{V}^{-1}X)^{-1}, \hat{V}^{-1} = \hat{\Sigma}_M^{-1} \otimes I_T, \quad (9)$$

This article estimates the elements of the unknown SUR error covariance matrix using the OR estimator following Abdel-Fattah (2013) instead of the usual estimation of Σ_M using the OLS method. Abdel-Fattah (2013) found out that estimating the elements of the unknown SUR error covariance matrix using the restricted residuals obtained from the OR estimator results in more efficient estimators than those based on the usual estimation of the error covariance matrix using the OLS method in terms of trace mean squared errors (TMSEs) criterion.

To deal with the problem of multicollinearity, Hoerl and Kennard (1970a,b) introduced the RR technique, then followed by numerous research studies. Srivastava and Giles (1987) defined the feasible general ridge estimator by;

$$\hat{\beta}_{FGR}(K) = (X'\hat{V}^{-1}X + K)^{-1}X'\hat{V}^{-1}Y, \quad (10)$$

$$\text{MSE}(\hat{\beta}_{FGR}(K)) = (X'\hat{V}^{-1}X + K)^{-1}(X'\hat{V}^{-1}X + K\beta\beta'K')(X'\hat{V}^{-1}X + K)^{-1}, \quad (11)$$

where; K is an unknown $p \times p$ diagonal matrix of nonnegative elements, $K = \text{diag}(K_1, K_2, \dots, K_M)$, $K_i = \text{diag}(k_{i1}, k_{i2}, \dots, k_{ip_i})$ and $k_{ij} > 0$ for $i = 1, 2, \dots, M$. By setting $K = 0$, we obtain $\hat{\beta}_{FGLS}$ and its covariance matrix.

2.1 Proposed SUR-ridge Estimators

We define the feasible SUR-ridge (FSR) estimator of β by;

$$\hat{\beta}_{FSR}(K) = \Gamma(\Lambda + K)^{-1}\Gamma'X'\hat{V}^{-1}Y, \quad (12)$$

$$\text{MSE}(\hat{\beta}_{FSR}(K)) = \Gamma(\Lambda + K)^{-1}(\Lambda + K\alpha\alpha'K')(\Lambda + K)^{-1}\Gamma', \quad (13)$$

where; $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ is the diagonal matrix of the eigenvalues of $X'\hat{V}^{-1}X$ and $\Gamma = (\gamma_{(1)}, \dots, \gamma_{(p)})$ is the orthogonal matrix of the standardized eigenvectors $\gamma_{(i)}$ of $X'\hat{V}^{-1}X$.

In the context of the SUR model, there exist numerous methods for selecting the SUR-ridge parameter, k_{ij} , and in order to describe them, we apply the suggested transformation of Srivastava and Giles (1987) on Eq. 1 and then rewrite the transformed model in the canonical form. Srivastava and Giles (1987) suggested transforming the model in Eq. 1 as;

$$Y^* = X^* \beta + u^*, \quad (14)$$

where; $Y^* = (\Sigma_M^{-1/2} \otimes I_T) Y$, $X^* = (\Sigma_M^{-1/2} \otimes I_T) X$ and $u^* = (\Sigma_M^{-1/2} \otimes I_T) u$.

Then we can get the canonical form of Eq. 14 as;

$$Y^* = Z \alpha + u^*, \quad (15)$$

where; $Z = X^* \Gamma$, $\alpha = \Gamma' \beta$ and $Z'Z = \Gamma' X^* \hat{V}^{-1} X^* \Gamma = \Lambda$.

The OLS and the SUR-ridge estimators of α are given by;

$$\hat{\alpha} = (Z'Z)^{-1} Z'Y^*, \quad (16)$$

$$\hat{\alpha}_{SUR}(K) = (Z'Z + K)^{-1} Z'Y^*, \quad (17)$$

$$TMSE(\hat{\alpha}_{SUR}(K)) = \sum_{i=1}^M \sum_{j=1}^P \frac{\lambda_{ij} + k_{ij}^2 \alpha_{ij}^2}{(\lambda_{ij} + k_{ij})^2}, \quad (18)$$

It follows that the value of k_{ij} which minimizes the $TMSE(\hat{\alpha}_{SUR}(K))$ is obtained by setting $\frac{\partial TMSE(\hat{\alpha}_{SUR}(K))}{\partial k_{ij}} = 0$ as;

$$k_{ij} = \frac{1}{\hat{\alpha}_{ij}^2}, \quad (19)$$

The SUR version of Hoerl and Kennard (1970a) OR parameter has been considered by Firinguetti (1997) as;

$$k_{ij(SHK)} = \frac{1}{\max_{ij}(\hat{\alpha}_{ij}^2)}, \quad (20)$$

The SUR version of Hoerl et al. (1975) harmonic mean has been considered by Firinguetti (1997) as;

$$k_{ij(Sharm)} = \frac{P}{\sum_{i=1}^M \sum_{j=1}^P \hat{\alpha}_{ij}^2}, \quad (21)$$

The SUR versions of the arithmetic mean, the geometric mean and the median proposed by Kibria (2003) have been considered by Alkhamisi and Shukur (2008) as;

$$k_{ij}(\text{Sarith}) = \frac{1}{P} \sum_{i=1}^M \sum_{j=1}^{P_i} \frac{1}{\hat{\alpha}_{ij}^2}, \quad (22)$$

$$k_{ij}(\text{Sgeom}) = \frac{1}{\left(\prod_{i=1}^M \prod_{j=1}^{P_i} \hat{\alpha}_{ij}^2 \right)^{\frac{1}{P}}}, \quad (23)$$

$$k_{ij}(\text{Smed}) = \text{median}_j \left(\frac{1}{\hat{\alpha}_{ij}^2} \right), \quad (24)$$

Alkhamisi et al. (2006) ridge parameter for the multiple regression model has been generalized by Alkhamisi and Shukur (2008) as;

$$k_{ij}(\text{Smax}) = \max_j \left(\frac{1}{\hat{\alpha}_{ij}^2} \right), \quad (25)$$

Alkhamisi and Shukur (2008) proposed the following;

$$k_{ij}(\text{Sqarith}) = \text{mean}_j \left(\frac{1}{\sqrt{\hat{\alpha}_{ij}^2}} \right), \quad (26)$$

$$k_{ij}(\text{Sqmax}) = \max_j \left(\frac{1}{\sqrt{\hat{\alpha}_{ij}^2}} \right), \quad (27)$$

Zeebari et al. (2012) modified the SUR-ridge parameters (Eq.19-27) by raising them to the power of $1/\bar{P}$, where; $\bar{P} = \sum_{i=1}^M P_i / M$. We refer to these modified parameters as $k_{ij}(\text{M}...)$.

Abdel-Fattah (2013) proposed k_{ij} as;

$$k_{ij}(\text{Sad}) = \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_j \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| \right)^{\frac{1}{\bar{P}}}, \quad (28)$$

$$k_{ij}(\text{Smad}) = \left(\frac{1}{P} \sum_{i=1}^M \sum_{j=1}^{P_i} \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_j \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| \right) \right)^{\frac{1}{\bar{P}}}, \quad (29)$$

$$k_{ij}(\text{Smedad}) = \left(\text{median} \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| \right) \right)^{\frac{1}{P}}, \quad (30)$$

$$k_{ij}(\text{Sminad}) = \left(\min \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| \right) \right)^{\frac{1}{P}}, \quad (31)$$

The current article proposes new SUR-ridge parameters as;

$$\text{(New)} k_{ij}(\text{MSad}) = \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| + \frac{1}{\lambda_{ij}} \right)^{\frac{1}{P}}, \quad (32)$$

$$\text{(New)} k_{ij}(\text{MSmad}) = \left(\frac{1}{P} \sum_{l=1}^M \sum_{j=1}^{P_l} \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| + \frac{1}{\lambda_{ij}} \right) \right)^{\frac{1}{P}}, \quad (33)$$

$$\text{(New)} k_{ij}(\text{MSmedad}) = \left(\text{median} \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| + \frac{1}{\lambda_{ij}} \right) \right)^{\frac{1}{P}}, \quad (34)$$

$$\text{(New)} k_{ij}(\text{MSminad}) = \left(\min \left(\left| \sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} - \text{mean}_y \left(\sqrt{\frac{1}{\hat{\alpha}_{ij}^2}} \right) \right| + \frac{1}{\lambda_{ij}} \right) \right)^{\frac{1}{P}}, \quad (35)$$

For the suggested SUR-ridge parameters defined by Eq. 32-35, we used the acronyms MSad, MSmad, MSmedad and MSminad; where; MS: denotes the modified SUR, ad: denotes the absolute deviation, m: denotes the mean, med: denotes the median and min: denotes the minimum.

3. An Application

In this section, we consider the data due to Theil (1971). This data have been analyzed by Judge et al. (1988) as an example for the estimation of two seemingly unrelated regressions, where;

- X_{12} and X_{22} are the market values and X_{13} and X_{23} are the capital stocks of General Electric and Westing-house, respectively, for the period 1934 through 1953. There are four exogenous variables in two regression equations.
- The endogenous variables Y_1 and Y_2 may be regarded as investment figures for the two companies, as tabulated in Judge et al. (1988).

We consider the following system of equations;

$$\begin{aligned} Y_1 &= \beta_{11} + X_{12}\beta_{12} + X_{13}\beta_{13} + u_1, \\ Y_2 &= \beta_{21} + X_{22}\beta_{22} + X_{23}\beta_{23} + u_2, T = 20 \end{aligned} \quad (36)$$

The correlation matrix of the X variables is;

	X_{12}	X_{13}	X_{22}	X_{23}
X_{12}	1	0.1182	<u>0.6852</u>	0.0916
X_{13}	0.1182	1	<u>0.7405</u>	<u>0.9744</u>
X_{22}	0.6852	0.7405	1	<u>0.7417</u>
X_{23}	0.0916	0.9744	0.7417	1

From the correlation matrix it is clear that the intercorrelations are across the equations and in the second equation. X_{13} and X_{23} are highly correlated, which is the same for X_{22} and X_{23} also for X_{13} and X_{22} . The value of the correlation coefficient between X_{12} and X_{22} indicate a moderate positive linear relationship.

Following Judge et al. (1988), we fit the SUR model with the intercept. Using the ordinary least squares estimation for each equation of the system, the TMSE for the first equation equals 1088.77, for the second equation equals 114.66 and for the full system equals 1203.43. The devisor $\text{trace}(P_{X_i} P_{X_j})$ was used to obtain $\hat{\Sigma}_M$ instead of $T - (P/M)$. Under the first devisor the TMSE of the FGLS estimator equal 1013.28, while under the second devisor it equals 1022.06. Under this model the condition number of $X' \hat{V}^{-1} X$ is 17902.86, which implies the existence of multicollinearity in the data set.

The estimated TMSEs of the SUR-ridge regression coefficients are presented in tables 2 and 3 (appendix b). From the tables we observed that the sign for the exogeneous variable X_{22} has been changed for the Sarith and the Smax estimators and the sign for the exogeneous variable X_{23} has been changed for the Sarith estimator. All proposed estimators performed better than the OLS and the FGLS in the sense that they have smaller TMSEs. None of them performed better than Srivastava and Giles (1987). Nevertheless, MSminad performed better than the other proposed estimators. In general, the proposed estimators performed the same as Abdel-Fattah (2013) estimators in terms of the TMSEs.

Table 3 shows the results of estimating the error covariance matrix using the restricted residuals obtained from the OR estimator. We have an indication that the data of the second regression equation is collinear. Thus the RR technique was applied on the second regression equation, where the ridge parameter ad performed better than the other estimators (appendix c). Then the SUR error covariance matrix was estimated by the OR estimator. As a result of estimating the SUR error covariance matrix using the restricted residuals obtained from the OR estimator, when compared with the OLS estimator, the efficiency of the FGLS and the SUR-ridge estimators increases which is clear from tables 2 and 3.

3.1 Conclusion

The main results revealed that estimating the elements of the unknown SUR error covariance matrix using the restricted residuals obtained from the OR estimator results in more efficient estimators than those based on the usual estimation of the error covariance matrix using the OLS method in terms of TMSE criterion. The proposed feasible SUR-ridge estimators based on $k_{ij}(MSad)$, $k_{ij}(MSmad)$, $k_{ij}(MSmedad)$ and $k_{ij}(MSminad)$ have been shown to give new better solutions to the multicollinearity problem in terms of TMSEs values. In general, the proposed estimators performed the same as Abdel-Fattah (2013) estimators in terms of the TMSEs. All proposed estimators performed better than the OLS and the FGLS in the sense that they have smaller TMSEs. None of them performed better than SK (Srivastava and Giles, 1987). Nevertheless, MSminad performed better than the other proposed estimators. The feasible SUR-ridge estimators based on $k_{ij}(Smax)$ and $k_{ij}(Sarith)$ have produced the worst TMSEs values among the other estimators.

Acknowledgment

The author would like to express his deep thanks to Prof. Amany M. Mohammed and Prof. Ahmed H. Youssef for their valuable advices during the preparation of this research.

Appendices:

Appendix (A)
Table 1 Formulae for Selecting the Ridge
Parameter in the Multiple Regression Model

SN	Author / Year	Ridge Parameters	
1.	Hoerl and Kennard (1970a)	$k_i = \frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}$	Eq. (A.1)
2.	Hoerl and Kennard (1970a)**	$k_{i(M)} = \left(\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2} \right)^{\frac{1}{P_i}}$	Eq. (A.1a)
3.	Hoerl and Kennard (1970a)	$k_{i(HK)} = \frac{\hat{\sigma}_i^2}{\max(\hat{\theta}_j^2)}$	Eq. (A.2)
4.	Hoerl and Kennard (1970a)**	$k_{i(MHK)} = \left(\frac{\hat{\sigma}_i^2}{\max(\hat{\theta}_j^2)} \right)^{\frac{1}{P_i}}$	Eq. (A.2a)
5.	Hoerl et al. (1975)	$k_{i(harm)} = \frac{P_i}{\sum_{j=1}^{P_i} \frac{1}{k_i}} = \frac{P_i \hat{\sigma}_i^2}{\sum_{j=1}^{P_i} \hat{\theta}_j^2}$	Eq. (A.3)
6.	Hoerl et al. (1975)**	$k_{i(Mharm)} = \left(\frac{P_i \hat{\sigma}_i^2}{\sum_{j=1}^{P_i} \hat{\theta}_j^2} \right)^{\frac{1}{P_i}}$	Eq. (A.3a)
7.	Kibria (2003)	$k_{i(arith)} = \frac{1}{P_i} \sum_{j=1}^{P_i} \frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}$	Eq. (A.4)
8.	Kibria (2003)**	$k_{i(Marith)} = \left(\frac{1}{P_i} \sum_{j=1}^{P_i} \frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2} \right)^{\frac{1}{P_i}}$	Eq. (A.4a)

SN	Author / Year	Ridge Parameters	
9.	Kibria (2003)	$k_{i(\text{geom})} = \frac{\hat{\sigma}_i^2}{\left(\prod_{j=1}^{p_i} \hat{\theta}_j^2\right)^{\frac{1}{p_i}}}$	Eq. (A.5)
10.	Kibria (2003)**	$k_{i(\text{Mgeom})} = \left(\frac{\hat{\sigma}_i^2}{\left(\prod_{j=1}^{p_i} \hat{\theta}_j^2\right)^{\frac{1}{p_i}}} \right)^{\frac{1}{p_i}}$	Eq. (A.5a)
11.	Kibria (2003)	$k_{i(\text{med})} = \text{median}\left(\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}\right)$	Eq. (A.6)
12.	Kibria (2003)**	$k_{i(\text{Mmed})} = \left(\text{median}\left(\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}\right) \right)^{\frac{1}{p_i}}$	Eq. (A.6a)
13.	Alkhamisi et al. (2006)	$k_{i(\text{max})} = \max\left(\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}\right)$	Eq. (A.7)
14.	Alkhamisi et al. (2006)**	$k_{i(\text{Mmax})} = \left(\max\left(\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}\right) \right)^{\frac{1}{p_i}}$	Eq. (A.7a)
15.	Alkhamisi and Shukur (2008)*	$k_{i(\text{arithsq})} = \text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right)$	Eq. (A.8)
16.	Alkhamisi and Shukur (2008)**	$k_{i(\text{Marithsq})} = \left(\text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right)^{\frac{1}{p_i}}$	Eq. (A.8a)
17.	Alkhamisi and Shukur (2008)*	$k_{i(\text{maxsq})} = \max\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right)$	Eq. (A.9)
18.	Alkhamisi and Shukur (2008)**	$k_{i(\text{Mmaxsq})} = \left(\max\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right)^{\frac{1}{p_i}}$	Eq. (A.9a)

SN	Author / Year	Ridge Parameters
19.		$k_{i(ad)} = \left(\left \sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}} - \text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right \right)^{\frac{1}{P_i}} \quad \text{Eq. (A.10)}$
20.	Abdel-Fattah (2013)	$k_{i(mad)} = \left(\frac{1}{P_i} \sum_{j=1}^{P_i} \left(\left \sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}} - \text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right \right)^{\frac{1}{P_i}} \right)^{\frac{1}{P_i}} \quad \text{Eq.(A.11)}$
21.		$k_{i(medad)} = \left(\text{median} \left(\left \sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}} - \text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right \right)^{\frac{1}{P_i}} \right)^{\frac{1}{P_i}} \quad \text{Eq.(A.12)}$
22.		$k_{i(minad)} = \left(\min \left(\left \sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}} - \text{mean}\left(\sqrt{\frac{\hat{\sigma}_i^2}{\hat{\theta}_j^2}}\right) \right \right)^{\frac{1}{P_i}} \right)^{\frac{1}{P_i}} \quad \text{Eq. (A.13)}$

* Extensions to the SUR model case.

** We extend Zeebari et al. (2012) by raising the parameters to the power of $1/P_i$.

Appendix (B): Table 2 The Estimated TMSEs and the Regression Coefficients of Different Estimators (Σ_M is estimated by OLS)

	OLS	Sad	Smad	Smedad	Smnad	SK	SHK	Sharm	Sarith	Sgeom	Smed	Smax	Sqarith	Sqmax
TMSE	1203.43	16.59	16.60	16.57	16.54	13.22	13.38	15.29	16.81	16.77	16.79	16.82	16.77	16.79
$\hat{\beta}_{11}$	5.2787	0.0072	0.0065	0.0076	0.0076	0.2272	0.2453	0.0577	0.0000	0.0008	0.0003	0.0000	0.0008	0.0002
$\hat{\beta}_{12}$	0.0236	0.0260	0.0260	0.0260	0.0260	0.0273	0.0256	0.0258	0.0386	0.0268	0.0279	0.0349	0.0268	0.0293
$\hat{\beta}_{13}$	0.1562	0.1587	0.1586	0.1586	0.1586	0.1490	0.1596	0.1592	0.0714	0.1546	0.1490	0.0238	0.1545	0.1419
$\hat{\beta}_{21}$	-0.5710	-0.0269	-0.0245	-0.0283	-0.0283	-0.9060	-0.9016	-0.2160	0.0001	-0.0030	-0.0011	0.0001	-0.0029	-0.0005
$\hat{\beta}_{22}$	0.0623	0.0608	0.0608	0.0608	0.0608	0.0656	0.0619	0.0609	0.0638	0.0621	0.0638	0.0442	0.0621	0.0656
$\hat{\beta}_{23}$	0.0721	0.0826	0.0823	0.0825	0.0825	0.0545	0.0827	0.0835	-0.0025	0.0719	0.0580	0.0025	0.0716	0.0424
	FGLS	(New) MSad	(New) MSmad	(New) MSmedad	(New) MSmnad	MSK	MSHK	MSharm	MSarith	MSgeom	MSmed	MSmax	MSqarith	MSqmax
TMSE	1013.28	16.59	16.60	16.57	16.54	15.45	15.45	15.96	16.75	16.58	16.64	16.77	16.59	16.67
$\hat{\beta}_{11}$	2.3522	0.0072	0.0065	0.0076	0.0076	0.0490	0.0507	0.0298	0.0015	0.0072	0.0053	0.0008	0.0072	0.0042
$\hat{\beta}_{12}$	0.0242	0.0261	0.0261	0.0261	0.0261	0.0260	0.0258	0.0259	0.0264	0.0260	0.0260	0.0268	0.0260	0.0261
$\hat{\beta}_{13}$	0.1607	0.1577	0.1577	0.1577	0.1577	0.1585	0.1592	0.1591	0.1566	0.1586	0.1584	0.1546	0.1586	0.1582
$\hat{\beta}_{21}$	-3.3555	-0.0269	-0.0245	-0.0283	-0.0283	-0.1901	-0.1896	-0.1118	-0.0055	-0.0270	-0.0199	-0.0030	-0.0268	-0.0158
$\hat{\beta}_{22}$	0.0655	0.0609	0.0609	0.0609	0.0609	0.0611	0.0609	0.0608	0.0615	0.0608	0.0609	0.0621	0.0608	0.0609
$\hat{\beta}_{23}$	0.0799	0.0810	0.0810	0.0810	0.0810	0.0816	0.0835	0.0835	0.0770	0.0825	0.0819	0.0721	0.0825	0.0814

Table 3 The Estimated TMSFs and the Regression Coefficients of Different Estimators (Σ_M is estimated by RR)

	FGLS	Sad	Smad	Smedad	Sminad	SK	SHK	Sharm	Sarith	Sgeom	Smed	Smax	Sqarith	Sqmax
TMSE	995.78	16.00	16.02	15.99	15.96	12.43	12.60	14.63	16.23	16.19	16.20	16.24	16.19	16.21
$\hat{\beta}_{11}$	2.1690	0.0072	0.0065	0.0075	0.0075	0.2381	0.2548	0.0594	0.0000	0.0008	0.0004	0.0000	0.0007	0.0002
$\hat{\beta}_{12}$	0.0242	0.0260	0.0260	0.0260	0.0260	0.0272	0.0256	0.0258	0.0391	0.0269	0.0279	0.0310	0.0270	0.0298
$\hat{\beta}_{13}$	0.1610	0.1588	0.1587	0.1588	0.1588	0.1493	0.1598	0.1594	0.0588	0.1542	0.1492	0.0176	0.1538	0.1394
$\hat{\beta}_{21}$	-3.3918	-0.0277	-0.0249	-0.0289	-0.0289	-0.9633	-0.9592	-0.2272	0.0001	-0.0029	-0.0012	0.0001	-0.0026	-0.0004
$\hat{\beta}_{22}$	0.0655	0.0608	0.0608	0.0608	0.0608	0.0655	0.0620	0.0609	0.0609	0.0622	0.0637	0.0379	0.0623	0.0662
$\hat{\beta}_{23}$	0.0803	0.0827	0.0824	0.0826	0.0826	0.0559	0.0829	0.0837	-0.0019	0.0711	0.0590	0.0027	0.0701	0.0379
	FGLS	(New) MSad	(New) MSmad	(New) MSmedad	(New) MSminad	MSK	MSHK	MSharm	MSarith	MSgeom	MSmed	MSmax	MSqarith	MSqmax
TMSE	995.78	16.00	16.02	15.99	15.96	14.76	14.76	15.32	16.17	15.99	16.04	16.19	16.00	16.09
$\hat{\beta}_{11}$	2.1690	0.0072	0.0065	0.0075	0.0075	0.0522	0.0538	0.0315	0.0014	0.0074	0.0057	0.0008	0.0072	0.0042
$\hat{\beta}_{12}$	0.0242	0.0261	0.0261	0.0261	0.0261	0.0260	0.0258	0.0258	0.0265	0.0260	0.0260	0.0269	0.0260	0.0261
$\hat{\beta}_{13}$	0.1610	0.1579	0.1579	0.1579	0.1579	0.1587	0.1593	0.1592	0.1564	0.1587	0.1586	0.1542	0.1587	0.1583
$\hat{\beta}_{21}$	-3.3918	-0.0277	-0.0249	-0.0289	-0.0289	-0.2063	-0.2059	-0.1205	-0.0052	-0.0284	-0.0218	-0.0029	-0.0276	-0.0161
$\hat{\beta}_{22}$	0.0655	0.0609	0.0609	0.0609	0.0609	0.0611	0.0609	0.0608	0.0615	0.0608	0.0609	0.0622	0.0608	0.0609
$\hat{\beta}_{23}$	0.0803	0.0812	0.0811	0.0812	0.0812	0.0819	0.0837	0.0836	0.0765	0.0826	0.0822	0.0710	0.0826	0.0815

Appendix (C): The Performance of some Ridge Estimators

Figure C.1

The Estimated TMSEs for Different Ridge Estimators

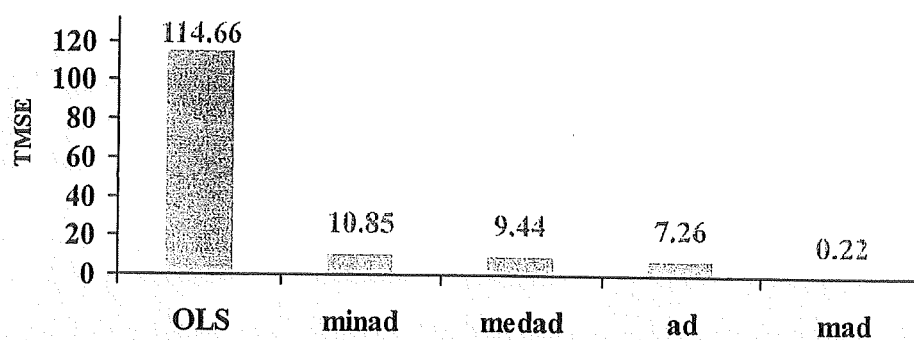


Figure C.2

The performance of Different Ridge Estimators

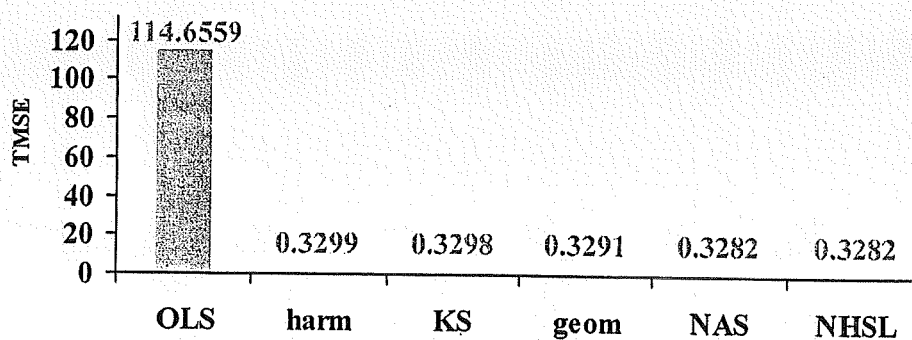


Figure C.3

The Performance of Modified Ridge Estimators

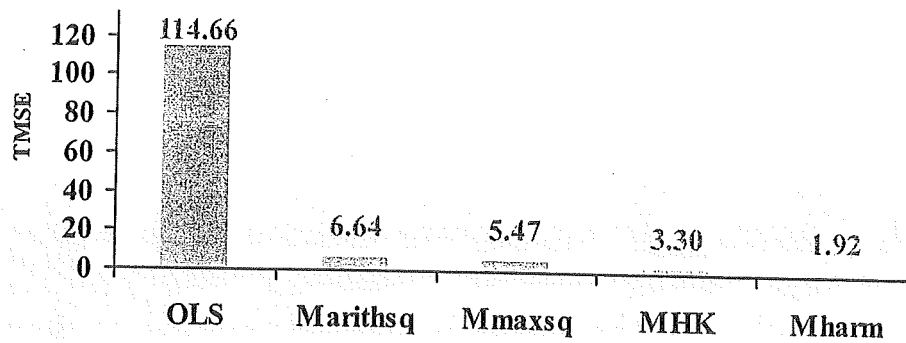
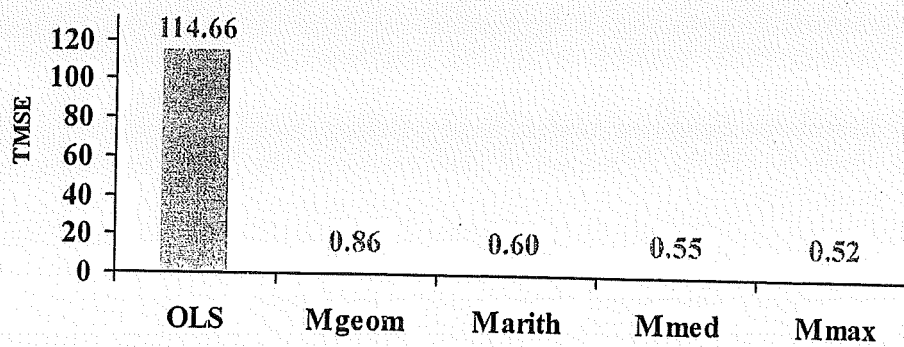


Figure C.4

The Estimated TMSEs for Modified Ridge Estimators



References

1. Abdel-Fattah, M. A. (2013), "A Statistical Approach to Handle Multicollinearity in SUR", Presented at the 25th Annual International Conference on Statistics and Modeling in Human and Social Sciences, Cairo University, Cairo, Egypt.
2. Alkhamisi, M. A. and Shukur, G. (2008), "Developing Ridge Parameters for SUR Model", *Communications in Statistics-Theory and Methods*, 37(4), 544-564.
3. Alkhamisi, M. A., Khalaf, G. and Shukur, G. (2006), "Some Modifications for Choosing Ridge Parameter", *Communications in Statistics-Theory and Methods*, 35, 1-16.
4. Firinguetti, L. (1997), "Ridge Regression in the Context of a System of Seemingly Unrelated Regression Equations", *Journal of Statistical Computation and Simulation*, 56(2), 145-162.
5. Hoerl, A. E. and Kennard, R. W. (1970a), "Ridge Regression: Biased Estimation for Nonorthogonal Problems", *Technometrics*, 12(1), 55-67.
6. Hoerl, A. E. and Kennard, R. W. (1970b), "Ridge Regression: Applications to Nonorthogonal Problems", *Technometrics*, 12(1), 69-82.
7. Hoerl, A. E., Kennard, R. W. and Baldwin, K.F. (1975), "Ridge Regression: Some Simulations", *Communications in Statistics-Simulation and Computation*, 4, 105-123.
8. Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. and Lee, T.C. (1988), *Introduction to the Theory and Practice of Econometrics*, (2nd ed.), John Wiley and Sons, Inc., New York.
9. Kibria, B. M. G. (2003), "Performance of Some New Ridge Regression Estimators", *Communications in Statistics-Simulation and Computation*, 32, 419-435.
10. R Development Core Team (2012), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
11. Srivastava, V. K. and Giles, D. E. A. (1987), *Seemingly Unrelated Regression Equations Models*, Marcel Dekker, Inc., New York.
12. Theil, H. (1971), *Principles of Econometrics*, John Wiley and Sons, Inc., New York.
13. Zeebari, Z., Shukur, G. and Kibria, B. M. G. (2012), "Modified Ridge Parameters for Seemingly Unrelated Regression Model", *Communications in Statistics-Theory and Methods*, 41(9), 1675-1691.
14. Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Relations and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 348-67.
15. Zellner, A. (1963), "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results", *Journal of the American Statistical Association*, 58, 977-992.

Time Series Analysis of Inflation in Egypt

Kandil, S.M*

Abstract

In this paper, inflation rate in Egypt using Box-Jenkins methodology and Linear Regression Analysis has been examined to reach the appropriate model during the period (January 1996 to December 2012). The fit model for the inflation rate was SARIMA (1,1,0)(0,0,1)₁₂ and the forecasted values of inflation rate of January to March 2013 was 4.93%, 4.81% and 4.80% respectively.

Keywords: Inflation, Consumer Price Index, SARIMA models, Dickey-Fuller test.

1. Introduction

The main Objectives for studying inflation are to understand the dynamics of the process generating it and to discovery the quantitative relationships to predict the path of inflation in the future. As well the importance of having variables help in understanding and interpretation it, meaning affording the possibility for the owner of the authority to control it up or down, it is known in economic literature as "inflation targeting."

Economic inflation: is one of the most economic prevalent terms, despite the prevalence of the term, it does not have an agreement between economists on its definition this is due to the division on defining the concept of inflation and is used to describe a number of different situations such as (excessive in the general level of price - Rising costs - excessive creation of cash balances).

* Statistical analyst, CAPMAS. Email: kandil.stat@gmail.com.

Inflation in Webster's New Universal Unabridged Dictionary defined by "An increase in the amount of currency in circulation, resulting in a relatively sharp and sudden fall in its value and rise in prices: it may be caused by an increase in the volume of paper money issued or of gold mined, or a relative increase in expenditures when the supply of goods fails to meet the demand". Other definition of inflation is "a continual increase in the price level, affects individuals, businesses, and the government" (Frederic S. Mishkin 1995).

Central Agency for Public Mobilization and Statistics (CAPMAS) produces the index of consumer prices (CPI) by collecting food prices and industrial goods and services a month from urban and rural areas during the period from Day 1 to Day 28 of each month, the number of sources (10085) from urban and (4337) from rural for the ninth series of CPI data start from August 2010.

The CPI is published according to The International Classification of individual Consumption by Purpose (COICOP-HBS 2003), which is included in the National Accounts System (SNA, 1993). COICOP is a recommended classification by many international organizations as International Labor Force organization (ILO) and International Monetary Fund (IMF). COICOP is also used in the International Comparison Program (ICP) (CAPMAS 2013). Weights of major components of derived for all items and main sections, Main sections of the used weights consists of 12 sections which include: (Food and Non Alcoholic Beverages - Alcoholic Beverages, Tobacco and Narcotics - Clothing and Footwear - Housing, Water, Electricity, Gas and other fuels - Furnishings, Household, Equipment and Routine Maintenance of The house -Health - Transport - Communications - Recreation and Culture - Education - Restaurants and Hotels - Miscellaneous Goods and Services). The share of each section distributed according to Table 1. The weights are calculated according to equation(1).

$$W_{ij} = \frac{A_{ij}}{\sum_i \sum_j A_{ij}} \times 10000 \quad (1)$$

Where: A_{ij} : value of item

$i=1,2, \dots, 12$ (Main Section) and

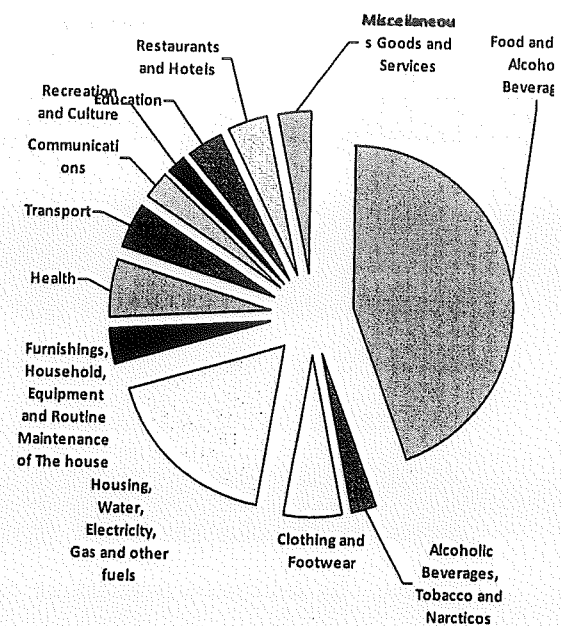
$j=1,2, \dots, 479$ (Commodity or Service item)

Table(1): Weights of CPI according to main sections

Main Section	Weight	%
Food and Non-Alcoholic	4421.3	44.2
Alcoholic Beverages, Tobacco	245.0	2.4
Clothing and Footwear	580.7	5.8
Housing, Water, Electricity, Gas and other fuels	1769.5	17.7
Furnishings, Household, Equipment and Routine	364.5	3.6
Health	587.7	5.9
Transport	462.0	4.6
Communications	258.0	2.6
Recreation and Culture	201.9	2.0
Education	351.0	3.5
Restaurants and Hotels	415.1	4.2
Miscellaneous Goods and	343.1	3.4
Total	10000	100

Source: CAPMAS, December 2012

Figure1: Weights of CPI



One of the main goals of the central banks is to maintain the inflation rate at a low, fixed level. Central banks followed the specific plans to achieve the inflation target or access to values close to the target rate for the effectiveness of monetary policy more. As well as economic units (private or public) monitors estimates prices closely in order to make decisions that will reach the highest payoff for the use of their available resources for that highlights the importance

of forecasting value of inflation and estimate a model for inflation to predict the correct rate of inflation by central banks or economic units.

2. Literature Review

Based on the importance and the need for analysis of inflation data, a lot of studies concerned with estimate inflation model such as Hamad (2007), Massimiliano (2008) and Richard and Stephen (2012). On the other hand, some studies analyzed the inflation using econometrical methods such as IDSC (2004), Hatem (2007) and Sharif (2012). Other studies have also dealt with inflation from an economic perspective such as Nevin (2001). Nevin (2001) studied the optimization rate of inflation which checks has economic growth target. According to her analysis, the monetary policy is specified basic rate of inflation in the long term and economic success fixed the exchange rate to control inflation but it caused on shocks of the currencies of these countries. IDSC (2004) studied inflation in Egypt after liberalization of the Egyptian pound in 29th January, 2003. This led to increase in the inflation rate with decrease in the standards of living and the inability of many segments of citizens to fulfill many of their basic needs. He aimed to measure the Pass-through effect, which was the impact of the changes in the Egyptian Pound exchange rate on domestic price levels, by estimated percentage change in inflation, due to changes in exchange rate using a Vector Auto Regression model. Hatem (2007) focused on the factors that could effect on inflation in the Gulf Cooperation Council (GCC) countries. For that purpose, he used a series of annual data for a (GCC) and econometric methods to identify the important variables that affect inflation, and estimated the relationship between inflation and those factors. Hamad (2007) used ARIMA models in studying and analyzing monthly time series for the inflation rate in Saudi Arabia for the period February 1980 to September 2004 to identify the pattern of change in the rate of inflation, then build a model from ARIMA family that helps to predict the values of the rate of inflation in the short term. The model of ARIMA showed some seasonal effects on the inflation rate during the study period.

Massimiliano (2008) focused on Predicting the future evolution of GDP growth and inflation in US for the period January 1959 to February 2004 through models based economic theory and time series model. It used bootstrap algorithm to evaluate the statistical significance of results. The main conclusion was -in general- linear time series models can be hardly beaten if they are carefully specified. Richard and Stephen (2012) discussed the time series analysis of Canada's inflation rate from 1995 to the end of 2011. The results shown significant lags in autoregressive process at lags 5,7,12 and 24, also, moving average process at lag 1,5,12 and 24. The appropriate model was SARMA(0,0,0)(2,0,2)₁₂. Sharif (2012) reviewed basic fundamentals inflation targeting and determined the criteria success of the Central Bank of Egypt in the implementation of these policies since 2005, date of the announcement of an official policy towards inflation targeting. He used simple vector autoregressive (VAR) model to analyze time series (January 2000 to December 2010) of inflation rate and tests the sensitivity of inflation to changes in selected monetary variable. Final result indicated that interest rate channel has a weak effect on inflation rate, but it was strengthened after the introduction of the corridor in 2005. Also, inflation rate has positive relation with exchange rate market, and this justifies using short run interest rate as an operational target by the CBE in the meantime. For better inflation targeting policy must met CBE independence which is the ultimate and the main prerequisite for adoption of inflation targeting.

3. Theoretical Background

Methodology of Box and Jenkins Based on two tools Autoregressive (AR) and Moving Average (MA), but Box and Jenkins Reunion in one methodology

$$z_t = \Phi_1 z_{t-1} + \dots + \Phi_p z_{t-p} - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q} + \varepsilon_t \dots (2)$$

known as ARIMA Models, where "I" standing for "Integrated", which level of stability in the series after a certain number of differences. The ARMA model has the following form:

Where:

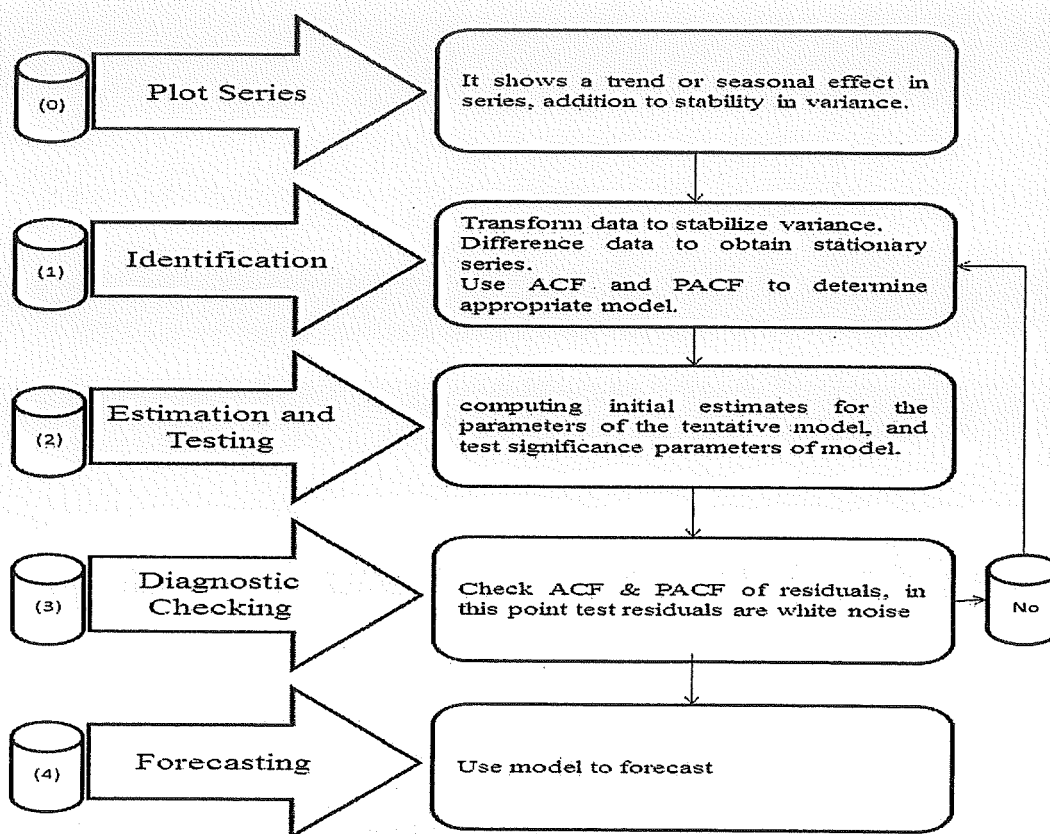
$Z_t, Z_{t-1}, Z_{t-2}, \dots$ the stationary data points;

$\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots$ the present and prior forecast errors;

$\mu, \phi_t, \phi_{t-1}, \phi_{t-2}, \dots, \Theta_t, \Theta_{t-1}, \Theta_{t-2}, \dots$ parameters of regression model.

Where ϕ_{t-i} a function in rank of Autoregressive (AR), i equal periods of deceleration, while Θ_{t-i} a function in value of error, model of ARMA built according to the following steps outlined in figure (2). Methodology Box-Jenkins can be extended to include seasonal autoregressive and seasonal moving average terms.

Figure2: THE BASIC STEPS IN THE BOX-JENKINS PROCEDURE



4. Model (Identification and Estimation)

Inflation is a measure as the percent of annual change in the consumer price index (CPI). So, the following formula (CSO of Yemen 2010) will be used:

$$\text{Inflation Rate} = [(CPI_t - CPI_{t-1}) / CPI_{t-1}] * 100 \quad (3)$$

Where:

CPI_t : CPI in current year

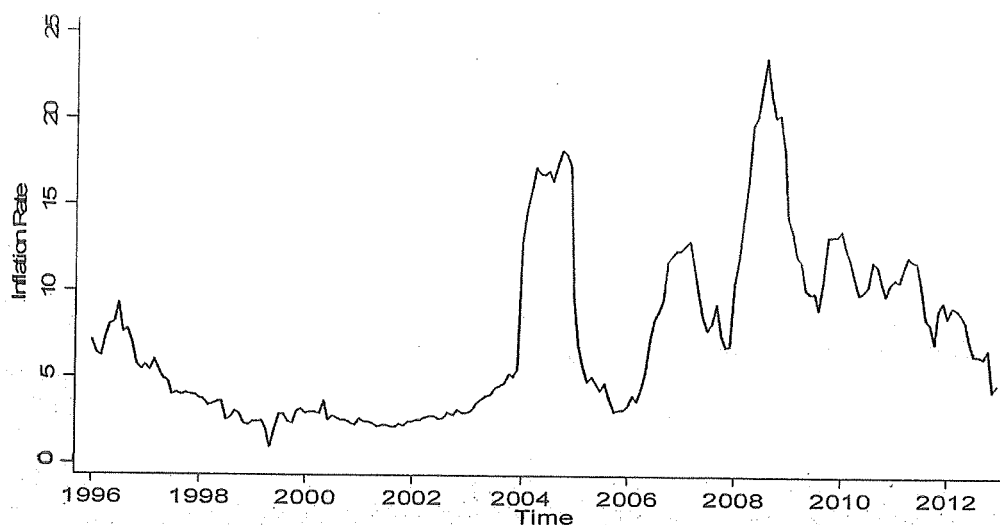
CPI_{t-1} : CPI in previous year, $t=2,3, \dots, T$

Calculate monthly inflation by using the following formula:

$$[(CPI_{\text{January 2008}} - CPI_{\text{January 2007}}) / CPI_{\text{January 2007}}] * 100 \quad (4)$$

Egypt passed through periods of stability and other periods rise in inflation through 1996 to 2012. The period from 1996 has high inflation as a result of setting fire of the financial crisis in Asia and consequent on that of decline in flows of foreign investment and Increase imbalance in the balance of trade, followed by the stability in rates of inflation even 2004 where exchange rate deregulation Led to rising inflation again ranged between 12% and 18%, and during 2008 and 2009 Egypt suffered from negative update for higher commodity prices globally and the subprime crisis in the United States, which led to high rates of inflation to reach 23.6% in August 2008. the rate of inflation during the From January to December of 2012 has decrease, where the rate moved from 8.5% January 2012 to 4.6% in December of the same year due to the economic slowdown and the decline in the purchasing power, and the decline in demand for commodities.

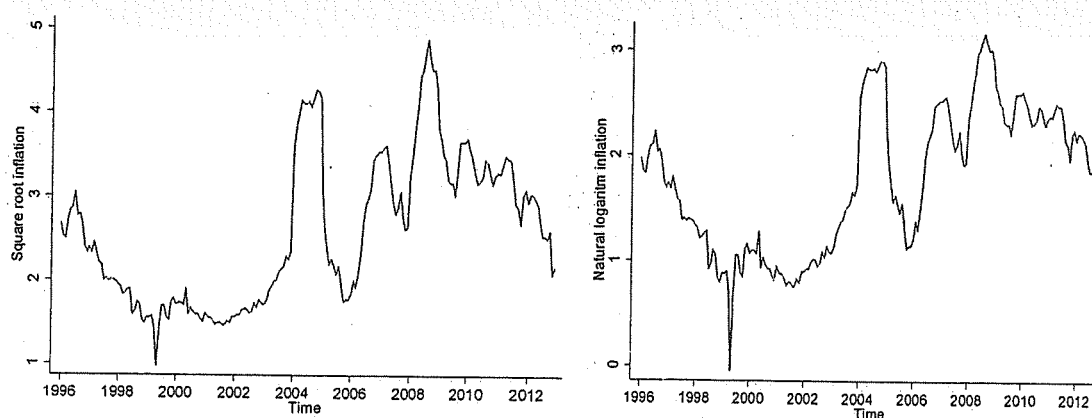
Figure3: Inflation Rate in Egypt (1996:2012) monthly data



4.1. Methodology of Box and Jenkins

According to figure (3), it can be noticed that Inflation series has instability in mean and variance. To solve instability in variance mathematical transformation such as square root or Natural logarithm will be used while ACF and PACF plots or Dickey-Fuller test will be used to check for any non-stationarity in mean, In the case of moral instability in the Mean, the differences will be used for soothing it. Figure (4) below shows transformation of Natural logarithm which has the same efficiency with square root in stationarity variance (model was action in the case of the square root, the result showed a high

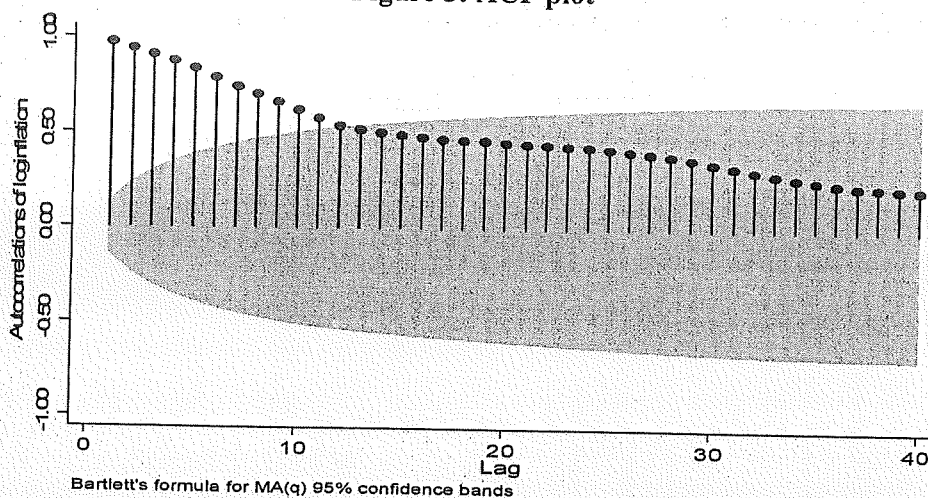
Figure 4: Transformation of Natural logarithm and square root



capacity for Natural logarithm about square root of the sample data modeling), we can choose Natural logarithm where majority of the literature of time series analysis indicates to the ability of transformation of logarithm in the soothing of the series compared with other transformations.

In this part we check the mean of inflation rate is constant or not through ACF plot and Dickey-Fuller test. It's clear from figure (5) that the series is not constant in mean.

Figure 5: ACF plot



Dickey-Fuller test indicate that the mean is not constant, so the first difference will be taken to produce series constant in mean.

Dickey-Fuller test for unit root				Number of obs = 203
----- Interpolated Dickey-Fuller -----				
Test	1% Critical	5% Critical	10% Critical	
Statistic	Value	Value	Value	
Z(t)	-1.609	-3.476	-2.883	-2.573
MacKinnon approximate p-value for Z(t) = 0.4790				

Figure 6: Stage of new series constant in (Variance and Mean)

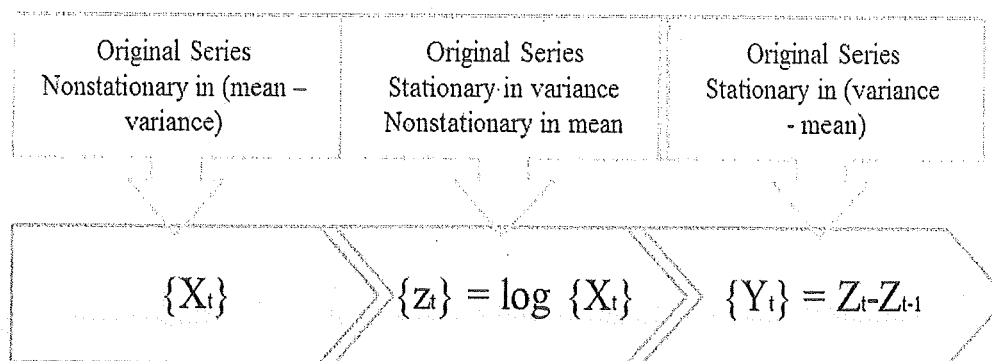
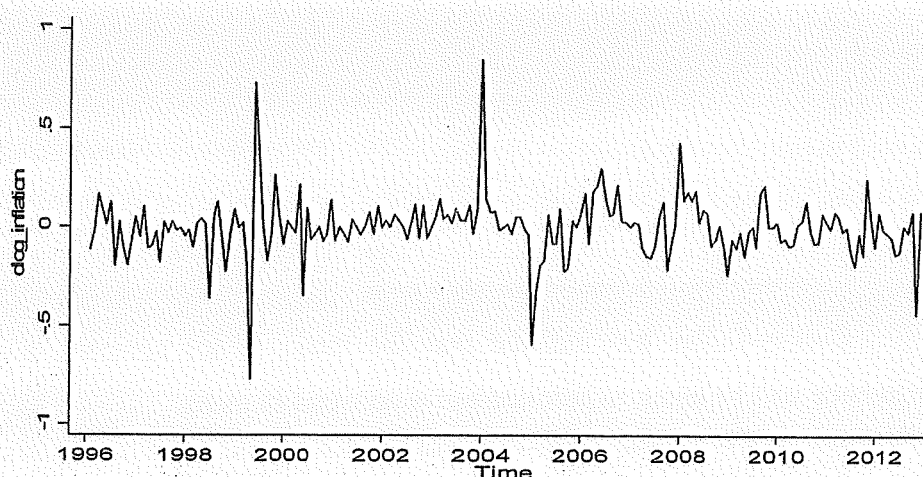


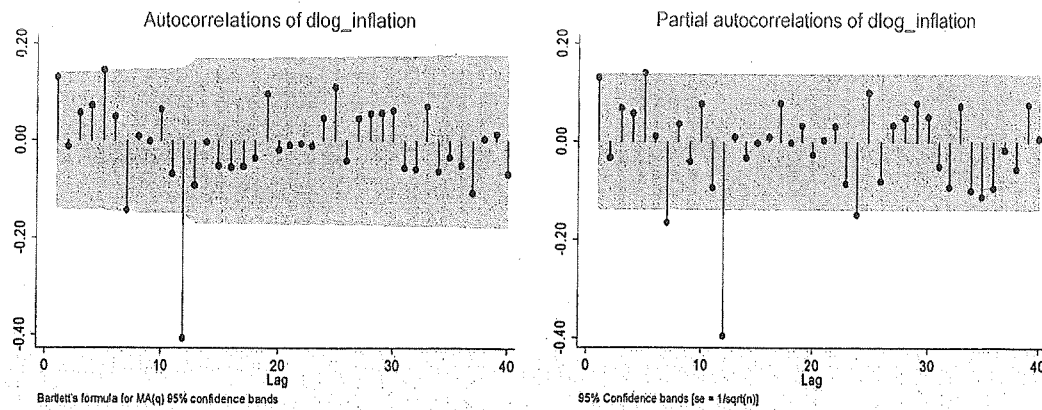
Figure (7) represents the new series Y_t , which is constant in mean.

Figure 7: First difference plot



According to figure(8) it can be noticed that ACF and PACF of $\{Y_t\}$ is large at lag 12 while lag 1 close to significance and PACF is large at lag 12 and lag 24 also lag 1 close to significance, indicating that $MA(1)_{12}$ model may exist. To fit SARIMA model based on ACF and PACF plots we will run different SARIMA models with Akaike Information Criteria (AIC) to select the best one.

Figure8: ACF plot and the PACF plot



AIC is given by $(-2 \log L + 2m)$ where $m = p+q+P+Q$ (P is the coefficient of Autoregressive seasonal and Q is the coefficient of Partial Autoregressive seasonal) and L is the likelihood function. The fit model that had lowest AIC among other models will be selected.

Table2: Value of AIC for several models

Model	p	q	P	Q	AIC
1	1	0	0	0	-172.991
2	1	0	0	0	-174.9629
3	1	0	1	0	-210.0458
4	1	0	1	0	-208.0551
5	1	1	0	0	-173.4959
6	1	1	0	0	-171.5227
7	1	0	0	1	-214.8936
8	1	0	1	1	-213.0122
9	1	0	1	1	-214.9832
10	1	1	1	1	-213.2193
11	0	0	1	0	-209.5837
12	0	0	1	1	-214.5785
13	1	0	0	1	-216.8478

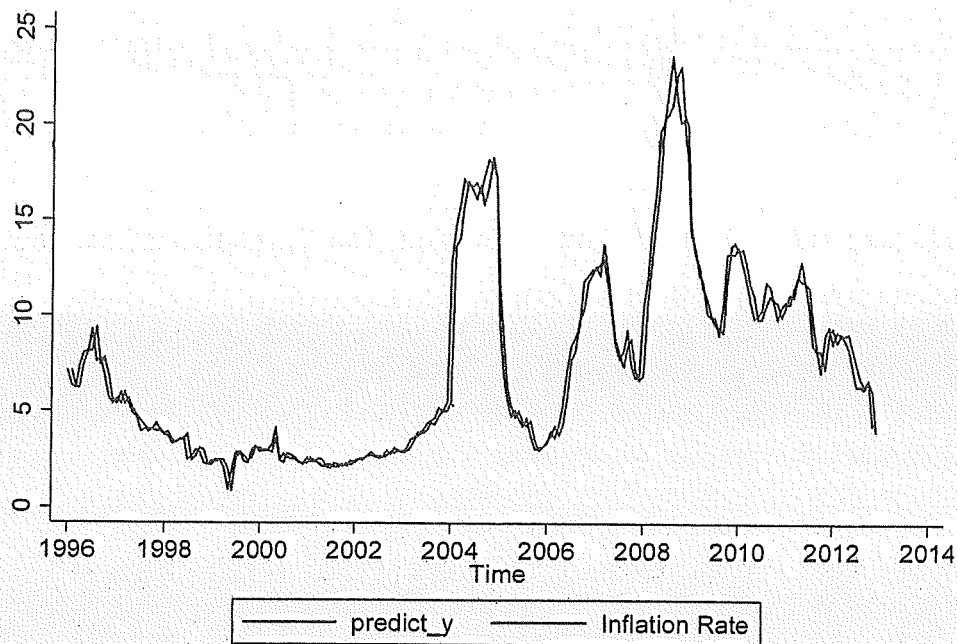
From table (2), the fit model for inflation rate in Egypt is

$$\text{SARIMA } (1,1,0) (0,0,1)_{12}$$

This can be written in the following form:

$$Y_t = 0.1089901 Y_{t-1} + 0.4937277 e_{t-12} \quad (5)$$

Figure 9: Origin Inflation Rate VS Predict Inflation



4.2. Linear Regression Analysis

Linear Regression Analysis can be used in all types of knowledge and application areas such as economic, social and environmental, where allows of assessing the impact of explanatory variables, which usually reflect the impact of regulations, policies and time

In this section, we estimate regression equation and compared it with SARIMA model where Random errors, the appropriate model using linear regression for sample data was third degree with log transfer, where the researcher built seven models (first, second and third degree) for original series and same degree for log transfer, addition to build nonlinear model (exponential regression) without significant difference fit indicator comparison linear models.

The indicator of fit model (F-test, R^2 , MSE) showed the fit model was log(inflation) with third degree in time.

Figure10: Scatter-diagram between Inflation and Time (third degree)

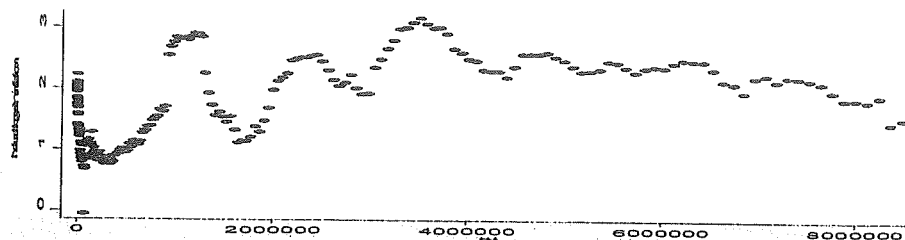


Figure11: Scatter-diagram between Inflation and Inflation at (first lag)

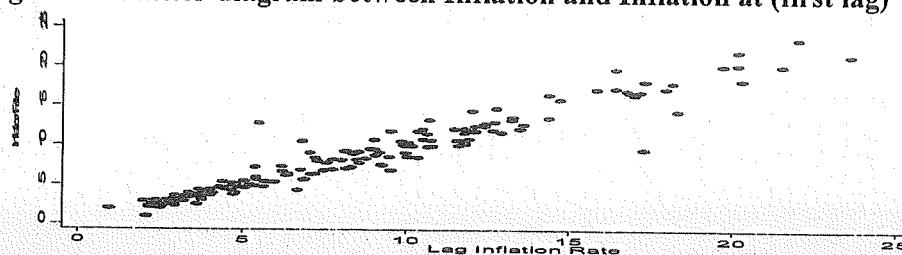
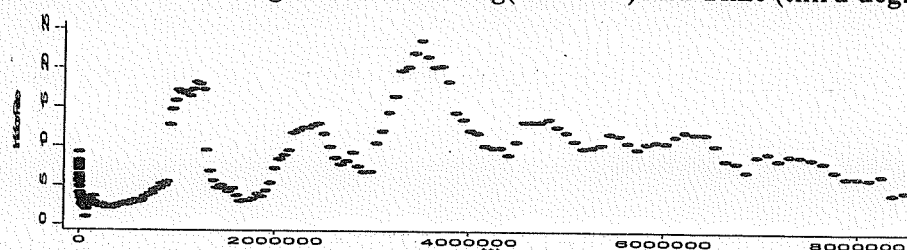


Figure12: Scatter-diagram between Log(Inflation) and Time (third degree)



. reg loginflation t tt ttt

Source	SS	df	MS
Model	67.8473534	3	22.6157845
Residual	29.8113464	200	.149056732
Total	97.6586998	203	.481077339

Number of obs = 204
 F(3, 200) = 151.73
 Prob > F = 0.0000
 R-squared = 0.6947
 Adj R-squared = 0.6902
 Root MSE = .38608

loginflation	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t	-.0596043	.0046416	-12.84	0.000	-.068757	-.0504515
tt	.0007764	.0000525	14.78	0.000	.0006728	.00088
ttt	-2.45e-06	1.69e-07	-14.54	0.000	-2.78e-06	-2.12e-06
_cons	2.284733	.1101421	20.74	0.000	2.067544	2.501922

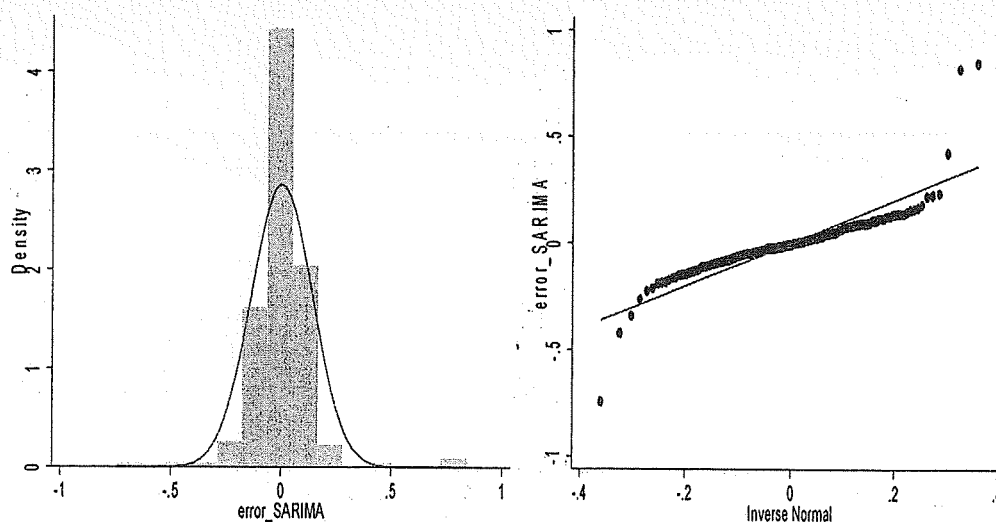
Table (4) shows indicator of error between the five linear regression models.

Table 4: Comparison between models linear regression

Model	Adj R ²	MSE	F-test	D-W	Sign Parameter ($\alpha=0.05$) From Total Parameter Model
Inflation= $\alpha+\beta_1$ Time+ β_2 Time ² + β_3 Time ³	0.535	11.878	78.98	.121	Sig(all)
Log(Inflation)= $\alpha+\beta_1$ Time	0.358	0.309	114.3	.081	Sig(all)
Log(Inflation)= $\alpha+\beta_1$ Time+ β_2 Time ²	0.366	0.305	59.58	.082	Sig(1 from 3)
Log(Inflation)= $\alpha+\beta_1$ Time+ β_2 Time ² + β_3 Time ³	0.69	0.149	151.73	.165	Sig(all)
Inflation= $\alpha+\beta_1$ Time(-1)	0.9443	1.43	3424.2	1.24	Sig(1 from 2)

5. Diagnostic checking

Figure13: Histogram of residual and Normal Q-Q plot

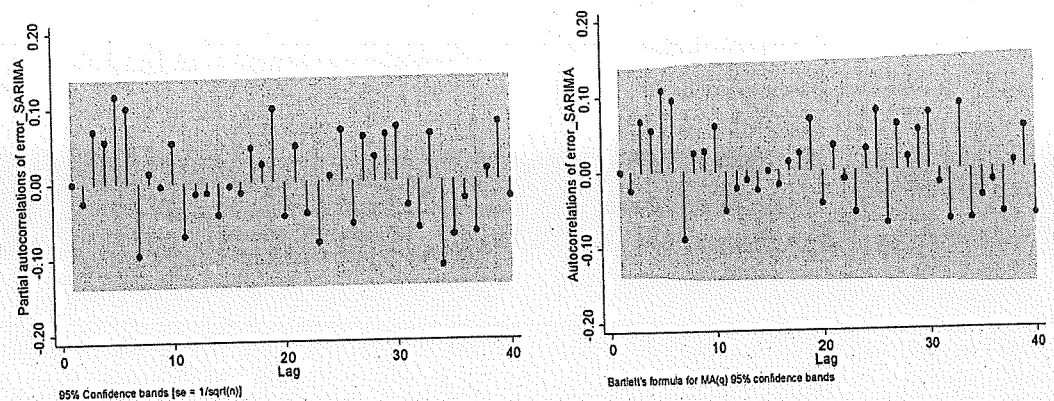


5.1.Methodology of Box and Jenkins

In this part run tests with the residuals for sure if the model of SARIMA built on sample data for log(Inflation) is actually a good fit. We examine the residuals for our SARIMA(1,1,0)(0,0,1)₁₂ model, if the model is appropriate, the residuals should distributed white noise model.

The histogram of the residuals appears look like distributed normally. The QQ plot of the residuals shows a fairly straight line, suggesting normality of the residuals. We can conclude that the residuals distributed white noise model.

Figure14: ACF plot and PACF plot of residuals



The ACF and PACF of the residuals equal approximately 0 at all lags if the model distributes white noise.

Figure15: ACF plot and PACF plot of residuals

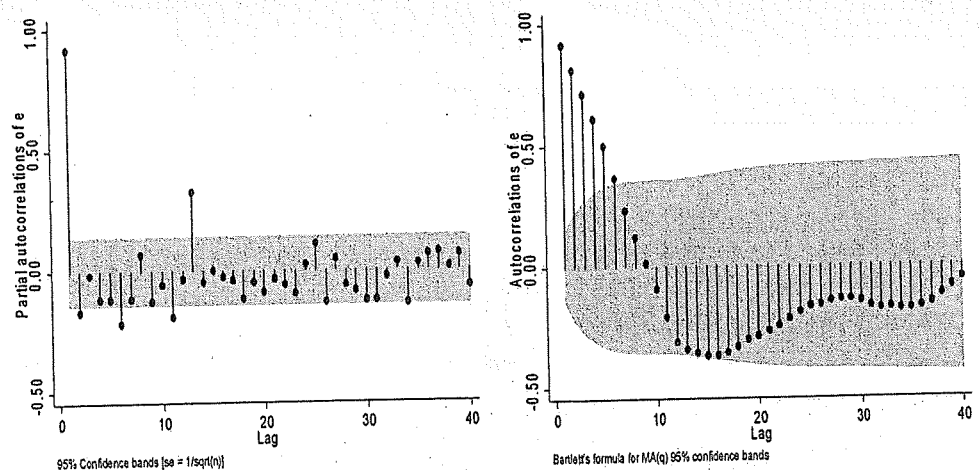


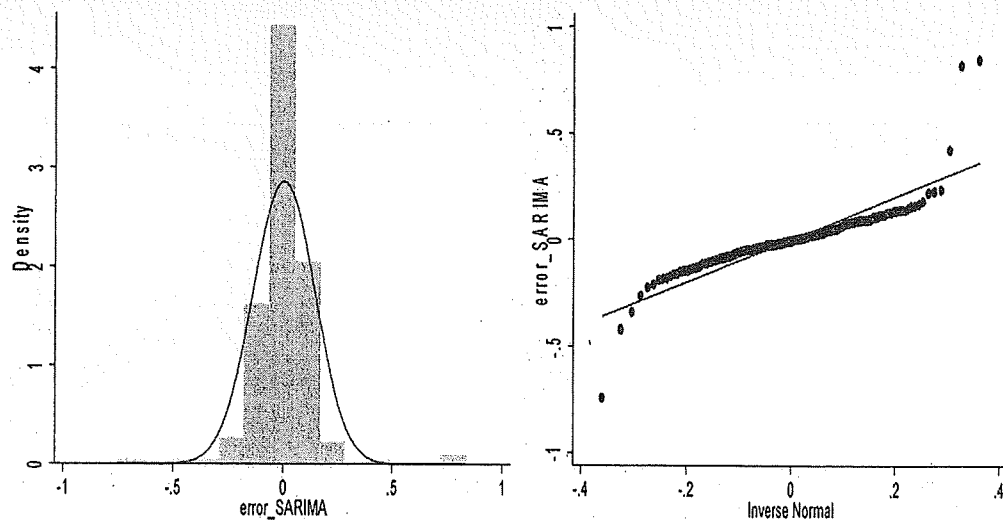
Table (4) shows indicator of error between the five linear regression models.

Table 4: Comparison between models linear regression

Model	Adj R ²	MSE	F-test	D-W	Sign Parameter ($\alpha=0.05$) From Total Parameter Model
Inflation= $\alpha+\beta_1\text{Time}+\beta_2\text{Time}^2+\beta_3\text{Time}^3$	0.535	11.878	78.98	.121	Sig(all)
Log(Inflation)= $\alpha+\beta_1\text{Time}$	0.358	0.309	114.3	.081	Sig(all)
Log(Inflation)= $\alpha+\beta_1\text{Time}+\beta_2\text{Time}^2$	0.366	0.305	59.58	.082	Sig(1 from 3)
Log(Inflation)= $\alpha+\beta_1\text{Time}+\beta_2\text{Time}^2+\beta_3\text{Time}^3$	0.69	0.149	151.73	.165	Sig(all)
Inflation= $\alpha+\beta_1\text{Time}(-1)$	0.9443	1.43	3424.2	1.24	Sig(1 from 2)

5. Diagnostic checking

Figure13: Histogram of residual and Normal Q-Q plot

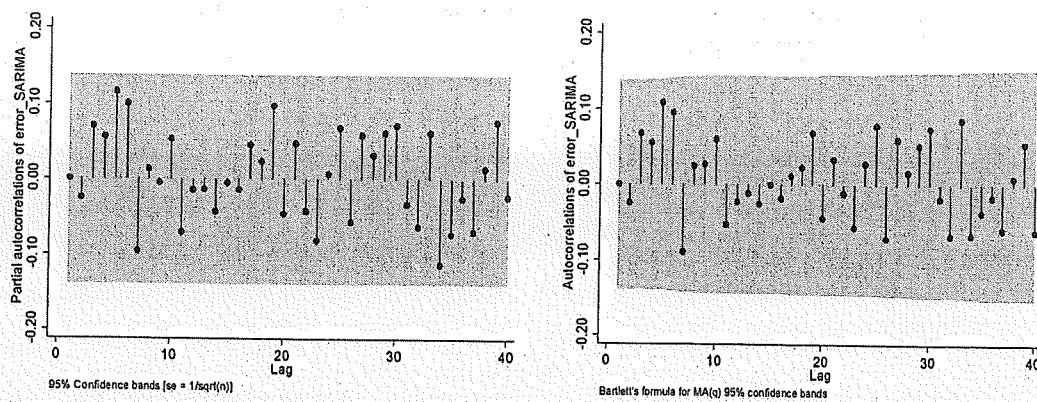


5.1.Methodology of Box and Jenkins

In this part run tests with the residuals for sure if the model of SARIMA built on sample data for log(Inflation) is actually a good fit. We examine the residuals for our SARIMA(1,1,0)(0,0,1)₁₂ model, if the model is appropriate, the residuals should distributed white noise model.

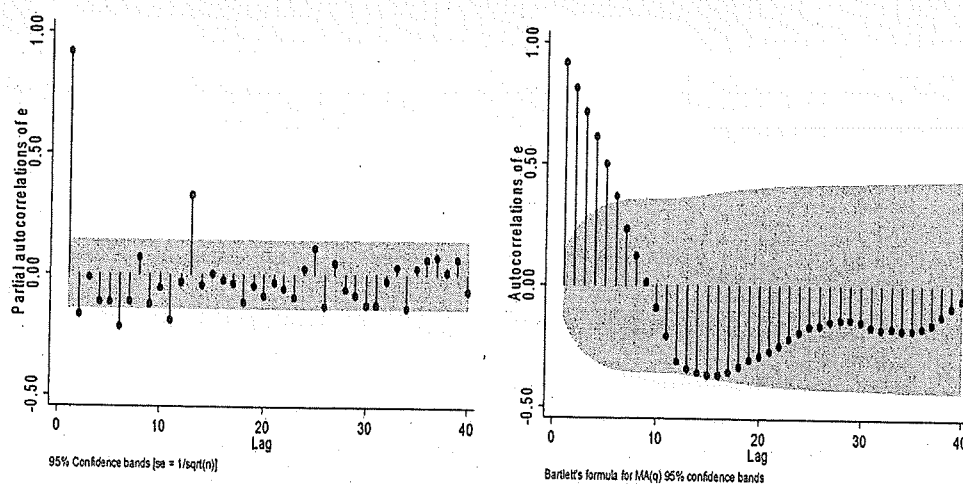
The histogram of the residuals appears look like distributed normally. The QQ plot of the residuals shows a fairly straight line, suggesting normality of the residuals. We can conclude that the residuals distributed white noise model.

Figure14: ACF plot and PACF plot of residuals



The ACF and PACF of the residuals equal approximately 0 at all lags if the model distributes white noise.

Figure15: ACF plot and PACF plot of residuals



5.2.Linear Regression Analysis

In this section we'll examine significant errors in the absence ($\rho \neq 0$) for $\log(\text{inflation})$ a function in time from first, second and third degree. There is no need for the rest of the tests determine the efficiency of the model in this case the Box- Jenkins methodology, and otherwise the rest of the tests will be conducted to reach the fit model.

It's clear from figure 14 and 15 the Box- Jenkins methodology more efficiency comparison with Linear Regression Analysis in time series analysis for inflation rate through period from January 1996 to December 2012.

6. Making forecasts

The fitting values of SARIMA model for inflation rate for first three month for 2013 year and some other mode were disused in paper such as the SARIMA model for square root inflation and model linear regression for log inflation from third degree were:

Month	SARIMA (1,1,0)(0,0,1) ₁₂ log (Inflation)	SARIMA (1,1,0)(0,0,1) ₁₂ Square.Root (Inflation)	Log(Inflation)= α + β_1 Time+ β_2 Time ² + β_3 Time ³	Actual value
January	4.926	5.242	4.889	6.318
February	4.814	5.146	4.646	8.258
March	4.803	5.109	4.409	7.586

From The predict inflation rate using SARIMA(1,1,0)(0,0,1)₁₂ after Natural logarithm transformation was found that far away as approximately 2% at least about the actual values for the rate of inflation. This can be explained insufficient the time analysis using SARIMA Models only to predict inflation,

which may indicate inflation is affected by economic variables more effective about time in a forecasting by the future path.

7. Conclusions

The paper studied inflation rate in Egypt to estimate a model to predict inflation values in the short run. Data from CAPMAS during the period from January 1996 to December 2012 has been used, where applied time series analysis. On the other hand, examined five linear regression models for modeling the sample data in order to ensure the efficiency of the final model to predict the inflation rate. The SARIMA $(1,1,0)(0,0,1)_{12}$ model using natural logarithm transformation is the fit model for this data.

The study found that a series of Egypt's inflation rate is nonstationary in mean during the study period, where required for stability take the first differences, also the series is nonstationary in variance therefore been using the natural logarithm transformation.

8. Recommendations

For more accurately in predict by inflation rate we can use other methodologies such as (ARMAX, VAR model), that reduce the difference was evident between predict and actual values.

From the high values of actual inflation about first three month in 2013 we can advise the government to reduce inflation by taking fiscal and monetary policies as:

A - Fiscal policy: Increasing taxes on luxury goods handled by a few residents of high-income earners and reducing government spending on luxury budget.

B - Monetary policy: Increasing the discount rate and increasing the legal reserve ratio to reduce ability banks in introduce Loans for citizens.

References

- [1] Anthony Young, "Time Series Analysis of Aluminum Prices and the London Metal Exchange Index", research paper.
"http://www.stat.ualberta.ca/~wiens/stat479/479project_ay.pdf"
- [2] Bulletin of indices of consumer prices (Dec 2012), the Central Agency for Public Mobilization and Statistics, Egypt, 2013.
- [3] Bulletin indices of consumer prices in 2009, Central Statistical Organization, Yemen, 2010.
- [4] Frederic S. Mishkin (1995), "The Economics of Money, Banking, and Financial Markets", Fourth Edition, HarperCollins College, United States of America.
- [5] Iron Dobre and Adriana Alexandru (2008), "Modeling unemployment rate using Box-Jenkins procedure", Journal of applied quantitative methods, 32, 156-166.
- [6] Massimiliano Marcellino (2008), "A Benchmark for Models of Growth and Inflation", Journal of Forecasting, 27, 305-340.
- [7] Michael H. Kutner; John Neter; Chris J. Nachtsheim and Christopher J. Nachtsheim (2005), "Applied Linear Statistical Models", McGraw-Hill companies, United States of America.
- [8] Nick T. Thomopoulos (1980), "Applied Forecasting Methods", Prentice-Hall, United States of America.
- [9] Richard Macklem and Stephen Lyman (2012), "Forecasting inflation rates using time series analysis", Queen's University, Canada.
- [10] Sherif Ahmed Maher (2012), "An Assessment the effectiveness of the inflation targeting policy on achieving the monetary policy goals in the Egyptian economy", Master Thesis, Faculty of Commerce, Department of Economics, Suez Canal University, Egypt.

[11] Stata Time-Series Reference Manual "Release12" (2011), College Station, United States of America.

مراجع باللغة العربية

- [12] الغنام، حامد بن عبدالله (٢٠٠٧)، "تحليل السلسلة الزمنية للتضخم في المملكة العربية السعودية"، جامعة الملك سعود، السعودية.
- [13] جيمس جوارتيني، ريجارد استروب- ترجمة عبدالفتاح عبدالرحمن، عبدالعظيم محمد (١٩٩٩)، "الاقتصاد الكلي: الاختيار العام والخاص"، دار المريخ، السعودية.
- [14] حاتم مهران (٢٠٠٧)، "التضخم في دول مجلس التعاون الخليجي ودور صناديق النفط في الاستقرار الاقتصادي"، سلسلة اوراق عمل، المعهد العربي للتخطيط، الكويت.
- [١٥] شعراوي، سمير مصطفى (٢٠٠٥)، "مقدمة في التحليل الحديث للسلاسل الزمنية"، مركز النشر العلمي بجامعة الملك عبدالعزيز، السعودية.
- [16] عيد، نيفين مختار (٢٠٠١)، "معدل التضخم الأمتل في الدول النامية"، رسالة ماجستير، كلية الاقتصاد والعلوم السياسية، جامعة القاهرة، مصر.
- [17] قطاع التحليل الاقتصادي بمركز المعلومات ودعم اتخاذ القرار (٢٠٠٤)، "أثر التغير في سعر الصرف على معدل التضخم في مصر"، مصر.
- [١٨] والتر فاندل -ترجمة: عبدالمرضي حاتم عزام وأحمد حسين هارون- (٢٠٠٥)، "السلاسل الزمنية ونماذج بوكس جنكز"، دار المريخ، السعودية.

Time Series: Identification and Modeling for Discrete Dynamic System

Zidan, A.I. *

Abstract

The paper takes advantage of the practical measurement from some price system and the time series theory to model and to analyze the system. Using the data dependent systems (DDS) to identification and modeling of easy stochastic dynamic systems, especially *The Egyptian Mobinil Stock Price*. Its purpose is to briefly characterize fundamental terms and equations of time series and to describe algorithm of a statistically adequate discrete model of stochastically discrete dynamic system. The paper discusses the stability and reversibility of the system. The Green function, the inverse function, and the Autocovariance function are also shown as characterize the ARMA (n, n-1) model. All the modeling and analysis procedures are realized by *DDS Toolbox-MATLABTM* software. The results show that Mobinil stock price system is stability and reversibility.

Key words

Data Dependent System (DDS), Green's Function, Time Series Analyses, System Identification, Modeling, ARMA (n, n-1) models, Criterion of Model Adequacy, System Stability, and Invertability.

1. Introduction

The goal of time series analysis is to develop suitable models and obtain accurate predictions for measurements over time of various kinds of phenomena. In statistics and econometrics, and in particular in time series analysis, an **autoregressive integrated moving average (ARIMA)** model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity.

The model is generally referred to as an ARIMA(p,d,q) model where parameters p , d , and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modelling.

In time series analysis, the **Box-Jenkins** methodology, named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasts.

*M.Sc. Student in ISSR - Department of Applied Statistics and Econometrics- Cairo university
Lecturer - Department of Basics Sciences - Modern Academy for Computer Since and Management
Technology - Cairo - Maadi

The original model uses an iterative three-stage modeling approach:

1. **Model identification and model selection:** making sure that the variables are stationary, identifying seasonality in the dependent series (seasonally differencing it if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which (if any) autoregressive or moving average component should be used in the model.
2. **Parameter estimation** using computation algorithms to arrive at coefficients which best fit the selected ARIMA model. The most common methods use maximum likelihood estimation or non-linear least-squares estimation.
3. **Model checking** by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time. (Plotting the mean and variance of residuals over time and performing a Ljung-Box test or plotting autocorrelation and partial autocorrelation of the residuals are helpful to identify misspecification.) If the estimation is inadequate, we have to return to step one and attempt to build a better model.

Data Dependent Systems (DDS) is a technique which through development of the difference/differential equations governing the system underlying the data directly without requiring any conjectures or guesses, Pandit and Shaw (1981), and Pandit and Wu (1983). For example suppose that an ARMA (5, 3) model is required for the data. Then we could have fitted as many as 27 models before reaching the adequate model. By using ARMA (n, n-1), whereas the 5 models in the system approach would automatically be checked by modeling procedure

The DDS simplify the two dimensions ARMA(n, m) to one dimension ARMA(n,n-1). S.M.Wu and S.M.Pandit (1983) pointed out each stable random sequence can be approximated to an required precision with an ARMA(n,n-1) model where n is large enough. According to this, the determination of can be dealt along the linear direction.

Time series Analysis: We can think of time series as being generated by a stochastic process, or the data generating process (DGP). A time series (sample) is a particular realization of the DGP (population). Time series analysis is the estimation of difference equations containing stochastic (error) terms.

A "solution" of a difference equation is a "**Function**" that reduces the equation to an identity. Solving a Difference Equation by Recursive Substitution or Dynamic Multipliers "**IMPLUSE RESPONSE FUNCTION**". The purpose of **DDS** is to introduce ARMA model in the form of stochastic linear difference equations to express this dependence. We will use the name **Green's function** to emphasize the basic theory of difference equations.

The real justification for the ARMA(n, n-1) strategy comes from treating the time series as the response of a dynamic system to random and independent or uncorrelated input, rather than trying to fit ARMA(n, m) models with unknown n and m to a series of dependent observations.

The DDS has been applied in many different fields in the Industrial and Management Systems, Engineering area, business, operations research, quality control, human factors, and ergonomics, for more details see: Amar (2010), Dong *et al.* (2003, 2007), Ermer and Notohardjono (1984),

Huggins (2012), Karanasos (2007), Keyvan (2001), Koustis *et al.* (2011), Leitner (2001, 2009), Leitner and Uriček (2010), Lu and Xin (2010), Nourbakhsh (2012), Oraby *et al.* (2005), Owen *et al.* (2001), Pandit (1981), Pandit and Wu (1983), Steudel *et al.* (1977), Tarawneh *et al.* (2012), Williams *et al.* (1994), Williams and Rajurkar (1989), and Zhu (2007).

2. Data Dependent Systems DDS.

The DDS methodology uses Non Linear least-square techniques to fit a series of difference/differential equations to the observed data until statistically adequate approximation is reached. The data is reduced to an uncorrelated residual series of white noise impulses that are free of any dynamics by the autoregressive moving average (ARMA) model function. The model itself contains the dynamics of the system and can be analyzed to find these dynamics and the corresponding properties of the system. No prior knowledge of the system is necessary for modeling it.

DDS reduces the correlated data to an uncorrelated or white noise process, which is the random component. The function which does the reduction is called the Green's function, Pandit (1981). The form of the DDS ARMA models easily yields the Green's function (impulse response function), and the autocovariance function Green's function in the form of a linear combination of exponentials and damped sinusoidal can be related to the physical phenomena generating the signal.

Pandit in 1981 found that the Green's function lends itself to more fundamental interpretation and analysis. Green's function can reduce the memory of a system to a single random disturbance. This method has been used to understand the system dynamics of a variety of systems, Rajurkar and Nissen (1985), Rajurkar *et al.* (1988), and Williams (1989).

3. Green's Function and the Orthogonal or Wold's Decomposition

The expansion of X_t with the help of Green's function, given by equation $x_t = \sum_{j=0}^{\infty} \phi_j^t a_{t-j}$ for an AR(1) system, can also be considered as an "Orthogonal Decomposition" of X_t as the sum of perpendicular vectors ($G_j a_{t,j}$) in an infinite dimensional space.

The concept of an "Orthogonal Decomposition" of stationary stochastic process or time series X_t into sum of uncorrelated random variables is very important, the expansion in equation (3-2). $x_t = \sum_{j=0}^{\infty} \phi_j^t a_{t-j}$ is known as Wold's decomposition it can be used to derive all statistical properties of the time series, [Bisgaard and Kulahci (2011), and Pandit and Wu (1983)]

Pandit and Wu (1983) pointed out, that each stable random sequence can be approximated to an required precision with an ARMA (n,n-1) model where n is large enough. According to this, the determination of ARMA (n, n-1) can be dealt along the linear direction. Namely, the DDS simplify the two dimensions (n, m) to one dimension. When talking about the value of n, it increases in twice. The principal idea of an autoregressive expression of a discrete stationary stochastic process is to express dependence of each immediate value of process X_t not as a function of time but as a function of former values.

Therefore, the simplest type of dependence will be the linear dependence of immediate values X_t on immediate previous values X_{t-1} is:

$$x_t = \phi_1 x_{t-1} + a_t \quad \dots\dots (1)$$

Which describes ϕ so called **autoregressive model of first order**, next signed **AR (1)**.

The measure ϕ of dependence from immediate values X_t proceeding ones can be determined using a linear least square procedure which minimizes the sum of deviation squares. This approach can be generalized on dependence of immediate X_t values on former n values, which can be described as autoregressive model of n -th orders - **AR (n)** as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots\dots + \phi_n x_{t-n} + a_t \quad \dots\dots (2)$$

The basic presumption of adequacy of **AR (n)** model is the independence of stochastic values a_t , which must form an independent series. If this presumption does not apply it means that a_t depends on a_{t-1} , a_{t-2} , a_{t-3} ... etc. The pure autoregressive models change to **Autoregressive Moving Average Models - ARMA**, generally of (n, m) order. With use of **ARMA** models one can express more complex types of internal dependencies, and as it will be shown further, their parameters have a very narrow dependence with the physical principal of followed processes .

A general type of **ARMA** dependence is a model of n -th order in an autoregressive part and $(n-1)$ -th order in moving average part – **ARMA (n, n-1)** described by :

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots\dots - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots\dots - \theta_{n-1} a_{t-n+1} \quad \dots\dots (3)$$

one can suppose, that residual deviations a_t are of normal distribution of probability with zero mean and dispersion of σ_a^2 , $a_t \sim NID(0, \sigma_a^2)$

The basic characteristics of **ARMA** models are **impulse response function** – so called **Greens function** which can express conditions of stability of models, and **inverse function** describing dynamics of models by expression if influence of former values of the process on the present ones . To express it more simply we can introduce a back shift operator **B** in general mode defined as:

$$B X_t = X_{t-1} \quad , \text{ or } B^j X_t = X_{t-j} \quad \dots\dots (4)$$

and using this the general **ARMA (n, n-1)** model form, from equation (3) gets the form:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots\dots - \phi_n B^n) X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots\dots - \theta_{n-1} B^{n-1}) a_t \quad \dots\dots (5)$$

Greens function G_j of difference equation (5) can be used to express values of X_t as a linear combination of stochastic values of a_t as:

$$x_t = \sum_{j=0}^{\infty} G_j a_{t-j} = \left[\sum_{j=-\infty}^t G_{t-j} B^j \right] a_j = (G_0 + G_1 + G_2 + \dots\dots) a_j \quad \dots\dots (6)$$

Similarly can be expressed the value of X_t as a linear combination of former values. Function of coefficients I_j in this expression is called inverse function "which is defined as:

$$X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + a_t \quad \text{..... (7)}$$

$$a_t = (1 - I_1 B - I_2 B^2 - \text{.....}) X_t \quad \text{..... (8)}$$

The second property function of the system is *inverse function*. It represents the current x_t by using the past response x_t . This function keeps some restraints on the parameters of MA model. That is the requirement of system's reversibility.

The condition of stability of ARMA (n, n-1) model is generally in form $|\lambda_k| < 1$, for $k = 1, 2, \dots, n$. Where λ_k are roots of characteristic equation on Left-hand side of equation (3) in form of:

$$\lambda^n - \phi_1 \lambda^{n-1} - \phi_2 \lambda^{n-2} - \text{.....} - \phi^n = 0 \quad \text{..... (9)}$$

Similarly we can defined a condition of inevitability as $|k_v| < 1$, for $k = 1, 2, \dots, n-1$. where v_k are roots of characteristic equation on right hand side of relationship (3) in form of:

$$v^{n-1} - \theta_1 v^{n-2} - \text{.....} - \theta_{n-1} = 0 \quad \text{..... (10)}$$

The third dynamic property of the system is *autocovariance function*. It represents the statistic character of dependence relationship among random variables x_t .

$$X_t, X_{t-1}, X_{t-2}, \dots \quad \text{.... (11)}$$

Just like Green's, Inverse function, and the theoretical autocovariance function completely characterizes an ARMA model. That is, given a model we can find any of these functions, and given a form of these functions we can find the ARMA model corresponding to it.

4. Algorithm of Statistically Adequate Discrete Model Determination

As it was indicated in the former section, the aim of identification is to determine order n of statistically adequate model ARMA (n, n-1), coefficients on left and right-hand side ($\phi_1, \phi_2, \dots, \phi_n, \theta_1, \theta_2, \dots, \theta_{n-1}$) and sum of squares of residual deviations $\sum a_i^2$ (or their dispersion σ_a^2). Because of the necessity of recurrent determination of deviations et from the start, the result is from the point of view of coefficients non-linear. So that it necessary to apply a nonlinear least square procedure looking in interactive steps for the minimum of sum of squares.

A special case of identification therefore forms pure autoregressive models of n -th orders AR(n) because for determination of its parameters a linear least squares procedure is suitable. Searched vector of coefficients:

$$\phi = [\phi_1 \quad \phi_2 \quad \dots \quad \phi_n]'$$

of AR(n) model one gets as a solution of matrix equation :

..... (12)

$$S.\phi = T$$

..... (13)

where: $S=X'X$ and $T=X'Y$

Where:

$$X = \begin{bmatrix} X_n & X_{n-1} & X_1 \\ X_{n+1} & X_n & X_2 \\ X_{N-1} & X_{N-2} & X_{N-n} \end{bmatrix}$$

..... (14)

$$Y = \begin{bmatrix} X_{n+1} \\ X_{n+2} \\ \vdots \\ X_N \end{bmatrix}$$

..... (15)

This gives the least squares estimated as: $\hat{\phi} = (X'X)^{-1}X'Y$

and:

$$\hat{\sigma}_a^2 = \frac{1}{(N-n)} \sum_{i=n+1}^N (X_i - \hat{\phi}_1 X_{i-1} - \hat{\phi}_2 X_{i-2} - \dots - \hat{\phi}_n X_{i-n})^2$$

$$\hat{\sigma}_a^2 = \sum a_i^2$$

..... (16)

Further steps in an identification procedure (determination of an optimum order n of AR (n) model) are similar to that of general ARMA (n, n-1) model.

5. Determination of starting guess of parameters

The main problem in solution of effective identification algorithm of ARMA (n, n-1) model is the guessing of $(\phi_1, \phi_2, \dots, \phi_n)^{(0)}$, $(\theta_1, \theta_2, \dots, \theta_{n-1})^{(0)}$ Values to ensure convergence of used iterative method and not too big a number of steps. The best of tested methods is the one based on expression of inverse function. These formulas are then linear for each type of ARMA (n, n-1) model.

Using general ARMA (n, n-1) in operator form (4) and for a_i 's giving such a definition of inverse function in form (8), comparing values by same powers of operators one gets system of equations as:

$$\begin{aligned}
 \phi_1 &= \theta_1 + I_1 \\
 \phi_2 &= \theta_2 - \theta_1 I_1 + I_2 \\
 \phi_3 &= \theta_3 - \theta_1 I_2 - \theta_2 I_1 + I_3 \\
 &\dots\dots\dots \\
 \phi_j &= \theta_j - \theta_1 I_{j-1} - \theta_2 I_{j-2} - \dots - \theta_{j-1} I_1 + I_j, \text{ for } k \geq j
 \end{aligned}
 \tag{17}$$

For all j with the assumption that $\theta_j = 0$ for $j > n-1$ and $\phi_j = 0$ for $j > n$ for the ARMA (n, n-1) model. In particular, then for $j > n$ is:

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_n B^n) I_j = 0 \tag{18}$$

I mean that parameters a_j and b_j one can get knowing values of inverse functions I_j . To solve the pure autoregressive model AR (p) can be used for a system of equations (13) is $I_j = 0$ for $j=1, 2, \dots, p$ and $I_j = 0$ for $(j > p)$.

6. Procedure of starting guess of parameters ϕ_j, θ_j of general ARMA (n, n-1) model is then as follows:

- a) Solutions of parameters a_j of an AR (p) model where $p=2n-1$ and values of inverse functions I_j simultaneously.
- b) Determination of moving average parameters θ_j by solving set of linear equations using equation (13) for $j = n+1, n+2, \dots, 2n-1$.
- c) Determination of autoregressive parameters ϕ_j using set of equations (13) and then parameters θ_j

7. Criterion of ARMA (n, n-1) model adequacy

To judge adequacy of chooses order n of ARMA (n, n-1) model (eventually AR (n) model) the procedure of statistical hypothesis testing was used.

The test of model adequacy is in a principle the test of statistical independence of deviations a_i . The standard F-test of statistical significance of two sample differences was chosen and this test was modified as a test of statistically significant decrease of the sum of squares. If one has to specify when by a regressive model with r parameters can be s of them counted as zero having N observations the test criterion becomes form as:

$$F = \frac{A_1 - A_0}{s} - \frac{A_0}{N - r} \sim F(s, N - r) \tag{19}$$

Where A_1 is sum of squares of higher order model, A_0 sum of squares of initial model. Resulting value of F is to be compared to value of F_{crit} found in the table of critical values of F - distribution for $(N - r)$ and s degrees of freedom and chosen level of probability (mostly 95 %). When we get that:

- a) $F > F_{crit}(S, N-r)_{0.95}$ then decrease the sum of squares during a change to higher order model was statistically significant, therefore, the initial model was not suitable.
- b) On the other hand, if $F < F_{crit}(S, N-r)_{0.95}$, the initial model was statistically adequate then an increase of its order does not make sense .

8. Algorithm of optimum model determination

Using former shown dependencies and formulas one can describe an optimum ARMA (n, n-1) model getting algorithm in words as follows.

1. Calculation of ARMA (2n, 2n-1) model parameters for n=1 and its sum of squares $A_0 = \sum a_i^2$
2. Increase of order $n \rightarrow (n+1)$ and calculation of model parameters and sum of squares A_1 .
3. Testing of statistically significance of sum of squares decrease $\Delta A = A_0 - A_1$. In case that decrease is statistically significant, go to previous step (2). In the other case the former model was statistical adequate.
4. Test of ϕ_{2n}, θ_{2n-1} parameters if their value is near zero or if their interval of confidence contains zero. When not ARMA (2n, 2n-1) model is suitable.
5. When ϕ_{2n}, θ_{2n-1} are zeros or near zero, calculation of ARMA (2n-1, 2n-2) model parameters.
6. Testing of moving average parameters θ_j of ARMA (2n-1, 2n-2) model. If some of them are near zero, create of ARMA (2n-1,m) model for $m < 2n-2$ and calculation of its parameters eventually of pure AR (2n-1) model.

9. Dissection and analysis the characters ARMA (n, n-1)

The model is:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = a_t - \theta_1 a_{t-1} \quad \dots (20)$$

Where: $\phi_1 = 0.7632$, $\phi_2 = 0.1810$, $\theta_1 = -0.3716$

Since the following relationships hold:

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad |\phi_2| < 1, \quad \text{and} \quad |\theta_1| < 1,$$

By definition of the system stability and inevitability, since the roots of ARMA (2, 1) are:

$$\lambda_1 = 0.9531, \lambda_2 = -0.1899, \nu_1 = -0.3716$$

The stability and inevitability conditions:

$$|\lambda_k| < 1 \quad , \quad |\nu_k| < 1 \quad \dots (21)$$

$$\lambda_1 = |0.9531| < 1, \lambda_2 = |-0.1899| < 1, \nu_1 = |-0.3716| < 1$$

One can conclude that this model is stable and irreversible which is a necessary condition to make appropriate conclusions about the model. The DDS model reduces the correlated data to an uncorrelated or white noise process. The function that does the reduction is called the Green's Function.

I Green's Function

The Green function is the first dynamic property of system. It expounds how the disturbance influences the system response. This function keeps some restraints on the parameters of AR model. That is the requirement of system's stability.

This function is the impulse response function of the ARMA model. The Green's Function, in the form of linear combination of exponentials and damped sinusoidal, can be related to physical phenomenon of the behavior of the **Egyptian Mobinil stock price**, and its model is given below.

$$G_j = g_1 \lambda_1^j + g_2 \lambda_2^j + \dots + g_n \lambda_n^j \quad \dots (22)$$

The project used the explicit method of Green's Function in calculus, and it is represented as:

$$G_j = g_1 \lambda_1^j + g_2 \lambda_2^j \quad \dots (23)$$

Since λ_1, λ_2 are *distinct roots*, then:

$$G_j = \left[\left(\frac{\lambda_1 - \theta}{\lambda_1 - \lambda_2} \right) \cdot \lambda_1^j + \left(\frac{\lambda_2 - \theta}{\lambda_2 - \lambda_1} \right) \cdot \lambda_2^j \right] \quad \dots (24)$$

$$g_1 = \left(\frac{\lambda_1 - \theta}{\lambda_1 - \lambda_2} \right), g_2 = \left(\frac{\lambda_2 - \theta}{\lambda_2 - \lambda_1} \right) \quad \dots (25)$$

That is,

$$\lambda_1 + \lambda_2 = \phi_1, \lambda_1 \lambda_2 = -\phi_2, \text{ or: } \lambda_1, \lambda_2 = \frac{1}{2}(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}) \quad \dots (26)$$

Where the result is:

$$G_j = 1.1589 * (0.9531)^j - 0.1589 * (-0.1899)^j$$

The λ_1 and λ_2 are the characteristic roots of the second order linear difference equation and for this model they are real roots. A system is stable when Green's Function is asymptotically unbiased in an exponential way.

One of the characteristic roots, $\lambda_1 = 0.9531$, is very close to unity and thus exhibits a nearly random price component. It should be noted that the DDS estimates the value of root λ_1 correctly instead of using a differencing operation before modeling. This strategy not only provides a precise representation of the data but also makes it possible to evaluate the contribution of the

stock price as a component in the development of the Egyptian Mobinil stock price in marketing.

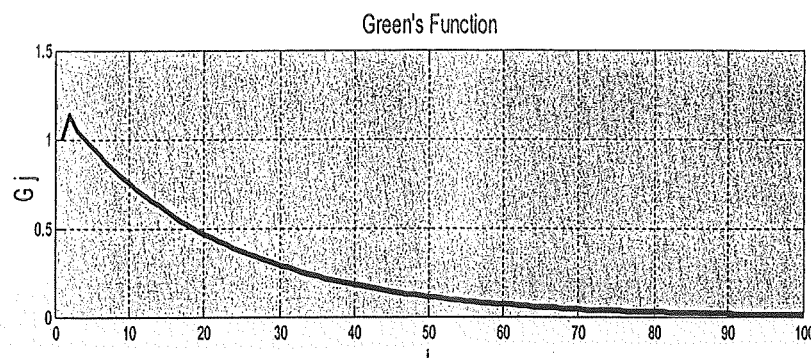


Figure 1: Green's Function

The Green's Function as system dynamics can be interpreted in two ways. The first is indicator of how well the shocks a_{t-j} are remembered by the system j time units back for $j = 0, 1, 2, \dots, t$. The second is how quickly the system will return to its equilibrium. The plot of the impulse response or G_j for the ARMA (2, 1) model of the price system is shown in Figure (1). Note that the figure shows the value $G_0 = 1$ and trend of $G_j \rightarrow 0$ when $j \rightarrow \infty$, [Amar (2010), Huggins (2012), and Nourbakhsh (2012)].

The Green's function returns to zero quickly. The figure demonstrates that the system returns to balance rapidly after *transient oscillation*. The Mobinil stock price System is stable. The figure (1) shows the Mobinil stock price rate impulse response which has a slow decay, and it exhibits a strong memory behavior.

The function of the line appears to be asymptotically stable. By definition whether the function curve approaches a finite value, it is called stable, and if this value is trending to zero we could conclude that the curve is asymptotically stable.

The equation:

$$G_j = 1.1589 * (0.9531)^j - 0.1589 * (-0.1899)^j$$

Reveals that the Mobinil stock price rate consists of two parts:

- **The first part**, contributed by the root $\lambda_1 = 0.9764$, effectively denotes slow decay, high inertia, or strong memory. **The second part**, which comes from the root $\lambda_2 = -0.1899$, expresses a dependent or correlated noise component.
- The root λ_2 (the correlated noise component) dominates the short-term response behavior of the Mobinil stock price before the correlation time of 247 days, whereas the root λ_1 , (the random walk component) determines the long-term response behavior.

II Inverse Function:

The second property function of the system is inverse function. It represents the current x_t by

$$\text{using the past response } x_t, \text{ i.e. } X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + a_t \quad \dots\dots (27)$$

The coefficient I_j in the equation is inverse function. This function keeps some restraints on the parameters of MA model. That is the requirement of system's reversibility Seen from the Figure (2).

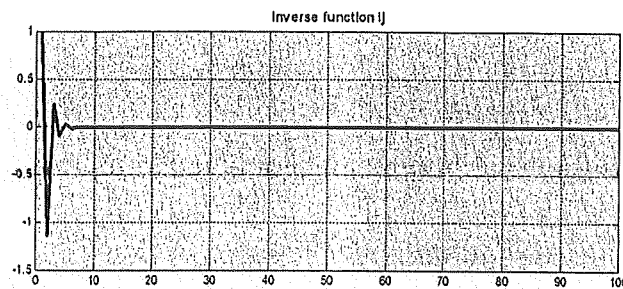


Figure 2: The Inverse Function

Seen from the Figure (2), the inverse function returns to zero quickly. The figure demonstrates that the system dynamic response attenuate quickly. So the system returns to balance rapidly. The **Mobinil stock price** System is reversible.

III Autocovariance Function

The third dynamic property of the system is autocovariance function Fig (3). It represents the statistic character of dependence relationship among random variables x_t .

$$\gamma_k = d_1 * (\lambda_1)^k + d_2 * (\lambda_2)^k + \dots\dots + d_n * (\lambda_n)^k \quad \dots\dots (28)$$

In this case:

$$\gamma_k = d_1 * (\lambda_1)^k + d_2 * (\lambda_2)^k \quad \dots\dots (29)$$

The weight of each AR roots in model variance: $d_1 = 181.2392$, $d_2 = -1.6195$ Then:

$$\gamma_k = 181.2392 * (0.9531)^k + -1.6195 * (-0.1899)^k, \text{ The variance:}$$

$$\gamma_0 = d_1 + d_2 = 179.6197$$

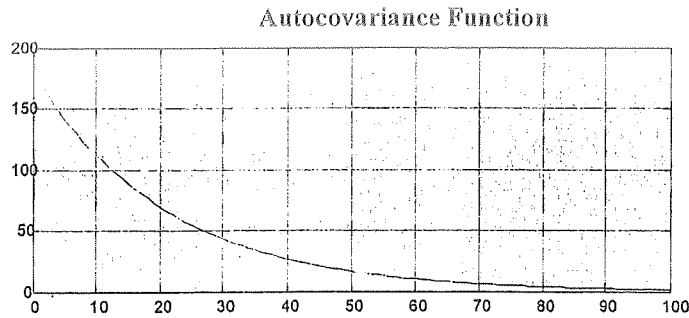


Figure (3): Autocovariance Function

Autocovariance function returns to zero quickly. The figure demonstrates that the system returns to balance rapidly after *transient oscillation*. Just as the Green function and the inverse function, the autocovariance function can completely characterize the ARMA(n,n-1) model.

10. Response Generation Using Green's Function.

A measure, which can be used to determine the system dynamic characteristics and behaviors, represented by the response amplitude X_t , is obtained by evaluating the system's response as a reaction to regular disturbances a_t . A dynamically stable system with no sudden variability is expected to behave in such a way that its output response amplitude X_t is with a similar pattern to the excited force. Deviation from such a state is usually attributed to some external effects such as **Mobinil stock price**, to a change in the system's dynamic features. To extract the tool's dynamic behavior under progressive **Mobinil stock price**, the system response X_t of the ARMA (2, 1) can

be generated according to relation:

$$x_t = \sum_{j=0}^{\infty} G_j a_{t-j} = \sum_{j=-\infty}^t G_{t,j} a_j \quad \dots\dots (30)$$

The result show that the response generation of disturbances using Green's function for ARMA (2, 1) with $\phi_1 = 0.7632$, $\phi_2 = 0.1810$, $\theta_1 = -0.3716$, become smooth as the positive values of ϕ 's ,and for large positive values there is a tendency towered trends and long runs.

11. Forecasting

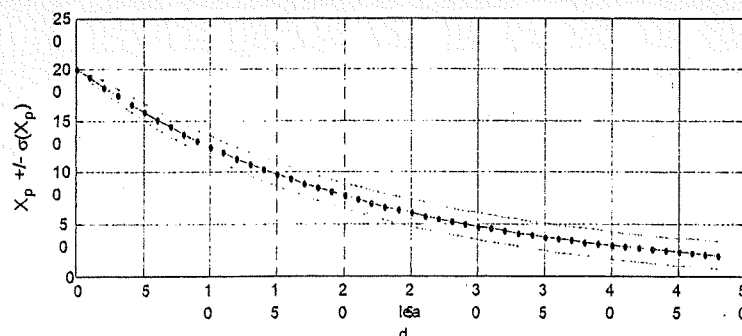


Figure 4: 25-Step ahead forecast

From the figure (4), the middle line is the predicted value; the upper and lower lines are the 95% confidence interval.

We can know that the predicted value has smooth and closed from the limits of confidence interval. Clear from the figure 6.8 that all the prediction values fall within the forecast period for 25 - ahead step forecasting, as that not widen in prediction period, [Bisgaard and Kulahci (2011), Box and Jenkins (1970), and Rajurkar and Nissen (1985)].

This refers to the appropriate model to predict accurately. The Function appears that high values of predict and then gradually decline prediction. The prediction within the sample also appears that close to the prediction values. By using Data Dependent Systems (DDS), the result arrived at the suitable model to be fitted. The results indicate that in the near future, by implementing the model arrived at in this thesis, there will not be any increase in the level of significance in Mobinil stock price.

12. CONCLUSIONS

An efficient algorithm for identification and modeling of Mobinil stock price System has been presented in this paper. The algorithm is based on DDS. As an improvement, concerning the obvious tendency in the system data, we extract the tendency before modeling. In doing so, the calculation complexity and the order of the model are decreased. It also shortens the modeling time.

The modeling research demonstrated that this model was stable, reversible and rational.. At the same time, because MATLABTM is good at calculation, we realize all the procedure and analysis by it.

By the modeling and discussing of the Egyptian Mobinil stock price System, we can draw the conclusions that the algorithm can assess identification of Mobinil stock price System quickly, accurately, easily and efficiently, it can completely realize the pre-assignment.

REFERENCES

1. Amar, M.R.S. (2010), *Estimation of Mechanical Properties of Soft Tissue Subjected to Dynamic Impact*, Industrial and Management Systems Engineering, University of Nebraska, Lincoln, Dissertations and Student research.
2. Bisgaard, S., Kulahci, M. (2011), *Time Series Analysis and Forecasting by Example*, John Wiley & Sons, New Jersey.
3. Box, G.E.P., Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
4. Dong Y., Liu, S., Wei, Z. (2003), Analyses and modeling of time series based on MATLAB, *computer engineering*, Vol. 9.

5. Dong, Y., Zhou, X., Shen, T. (2007), The Application of Time Series Analysis In Precision Measurement of Airborne Fire Control System, *International Symposium on Test and Measurement*, Vol. 3, pp. 1827-1830.
6. Ermer, D.S., Notohardjono, B.D. (1984), Fatigue failure identification by time series model, *Engineering Fracture Mechanics*, Vol.20, issue 5-6, pp.705-718.
7. Huggins, A. (2012), A Model for Optimizing the Allocation of Ethanol Production in the U.S., Department of Industrial and Management Systems Engineering, University of Nebraska–Lincoln, Dissertations and Student research.
8. Karanasos, M. (2007), The Correlation Structure of Some Financial Time Series Models, *Quantitative and Qualitative Analysis in Social Sciences*, Vol. 1, Issue 2, pp.71-87.
9. Keyvan, S. (2001), Traditional signal pattern recognition versus artificial neural networks for nuclear plant diagnostics, *Progress in Nuclear Energy*, vol. 39, Issue 1, pp. 1-29.
10. Koustis, G., Hadjipaschalis, I., Poulliskkas, A. (2011), An Overview of Load Demand and Price Forecasting Methodologies. *International Journal of Energy and Environment*, Vol. 2, Issue 1, pp. 123-150.
11. Leitner, B. (2001), Stochastic Time Series and Possibilities of Their Use for selected Modal Characteristics Determination of Mechanical Structures. *Scientific Papers of University of Žilina*.
12. Leitner, B. (2009), Data Dependent System and Their Use in Dynamic System Identification and Modeling, Academic journal: *Mechanics Transport Communications*, issue 3.
13. Leitner, B., Uriček, J. (2010), A Method for Adaptive Identification of Stochastically Loaded parts of Mechanical Systems. *Proceedings of the 3.rd international multi-conference on engineering and technological innovation IMETI*, Orlando, Florida, USA, Vol. II, pp. 174-179.
14. Lu, C., Xin, Z. (2010), *Impulse-Response function Analysis: An application to macroeconomy data of China*, School of Economics and Social Sciences, Höskolan Dalarna, D-Level Essay in Statistics for M.S. Degree.
15. Nourbakhsh, F. (2012), *Machining Stability of Wire EDM of Titanium* Industrial and Management Systems Engineering, University of Nebraska, Lincoln, Dissertations and Student research.
16. Oraby, S.E., Al-Modhuf, A.F. , Hayhurst, D.R. (2005), A Diagnostic Approach for Turning Tool Based on the Dynamic Force Signals, *Journal of Manufacturing Science and Engineering*, Vol. 127 ,Issue 3, pp. 463-475.
17. Owen, J.S., Eccles, B.J. , Choo, B.S., Woodings, M.A. (2001), The application of autoregressive time series modelling for the time-frequency analysis of civil engineering structures, *Engineering Structures*, Vol. 23, Issue 5, pp. 521–536.

18. Pandit, S.M., Shaw, M.C. (1981), Characteristic Shapes and Wavelength Decomposition of Surfaces in Machining, *CIRP Annals - Manufacturing Technology*, Vol. 30, Issue 1, pp. 487-492.
19. Pandit, S.M., Wu, S.M. (1983), *Time Series and System Analysis with Applications*, Wiley
20. Rajurkar, K.P., Nissen, J.L. (1985), Data Dependent Systems Approach to Short-Term Load Forecasting. *Systems, Man and Cybernetics, IEEE Transactions on*, Vol. SMC-15, Issue 4, pp. 532-536.
21. Rajurkar, K.P., Schnacker, C.L., Lindsay, R.P. (1988), Some Aspects of ECM Performance and Control, *CIRP Annals - Manufacturing Technology*, Vol. 37, Issue 1, pp.183-186.
22. Steudel, H.J., Pandit, S.M., Wu, S.M. (1977), A Multiple Time Series Approach to Modeling the Manufacturing Job-Shop As a Network of Queues, *Management Science*, Vol. 24, Issue 4, pp. 456-463.
23. Steudel, H.J., Wu, S.M. (1977), A Time Series Approach to Queuing Systems with Applications for Modeling Job-Shop In-Process Inventories, *Management Science*, Vol. 7, Issue 7, pp. 745-755.
24. Tarawneh, I.S., Williams, R.E., Bishu, R.R. (2012), Data Dependent System (DDS) Technique: Can it Be Used in Ergonomics? *Proceeding of the Human Factors and Ergonomics Society 3rd Annual Meeting*.
25. Williams, R.E. (1989), *Metal Removal and Surface Finish Characteristics of Abrasive Flow Machining*, University of Nebraska, Lincoln.
26. Williams, R.E. (1989), *Metal Removal and Surface Finish Characteristics of Abrasive Flow Machining*, University of Nebraska, Lincoln.
27. Williams, R.E., Bishu, R.R., Rajurkar, K.P. (1994), *Application of Data Dependent Systems (DDS) Methodology to the Analysis of Lifting Styles*, Unpublished Department Paper, Industrial and Management Systems Engineering Department, University of Nebraska-Lincoln, Nebraska.
28. Zhu, J. (2007), Detection of cylinder pressure in diesel engines using cylinder head vibration and time series methods, *Journal of Marine Science and Application*, Vol.6, Issue4, pp.8-12.

Calibration Estimation in Stratified Sampling

El-Sheikh A.A*, Abdel-Rahman H. Mohamed**

Abstract

Calibration is commonly used in survey sampling to include auxiliary information to increase the precision of the estimates of population parameters as total and/or means. In this paper, the most commonly distances that introduced in the literatures will be reviewed. Also, we newly propose calibration estimators in stratified random sample for equal and proportion allocation by using three suggested distances

Keywords: Auxiliary Information – Calibration - Distance Measures -.Stratified Sampling.

1- Introduction

Calibration has become a widely used procedure for estimation in sample survey. Deville and Särndal (1992) presented calibration estimators in survey. It used auxiliary information to produce efficient estimates. Calibration requires that the population totals for one or more auxiliary variables have been known. The efficiency of the calibration estimator depends on how well auxiliary variables explain the variability of y (The variable of interest).

There are four major advantages of calibration estimation in survey sampling:.

(1) Calibration as a linear weighting method. The fixation on weighting methods on the part of the leading national statistical agencies was a powerful driving force behind calibration.

* Assoc.Prof. of Applied Statistics and Econometrics, I.S.S.R, Cairo University

** Ph.D Student of Applied Statistics and Econometrics, I.S.S.R, Cairo University

(2) Calibration as a systematic way to use auxiliary information. Calibration provided a systematic way to take auxiliary information into account.

(3) Calibration to achieve consistency. The calibration equations imposed consistency on the weight system, so that, when applied to the auxiliary variables, it will confirm (be consistent with) known aggregates for those same auxiliary variables.

(4) Calibration for convenience. **Harms and Duchesne (2006)** pointed out, "The calibration approach has gained popularity in real applications because the resulting estimates are easy to interpret and to motivate, relying, as they do, on design weights and natural calibration constraints." Calibration on known totals strikes the typical user as transparent and natural.

In this paper, new calibration estimator for improving variance estimator with the aid of auxiliary information in stratified random sampling for different scheme will be introduced. The paper is organized as follows: Section 2 reviews the calibration method in survey sampling. Section 3 suggests calibration estimation for stratified sampling. We introduced a new calibration estimator using five distance measures in stratified sampling with the aid of auxiliary information and proposed the estimator of variance of calibration estimator for each case.

2. Literature Review of Calibration Estimation in Sampling Survey

Wu (1999), provided a unified framework to deal with the effective use of complete auxiliary information from survey data at the estimation stage. The proposed method involved modeling the relationship between the variable of interest and the auxiliary variables, and then incorporating the auxiliary information into design-consistent estimators of finite population means, totals, distribution functions and quantiles through the predicted values using calibration and empirical likelihood methods. Estevao and Särndal (2000), noted that distance minimization is not the only possible starting point for calibration. They defined and developed an alternative, the functional form approach. Also, in (2002) they showed that there are exactly nine different subsets of the complete information, for a total ten different cases of auxiliary information. Kim, et al (2006), proposed various calibration approach ratio estimators and derived the estimator of the variance of the calibration approach ratio-type estimators in stratified sampling for four estimators to the use of complete auxiliary information to estimate ratio estimator and derived the estimator of variance of calibration approach ratio estimators.

2.1 General Set-Up

Suppose that the finite population consists of N identifiable units. Associated with the i^{th} unit are, the study variable, y_i , and a vector of auxiliary variables, x_i . The values $x_1, x_2, \dots, x_N$ are known for the entire population but y_i is known only if the i^{th} unit is selected in the sample, s . A sample s of size n is drawn without replacement according to a probabilistic sampling plan with inclusion probabilities $\pi_i = \text{pr}(i \in s)$.

2.2 Review of the Calibration Estimator

Let, d_i be the design-weight associated with unit i , $d_i = \frac{1}{\pi_i}$. The theory of calibration estimators is developed in Deville and Särndal (1992). They proposed a general method of deriving so called calibration estimators by first choosing a distance measure Φ_s between the basic design weights d_i and the revised calibration weights w_i and then minimizing this distance subject to specified benchmark constraints.

The most commonly used distance measure is the chi-square distance,

$$\Phi_s = \sum_{i=1}^n \frac{(w_i - d_i)^2}{2d_i q_i} \quad i = 1, 2, \dots, n \quad (2-1)$$

Where the q_i 's are known positive weights unrelated to d_i . Note that $w_i = d_i = 0$ for $i \notin s$. The uniform weights $q_i = 1$ are used in most applications. A distance measure Φ_s between the weights d_i and the "calibrated" weights w_i has to satisfy the following criteria:

1. $\Phi_s(w_i, d_i) \geq 0$; Φ_s is differentiable with respect to w_i ; strictly convex; and $\Phi_s(d_i, d_i) = 0$.

2. $\frac{\partial \Phi_s}{\partial w_i} = \Phi(w_i, d_i)$ is continuous and one-to-one.

The calibration estimator of y is constructed as

$$\hat{y}_c = \sum_{i=1}^n w_i y_i, \quad (2-2)$$

where the calibration weights, w_i , are chosen to minimize Φ_s subject to the constraint

$$\sum_{i=1}^n w_i x_i = \sum_{i=1}^N x_i \quad (2-3)$$

For the chi-square distance, the resulting calibration estimator is

$$\hat{y}_C = \sum_{i=1}^n w_i y_i = \hat{y}_{HT} + \left(\sum_{i=1}^N x_i - \hat{x}_{HT} \right)^T \hat{B} \quad (2-4)$$

Where $\hat{y}_{HT}$ the Horvitz-Thompson estimator for total y , $\hat{x}_{HT} = \sum_{i=1}^n d_i x_i$

The Horvitz-Thompson estimator for total x , $\hat{B} = [T]^{-1} \sum_{i=1}^n d_i q_i x_i^T y_i$.

where, $T = \sum_{i=1}^n d_i q_i x_i^T x_i$ and assuming that the inverse of T exist.

Proof:

Consider the Lagrange function:

$$L(\Phi_s; w_i) = \sum_{i=1}^n \frac{(w_i - d_i)^2}{2d_i q_i} - \lambda^T \left(\sum_{i=1}^n w_i x_i - \sum_{i=1}^N x_i \right)$$

So,

$$\frac{\partial L}{\partial w_i} = \frac{2(w_i - d_i)}{2d_i q_i} - \lambda^T x_i$$

By equating the partial derivative to 0, it yields

$$w_i = \lambda^T x_i d_i q_i + d_i \quad (2-5)$$

By substituting equation (2-5) in (2-3) to get,

$$\lambda^T = \left[\sum_{i=1}^N x_i - \sum_{i=1}^n d_i x_i \right]^T \left[\sum_{i=1}^n d_i q_i x_i^T x_i \right]^{-1} \quad (2-6)$$

By substituting equation (2-6) in (2-5) to get,

$$w_i = d_i + \left[\sum_{i=1}^N x_i - \sum_{i=1}^n d_i x_i \right]^T \left[\sum_{i=1}^n d_i q_i x_i^T x_i \right]^{-1} d_i q_i x_i \quad (2-7)$$

By substituting equation (2-7) in (2-2) to get,

$$\hat{y}_C = \sum_{i=1}^n d_i y_i + \left[\sum_{i=1}^N x_i - \sum_{i=1}^n d_i x_i \right]^T \left[\sum_{i=1}^n d_i q_i x_i^T x_i \right]^{-1} \sum_{i=1}^n d_i q_i x_i y_i$$

3- Calibration Estimation for Stratified Sampling

Suppose the population consists of H strata with N_h units in the h^{th} stratum from which a simple random sample of size n_h is taken without replacement. Let, total population size be $N = \sum_{h=1}^H N_h$ and sample size be $n = \sum_{h=1}^H n_h$, respectively.

Associated with the i^{th} unit of the h^{th} stratum there are two values y_{hi} and x_{hi} with $x_{hi} > 0$ being the covariate which the i^{th} unit selected from the h^{th} stratum, where $i = 1, 2, \dots, n_h$. For the h^{th} stratum, let $W_h = N_h / N$ be the stratum weights, $N = \sum_{h=1}^H N_h$. $X = \sum_{h=1}^H W_h X_h$ and $Y = \sum_{h=1}^H W_h Y_h$, $f = n_h / N_h$ the sample fraction, Assume x_{hi} denote the value of the i^{th} unit of the auxiliary variable in the h^{th} stratum about which information known at the unit level or at the stratum level.

The main aim of this section is to derive the calibration estimator for the parameters of interest with the case of stratified simple random sampling when only X is known for two kind of stratified sample, equal and propotion allocation.

3-1 The Calibration Estimator for stratified sample for generalized distance

A- Equal Allocation

The calibration estimator for the above sampling scheme is defined as

$$\hat{y}_{CST.EA} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-1)$$

Where w_{hi} are the calibration weights for stratified sample for the i^{th} unit in the h^{th} stratum, which are chosen to minimize Φ_s and the subscript of $\hat{y}$ is $CST.EA$ (calibration for stratified sample in equal allocation).

The most commonly used distance measure is the chi-square distance,

$$\Phi_s = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi}q_{hi}} \quad (3-2)$$

Where the q_{hi} s are known positive weights unrelated to d_{hi} . The uniform weights $q_{hi} = 1$ are used in most applications. Minimize (3-2) subject to the constraint

$$\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} \quad (3-3)$$

Proof:

Consider the Lagrange function

$$L(\Phi_s; w_i) = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi}q_{hi}} - \lambda^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} - \sum_{i=1}^{N_h} x_{hi} \right)$$

So,

$$\frac{\partial L}{\partial w_h} = \frac{2W_h(w_{hi} - d_{hi})}{2d_{hi}q_{hi}} - \lambda^T W_h x_{hi}$$

Equating the partial derivative to 0 yields,

$$w_{hi} = d_{hi} (1 + \lambda^T x_{hi} q_{hi}) \quad (3-4)$$

By substituting equation (3-4) in (3-3) to get,

$$\lambda^T = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1}$$

Substituting by λ^T results

$$w_{hi} = [d_{hi} + \left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1} d_{hi} q_{hi} x_{hi}] \quad (3-5)$$

Then substituting equation (3-5) in (3-1) to get,

$$\hat{y}_{CST} = \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} y_{hi} + \left[\left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T y_{hi} \right]$$

B- Proportional Allocation

According to proportional allocation, the sample is allocated according to

$$f = n/N. \text{ This implies that } \frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n}{N} = f$$

Since using simple random sampling in each stratum the probability of any sampling unit in stratum h being included in the subsample n_h is $\frac{n_h}{N_h} = f$ for all strata, then

$$n_h = \frac{N_h}{N} n, \text{ this simply shows that } n \text{ should be allocated according to } N_h/N.$$

The calibration estimator is defined as

$$\hat{y}_{CST.PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-6)$$

Where $\hat{y}_{CST.PA}$ is calibration estimator for stratified sample in case of proportional allocation.

The most commonly used distance measure is the chi-square distance,

$$\Phi_s = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi}q_{hi}} \quad (3-7)$$

Minimize (3-7) subject to the constraint

$$\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} \quad (3-8)$$

Proof:

Consider the Lagrange function

$$L(\Phi_s; w_{hi}) = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi}q_{hi}} - \lambda^T \left(\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} \right)$$

So,

$$\frac{\partial L}{\partial w_{hi}} = \frac{2w_{hi}(w_{hi} - d_{hi})}{2d_{hi}q_{hi}} - \lambda^T w_{hi} x_{hi}$$

Equating the partial derivative to 0 yields,

$$w_{hi} = d_{hi} (1 + \lambda^T x_{hi} q_{hi}) \quad (3-9)$$

By substituting equation (3-9) in (3-8) to get,

$$\lambda^T = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1}$$

Substituting by λ^T results

$$w_{hi} = [d_{hi} + \left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1} d_{hi} q_{hi} x_{hi}] \quad (3-10)$$

Then substituting equation (3-10) in (3-6) to get,

$$\hat{y}_{CST,PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} y_{hi} + \left[\left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^T \left(\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T x_{hi} \right)^{-1} \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} x_{hi}^T y_{hi} \right]$$

3-2 Minimum entropy distance

A- Equal Allocation

The calibration estimation is

$$\hat{y}_{CST} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-11)$$

By using minimum entropy distance

$$\Phi_s = \sum_{h=1}^H \sum_{i=1}^{n_h} (-d_{hi} \log(\frac{w_{hi}}{d_{hi}}) + w_{hi} - d_{hi}) / q_{hi} \quad (3-12)$$

and calibration condition

$$\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} \quad (3-13)$$

By using (3-11), (3-12) and (3-13) calibration estimator of minimum entropy distance

can be obtained as,

$$\hat{y}_{CST,EA} = \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} y_{hi} + \left[\left(\sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right) \left[\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right]^{-1} \left[\sum_{h=1}^H \sum_{i=1}^{n_h} q_{hi} x_{hi} \right]^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} y_{hi} \right]$$

B- Proportion Allocation

The calibration estimator is defined as

$$\hat{y}_{CST.PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-14)$$

By using minimum entropy distance

$$\Phi_s = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} (-d_{hi} \log(\frac{w_{hi}}{d_{hi}}) + w_{hi} - d_{hi}) / q_{hi} \quad (3-15)$$

and calibration condition

$$\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} x_{hi} \quad (3-16)$$

Using (3-14), (3-15) and (3-16) calibration estimator of minimum entropy distance can be obtained as,

$$\hat{y}_{CST.PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} y_{hi} + \left[\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right] \left[\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right]^{-1} \left[\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi} \right] \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} y_{hi}$$

3-3 New Distance (1)

A- Equal Allocation

Calibration estimator using new distance 1 will be derived as,

$$\hat{y}_{CST.EA} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-17)$$

The new distance 1 measure is,

$$\Phi_s = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi} q_{hi} \rho s_{hi}^2} \quad (3-18)$$

Where ρ is the correlation between x_{hi}, y_{hi} , it takes the value $\rho=1$. Or $\rho < 1$ or $\rho > 1$, we used $\rho=1$. s_{hi}^2 is the variance between auxiliary variable For the h^{th} stratum. For every distance measure there is a corresponding set of calibrated weights and a calibration estimator. Minimize (3-18) subject to the constraint

$$\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} \quad (3-19)$$

Proof:

By using equations (3-18) and (3-19), the Lagrange function will be,

$$L(\Phi_s; w_i) = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi} q_{hi} \rho s_{hi}^2} - \lambda^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} - X \right)$$

So,

$$\frac{\partial L}{\partial w_h} = \frac{2W_h(w_{hi} - d_{hi})}{2d_{hi}q_{hi}\rho s_{hi}^2} - \lambda^T W_h x_{hi}$$

Equating the partial derivative to 0, it yields

$$w_{hi} = d_{hi} (1 + \lambda^T x_{hi} q_{hi} \rho s_{hi}^2) \quad (3-20)$$

By substituting equation (3-20) in (3-19) to get,

$$\lambda_{hi}^T = (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^T (\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho s_{hi}^2 x_{hi}^T x_{hi})^{-1} \quad (3-21)$$

By substituting equation (3-21) in (3-20) to get,

$$w_{hi} = d_{hi} [1 + (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^T (\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho s_{hi}^2 x_{hi}^T x_{hi})^{-1} x_{hi} q_{hi} \rho s_{hi}^2] \quad (3-22)$$

Then substituting equation (3-22) in (3-17) to get,

$$\hat{y}_{CST,PA} = \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} y_{hi} + (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^T (\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho s_{hi}^2 x_{hi}^T x_{hi})^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} q_{hi} \rho s_{hi}^2 y_{hi}$$

B- Proportional Allocation

The calibration estimator is defined as

$$\hat{y}_{CST} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-23)$$

The new distance 1 measure is,

$$\Phi_s = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2d_{hi}q_{hi}\rho s_{hi}^2} \quad (3-24)$$

Minimize (3-24) subject to the constraint (3.19) and by using equation (3-23) calibration estimator for new distance 1 can be derived as,

$$\hat{y}_{CST,PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} y_{hi} + (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi})^T (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho s_{hi}^2 x_{hi}^T x_{hi})^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} x_{hi} q_{hi} \rho s_{hi}^2 y_{hi}$$

3-4 New distance (2)

A-Equal Allocation

The calibration estimator for equal allocation is defined as

$$\hat{y}_{CST} = \sum_{h=1}^H \sum_{i=1}^{nh} w_{hi} y_{hi} \quad (3-25)$$

By using new distance (2)

$$\Phi_s = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{2(\sqrt{w_{hi}} - \sqrt{d_{hi}})^2}{q_{hi} \rho s_{hi}^2} \quad (3-26)$$

and calibration condition

$$\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} \quad (3-27)$$

Calibration estimator will be derived by minimize (3-26) subject to the constraint (3-27), considering the Lagrange function

$$L(\Phi_s, w_i) = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{2(\sqrt{w_{hi}} - \sqrt{d_{hi}})^2}{q_{hi} \rho s_{hi}^2} - 2\lambda^T \left(\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} \right)$$

So,

$$\frac{\partial L}{\partial w_{hi}} = 2 \left(1 - \frac{\sqrt{d_{hi}}}{\sqrt{w_{hi}}} \right) \frac{1}{q_{hi} \rho s_{hi}^2} - \lambda^T x_{hi}$$

Equating the partial derivative to 0, it yields

$$w_{hi} = d_{hi} \left(1 - \frac{\lambda^T x_{hi} q_{hi} \rho s_{hi}^2}{2} \right)^{-2} \quad (3-28)$$

By substituting equation (3-28) in (3-27) to get,

$$\lambda' = - \left(\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^{-1} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \frac{2 d_{hi} x_{hi}}{q_{hi} \rho s_{hi}^2 x_{hi}} \right) \quad (3-29)$$

By substituting equation (3-29) in (3-28) to get,

$$w_{hi} = d_{hi} \left(1 / \left(1 + q_{hi} \rho s_{hi}^2 x_{hi} \left(\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^{-1} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \frac{2 d_{hi} x_{hi}}{q_{hi} \rho s_{hi}^2 x_{hi}} \right) \right) \right)^{-2} \quad (3-30)$$

Substituting by (3-30) in (3-25) to get,

$$\hat{y}_{CALEA} = \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} y_{hi} \left[1 / \left(1 + \left(\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi} \right)^{-1} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \frac{2 d_{hi} x_{hi}}{q_{hi} \rho s_{hi}^2 x_{hi}} \right) \sum_{h=1}^H \sum_{i=1}^{n_h} q_{hi} \rho s_{hi}^2 x_{hi}^2 \right) \right]$$

B- Proportion Allocation

The calibration estimator is defined as

$$\hat{y}_{CST} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-31)$$

By using new distance 2

$$\Phi_s = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{2(\sqrt{w_{hi}} - \sqrt{d_{hi}})^2}{q_{hi} \rho s_{hi}^2} \quad (3-32)$$

and calibration condition

$$\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{N_h} x_{hi} \quad (3-33)$$

Using (3-31), (3-32) and (3-33) . The calibration estimator will be:

$$\hat{y}_{CALPA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} y_{hi} [1 / (1 + (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi})^{-1} (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{2d_{hi} x_{hi}}{q_{hi} \rho_h s_{hi}^2} (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} q_{hi} \rho_h s_{hi}^2 x_{hi}^2)))]$$

3-5 New Distance (3)

A- Equal Allocation

The calibration estimator is defined as

$$\hat{y}_{CST.EA} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-34)$$

By using New distance 3

$$\Phi_s = \sum_{h=1}^H \sum_{i=1}^{n_h} (w_{hi} - d_{hi})^2 / 2 w_{hi} q_{hi} \rho s_{hi}^2 \quad (3-35)$$

and calibration condition

$$\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} \quad (3-36)$$

we deriving calibration estimator by minimize (3-35) subject to the constraint (3-36)

Proof:

if consider the Lagrange function

$$L(\Phi_s; w_{hi}) = \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{(w_{hi} - d_{hi})^2}{2 w_{hi} q_{hi} \rho s_{hi}^2} - \lambda^T [\sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi}]$$

With partial derivative $\partial L / \partial w_{hi}$

$$\frac{\partial L}{\partial w_{hi}} = \frac{1 - \frac{d_{hi}^2}{w_{hi}^2}}{2 q_{hi} \rho s_{hi}^2} - \lambda^T x_{hi}$$

Equating the partial derivative to 0, it yields

$$w_{hi} = d_{hi} [1 - 2 q_{hi} \rho s_{hi}^2 \lambda^T x_{hi}]^{-1/2} \quad (3-37)$$

By substituting equation (3-37) in (3-36) to get,

$$\lambda' = -(\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^{-1} (\frac{\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi}}{\sum_{h=1}^H \sum_{i=1}^{n_h} q_{hi} \rho_h s_{hi}^2 x_{hi}}) \quad (3-38)$$

Substituting by (3-38) in (3-37) leading to

$$w_{hi} = d_{hi} / (1 + q_{hi} \rho_h s_{hi}^2 x_{hi} (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^{-1} (\frac{\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi}}{\sum_{h=1}^H \sum_{i=1}^{n_h} q_{hi} \rho_h s_{hi}^2 x_{hi}})) \quad (3-39)$$

Substituting by (3-39) in (3-34) leading to

$$\hat{y}_{CAL EA} = \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} y_{hi} / [(1 + (\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho_h s_{hi}^2 x_{hi}) (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi})^{-1} (\frac{\sum_{h=1}^H \sum_{i=1}^{n_h} d_{hi} x_{hi}}{\sum_{h=1}^H \sum_{i=1}^{n_h} q_{hi} \rho_h s_{hi}^2 x_{hi}}))]$$

B- Proportion Allocation

The calibration estimator is defined as

$$\hat{y}_{CST .PA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} y_{hi} \quad (3-40)$$

By using new distance 3

$$\Phi_s = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} (w_{hi} - d_{hi})^2 / 2 w_{hi} q_{hi} \rho_h s_{hi}^2 \quad (3-41)$$

and calibration condition

$$\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{N_h} x_{hi} \quad (3-42)$$

Using (3-40),(3-41) and (3-42)The calibration estimator is

$$\hat{y}_{CALPA} = \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} y_{hi} / [(1 + (\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} q_{hi} \rho_h s_{hi}^2 x_{hi}) (\sum_{h=1}^H \sum_{i=1}^{N_h} x_{hi} - \sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi})^{-1} (\frac{\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} d_{hi} x_{hi}}{\sum_{h=1}^H \frac{N_h}{N} \sum_{i=1}^{n_h} q_{hi} \rho_h s_{hi}^2 x_{hi}}))]$$

References

- 1- Abel-Rahman H. Mohamed (2013), The Calibration Approach in Sampling Technique. PhD. Unpublished doctoral dissertation, ISSR -Cairo-Univ. Egypt
- 2- Alain F .Zuur, Elena N. Ieno and Erik Meesters (2009), A Beginners Guide to R. Springer Verlag. New York
- 3- Cassel, C., Särndal, C. and Wretman, J. H. (1977). Foundations of Inference in survey Sampling. John Wiley and Sons, Inc. New York.
- 4- Cochran, W.G. (1977). Sampling Techniques, 3rd ed., John Wiley and Sons, Inc. New York.
- 5- Estevao, V.M., and Särndal, C.-E. (2000). A Functional Approach to Calibration. Journal of Official Statistics, 16, 379-399.
- 6- Estevao, V.M., and Särndal, C.E. (2002). The Ten Cases of Auxiliary Information for Calibration in Two-phase Sampling. Journal of Official Statistics, 18, 233-255.
- 7- Fuller, W.A. (2002). Regression estimation for survey samples. Survey Methodology, 28, 5-23.
- 8- Fuller, W.A. (2009). Sampling Statistics. John Wiley & Sons, Inc. Hoboken, New Jersey.
- 9- Harms, T. and Duchesne, P. (2006). On Calibration Estimation for Quantiles . Survey Methodology, 32, 37-52.
- 10- Hartely, H. O. and Rao, J. N. K. (1968). A new Estimation Theory for Sample Surveys. Biometrika , 55, 547-57.
- 11- Jessen, R.J (1969). Some Methods of Probability Non-replacement Sampling. JASA, 64, 175-193.
- 12- Kim, J.K. (2009). Calibration Estimation using Empirical Likelihood in Survey Sampling. Statistica Sinica, 19, 145-157.
- 13- Kim, J.K. and Park, M. (2010). Calibration Estimation in Survey Sampling. International Statistical Review, 78(1), 21-39
- 14- Kim, J.M. Sungura, E.A and Heo, T.Y.(2007). Calibration Approach Estimators in Stratified Sampling. Statistics and Probability Letters 77, 99-103.
- 15- Kish, L. (1965). Survey Sampling. Wiley- Interscience Publication: New York.
- 16- Kott, P. (2003). A practical Use for Instrumental Variable Calibration. Journal of Official Statistics, 19(3), 265-272.

- 17-Maria L. Rizzo (2008), Statistical Computing with R, Chapman and Hall/CRC
- 18- Särndal, C.E Swensson, B and Wretman, J.H (1992). Model assisted survey sampling.: Springer-Verlag. New York
- 19-Särndal, C.E.(1996). Efficient Estimators with Simple Variance in Unequal probability sampling. JASA, 91, 1289-1300.
- 20- Särndal, C.E (2007). The Calibration Approach in Survey Theory and Practice. Survey Methodology 33, 99-119.
- 21-Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the general regression estimator: Higher level calibration approach. Survey Methodology, 24, 41-50.
- 22-Wu, C. (1999). "The Effective Use of Complete Auxiliary Information From Survey Data," Unpublished doctoral dissertation, Simon Fraser University, Canada.
- 23- Wu, C. (2003). Optimal Calibration Estimators in Survey Sampling. Biometrika, 90, 937-951.
- 24- Wu, C. and Luan, Y. (2003). Optimal Calibration Estimators Under Two-phase Sampling. Journal of official statistics, 19, 119-131.
- 25- Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. JASA, 96, 185-193
- 26- Wu, C., and Sitter, R.R. (2001). Variance estimation for the finite population distribution function with complete auxiliary information. Canadian Journal of Statistics, 29, 289-307

المراجع:

١- نشرة القوى العاملة السنوية لعامى ٢٠١٠ - ٢٠١١ - الجهاز المركزى للتعبئة العامة والإحصاء.

- 2- Soliman,A (1995), Some aspects of unemployment in Egypt in 1988, with prospects of future unemployment up to 1997,Cairo Demogriphic Center(CDC).
- 3- Mohamed,H (2007), Labor Force and Unemployment In The Sultanate of Oman, Cairo Demogriphic Center(CDC).
- 4- Ahmed,E (2003) , *Egyptian Labor Force Dimensions and Changing Structure During (1979-1996)* ,Cairo Demogriphic Center(CDC).

تفسير معاملات المتغيرات المعنوية:

- مفاضلة أن يكون الفرد مشغول وحاصل على مؤهل أقل من المتوسط هو (٠,٦٢٧)
- مفاضلة الفرد المشغول والذي يكون حالته التعليمية أسمى.
- مفاضلة أن يكون الفرد مشغول وحاصل على مؤهل متوسط أو فوق المتوسط هو (٠,٥١٣)
- مفاضلة الفرد المشغول والذي يكون حالته التعليمية أسمى.
- مفاضلة أن يكون الفرد مشغول وحاصل على مؤهل جامعى فأعلى هو (٠,٢٤٧) مفاضلة الفرد المشغول والذي يكون حالته التعليمية أسمى.
- مفاضلة أن يكون الفرد مشغول ومتزوج هو (٢,٣٢٥) مرة مفاضلة الفرد المشغول والذي يكون حالته الزوجية لم يتزوج.
- مفاضلة أن يكون الفرد مشغول ومطلق / أرمل هو (١,٤٨٧) مرة مفاضلة الفرد المشغول والذي يكون حالته الزوجية لم يتزوج.
- مفاضلة أن يكون الفرد مشغول وعمره فى الفئة العمرية أكثر من ٢٩ سنة هو (٣,٠٥٣) مرة مفاضلة الفرد المشغول والذي يكون عمره فى الفئة العمرية (١٥ - ٢٩ سنة) .

من خلال التحليل يتبين لنا أنه كلما زاد المستوى التعليمى لدى الفرد كلما قلت فرصة حصوله على وظيفة وأن الفرد الذى يقع فى الفئة العمرية (أكثر من ٢٩ سنة) تكون فرصته أكبر بكثير من الفرد الذى يقع فى الفئة العمرية (١٥ - ٢٩ سنة) ، وهذا يبين لنا أن مشكلة البطالة تتركز فى الأفراد حملة المؤهلات والأفراد فى الفئة العمرية (١٥ - ٢٩ سنة) والأفراد اللذين لم يتلقوا التدريب. وبالمقارنة بنتائج عام ٢٠١٠ يتضح لنا إرتفاع فرص حصول الفرد الحاصل على مستوى تعليمى عالى سواء كان مؤهل فوق متوسط أو مؤهل جامعى وهذا بسبب تعيين أوائل الجامعات نتيجة للقرارات التى إتخذت من قبل الدولة بعد الإحداث التى شهدتها مصر خلال هذا العام كما نلاحظ أيضاً إنخفاض فرصة الفرد الذى يقع فى الفئة العمرية (أكثر من ٢٩ سنة) وذلك لأن فئة أوائل الخريجين والذي كان لها النصيب الأكبر فى الحصول على فرصة عمل تقع فى الفئة العمرية (١٥ - ٢٩ سنة) .

تفسير معاملات المتغيرات المعنوية:

- مفاضلة أن يكون الفرد مشتغل وحاصل على مؤهل أقل من المتوسط هو (٠,٥٠٢)
- مفاضلة الفرد المشتغل والذي يكون حالته التعليمية أمة.
- مفاضلة أن يكون الفرد مشتغل وحاصل على مؤهل متوسط أو فوق المتوسط هو (٠,٠٩٥) مفاضلة الفرد المشتغل والذي يكون حالته التعليمية أمة.
- مفاضلة أن يكون الفرد مشتغل وحاصل على مؤهل جامعي فأعلى هو (٠,٠٤٨) مفاضلة الفرد المشتغل والذي يكون حالته التعليمية أمة.
- مفاضلة أن يكون الفرد مشتغل ومتزوج هو (٢,٢٥٧) مرة مفاضلة الفرد المشتغل والذي يكون حالته الزوجية لم يتزوج.
- مفاضلة أن يكون الفرد مشتغل وعمره في الفئة العمرية أكثر من ٢٩ سنة هو (٥,٣١٣) مرة مفاضلة الفرد المشتغل والذي يكون عمره في الفئة العمرية (١٥ - ٢٩ سنة) .
- مفاضلة أن يكون الفرد مشتغل ولم يتلقى التدريب هو (٠,٥٦١) مفاضلة الفرد المشتغل والذي تلقى التدريب .

من خلال التفسير يتبين لنا أن كلما زاد المستوى التعليمي لدى الفرد كلما قلت فرصة حصوله على وظيفة وأن الفرد الذي يقع في الفئة العمرية (أكثر من ٢٩ سنة) تكون فرصته أكبر بكثير من الفرد الذي يقع في الفئة العمرية (١٥ - ٢٤ سنة) بكثير وأيضاً الفرد الذي يتلقى تدريب ما تكون فرصته أكبر من الفرد الذي لم يتلقى التدريب، وهذا يبين لنا أن مشكلة البطالة تتركز في الأفراد حملة المؤهلات والأفراد في الفئة العمرية (١٥ - ٢٩ سنة) والأفراد اللذين لم يتلقوا التدريب.

• تحليل نتائج الانحدار خلال عام ٢٠١١ طبقاً لإجمالي الجمهورية:

من جدول التصنيف للمتغير التابع (موقف الفرد من العمل) نجد أن نسبة التصنيف الصحيح باستخدام هذا النموذج في التقدير ٨٨,١% من جملة المشاهدات .

المتغيرات داخل النموذج

المتغيرات المستقلة	B	S.E.	Sig.	Exp(B)
الحالة التعليمية				
يقراً ويكتب	-.٤٦٦	.٠٥٩	.٠٠٠	.٦٢٧
مؤهل أقل من المتوسط	-.٦٦٨	.٠٤٩	.٠٠٠	.٥١٣
مؤهل متوسط وفوق المتوسط	-١,٣٩٩	.٠٣٨	.٠٠٠	.٢٤٧
مؤهل جامعي فأعلى	-١,٧٦٢	.٠٤٠	.٠٠٠	.١٧٢
الحالة الزوجية				
متزوج	.٨٤٤	.٠٢٥	.٠٠٠	٢,٣٢٥
مطلق / أرمل	.٣٩٧	.٠٦٨	.٠٠٠	١,٤٨٧
العمر أكثر من ٢٩ سنة	١,١١٦	.٠٢٦	.٠٠٠	٣,٠٥٣
Constant	٢,١٥٢	.٠٣٨	.٠٠٠	٨,٦٠٢

P: هي احتمال أن يكون الفرد مشتغل.

٥,٤ دراسة العوامل المؤثرة على التدفقات فى سوق العمل المصرى خلال الفترة (٢٠١٠ - ٢٠١١):

فى إطار تحليل البيانات تم إستخدام نموذج الانحدار اللوجستى الثنائى (نظراً لأن المتغيرات التابعة تتكون من فئتين) بالاعتماد على طريقة forward stepwise لأختيار علاقة المتغير التابع وهو الحالة العملية للفرد وبين المتغيرات المستقلة وهى : الحالة التعليمية ، الحالة الزوجية ، العمر ، التدريب. (سليمان ١٩٩٥).

وعند إجراء التحليل تم إستخدام الفئة الأولى كفئة مرجعية فى كافة المتغيرات ، فالنسبة للمتغيرات كانت الفئة المرجعية (أمى - لم يتزوج أبداً - (١٥ - ٢٩ سنة) - تلقى التدريب).

نتائج الانحدار اللوجيستى المقدّر:

فيما يلى إستعراض نتائج الانحدار اللوجيستى المقدّر للمتغير التابع وعلاقته بالمتغيرات المستقلة (الحالة التعليمية - الحالة الزوجية - العمر - التدريب)، حيث يأخذ المتغير التابع (موقف الفرد من العمل) القيمة (١) إذا كان الفرد مشغول ، ويأخذ القيمة (٠) إذا كان الفرد متعطّل.

النموذج اللوجيستى:

$$\text{Log} (p/1-p) = B_0 + B_1x_1 + \dots + B_px_p$$

- تحليل نتائج الانحدار خلال الفترة من ٢٠١٠ - ٢٠١١ طبقاً لإجمالى الجمهورية:
- تحليل نتائج الانحدار خلال عام ٢٠١٠ طبقاً لإجمالى الجمهورية:

من جدول التصنيف للمتغير التابع (موقف الفرد من العمل) نجد أن نسبة التصنيف الصحيح بإستخدام هذا النموذج فى التقدير ٩١,٢% من جملة المشاهدات .

المتغيرات داخل النموذج

المتغيرات المستقلة	B	S.E.	Sig.	Exp(B)
الحالة التعليمية				
يقرأ ويكتب	.٠٦٣	.١٣١	.٦٣٢	١,٠٦٥
مؤهل أقل من المتوسط	-.٦٨٩	.٠٩٤	.٠٠٠	.٥٠٢
مؤهل متوسط وفوق المتوسط	-٢,٣٥٤	.٠٧٣	.٠٠٠	.٠٩٥
مؤهل جامعى فاعلى	-٣,٠٢٦	.٠٧٣	.٠٠٠	.٠٤٨
الحالة الزوجية				
متزوج	.٨١٤	.٠٢٩	.٠٠٠	٢,٢٥٧
مطلق / أرمل	.٠٥٤	.٠٩٥	.٥٦٩	١,٠٥٦
العمر أكثر من ٢٩ سنة	١,٦٧٠	.٠٣٤	.٠٠٠	٥,٣١٣
الفرد لم يتلقى التدريب	-.٥٧٨	.٢٢٤	.٠١٠	.٥٦١
Constant	٣,٩٣٥	.٢٣٥	.٠٠٠	٥١,١٤٠

P: هى إحتمال أن يكون الفرد مشغول.

جدول رقم (١٧) مصفوفة الإنتقال للمشتغلين طبقاً لنوع القطاع خلال الربعين الأول والثاني لعام ٢٠١١

القطاع للربع الثاني ٢٠١١ %						القطاع للربع الأول ٢٠١١
الإجمالي	أخرى وتشمل التعاونيات	استثماري	خاص	عام	حكومي	
١٠٠	٠,٢٥	١,٧٤	٦١,٢٢	٤,٠٢	٣٢,٧٨	حكومي
١٠٠	٠,٥٤	٣,٥٠	٥٧,٣١	٨,٧٤	٢٩,٩١	عام
١٠٠	٠,٣٦	١,٥٢	٧٤,٠٣	٢,٦٩	٢١,٤١	خاص
١٠٠	١,٠٨	٥,٩١	٥٩,٧٥	٧,٦٠	٢٥,٦٦	استثماري
١٠٠	٠,٠	١,٢٣	٦٠,٦٧	١,٠	٣٧,١٠	أخرى وتشمل (التعاونيات والأجنبي)
١٠٠	٠,٣٥	١,٧٠	٧٠,٢٠	٣,٢٥	٢٤,٥١	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء.

توضح بيانات الجدول السابق أن الغالبية العظمى ٧٤% من العاملين بالقطاع الخاص خلال الربع الأول عام ٢٠١١ إستمروا يعملون في نفس القطاع خلال الربع الثاني من نفس العام. تليها نسبة العاملين بالقطاع الحكومي حيث ان حوالي ٣٢% من العاملين بهذا القطاع إستقروا في العمل بالقطاع الحكومي خلال فترة الربعين عام ٢٠١١. بينما نجد أن حوالي ٦١% ممن كانوا يعملون بالقطاع الحكومي خلال الربع الأول عام ٢٠١١ إنتقلوا للعمل بالقطاع الخاص خلال الربع الأول من نفس العام، وبالنسبة للعاملين في القطاع العام خلال الربع الأول عام ٢٠١١، تبين ان حوالي ٨% منهم ظلوا يعملون في نفس القطاع، حوالي ٥٧% منهم إنتقلوا للعمل بالقطاع الخاص وحوالي ٢٩% إنتقلوا للعمل بالقطاع الحكومي خلال الربع الثاني من نفس العام، وبالنسبة للعاملين بالقطاع الاستثماري خلال الربع الأول عام ٢٠١١، فإن حوالي ٥% منهم استمروا في العمل بنفس القطاع، بينما أكثر من نصفهم حوالي ٥٩% منهم إنتقلوا للعمل بالقطاع الخاص خلال الربع الثاني من نفس العام، وأيضا بالنسبة للعاملين بالقطاعات الأخرى وتشمل التعاونيات والأجنبي خلال الربع الأول عام ٢٠١١، نجد أن حوالي ١% منهم فقط إستمروا في العمل بنفس القطاعات، بينما أكثر من ٦٠% منهم إنتقلوا للعمل بالقطاع الخاص، ٣٧% منهم إنتقلوا للعمل بالقطاع الحكومي خلال الربع الثاني من نفس العام.

مما سبق يتبين ما للقطاع الخاص من أهمية في إستيعاب العاملين من جميع القطاعات الأخرى.

بالقطاع الحكومي حيث ان حوالي ٦٤% من العاملين بهذا القطاع إستقروا في العمل بالقطاع الحكومي خلال فترة الربعين عام ٢٠١٠. بينما نجد أن حوالي ٣٠% ممن كانوا يعملون بالقطاع الحكومي خلال الربع الأول عام ٢٠١٠ إنتقلوا للعمل بالقطاع الخاص خلال الربع الثاني من نفس العام. وبالنسبة للعاملين في القطاع العام خلال الربع الأول عام ٢٠١٠، تبين ان حوالي ٣٥% منهم ظلوا يعملون في نفس القطاع، حوالي ٣٧% منهم إنتقلوا للعمل بالقطاع الخاص وحوالي ٢١% إنتقلوا للعمل بالقطاع الحكومي خلال الربع الثاني من نفس العام. وبالنسبة للعاملين بالقطاع الاستثماري خلال الربع الأول عام ٢٠١٠، فإن حوالي ١٩% منهم استمروا في العمل بنفس القطاع، بينما نصفهم حوالي ٥٠% منهم إنتقلوا للعمل بالقطاع الخاص خلال الربع الثاني من نفس العام. وأيضا بالنسبة للعاملين بالقطاعات الأخرى وتشمل التعاونيات والاجنبى خلال الربع الأول عام ٢٠١٠، نجد أن حوالي ١٠% منهم فقط إستمروا في العمل بنفس القطاعات، بينما أكثر من ٥٠% منهم إنتقلوا للعمل بالقطاع الخاص، ٣٣% منهم إنتقلوا للعمل بالقطاع الحكومي خلال الربع الثاني من نفس العام.

جدول رقم (١٦) مصفوفة الإنتقال للمشتغلين طبقاً لنوع القطاع خلال الربعين الثالث والرابع لعام ٢٠١٠

القطاع للربع الرابع ٢٠١٠ %						القطاع للربع الثالث ٢٠١٠
الإجمالي	أخرى وتشمل التعاونيات	استثمارى	خاص	عام	حكومى	
١٠٠	٠,٢١	١,٦١	٤٧,١٦	٤,١٢	٤٦,٨٩	حكومى
١٠٠	٠,٠	٣,٠٢	٥٠,٥٧	٢٠,١١	٢٦,٣٠	عام
١٠٠	٠,١٦	١,٢٥	٨٠,١٨	٢,٦٠	١٥,٨١	خاص
١٠٠	٠,٠	١٥,٣٥	٥٩,٠	٥,٨٩	١٩,٧٦	استثمارى
١٠٠	١,٨٢	٥,٣٥	٥٣,٥٤	٣,٣٤	٣٥,٩٥	أخرى وتشمل (التعاونيات والأجنبي)
١٠٠	٠,١٧	١,٦٩	٧٠,٦٨	٣,٦٣	٢٣,٨٣	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزى للتعبئة العامة والإحصاء.

توضح بيانات الجدول السابق أن الغالبية العظمى ٨٠% من العاملين بالقطاع الخاص خلال الربع الثالث عام ٢٠١٠ إستمروا يعملون في نفس القطاع خلال الربع الرابع من نفس العام. تليها نسبة العاملين بالقطاع الحكومي حيث ان حوالي ٤٦% من العاملين بهذا القطاع إستقروا في العمل بالقطاع الحكومي خلال فترة الربعين عام ٢٠١٠. بينما نجد أن حوالي ٤٧% ممن كانوا يعملون بالقطاع الحكومي خلال الربع الثالث عام ٢٠١٠ إنتقلوا للعمل بالقطاع الخاص خلال الربع الرابع من نفس العام. وبالنسبة للعاملين بالقطاع الاستثماري خلال الربع الثالث عام ٢٠١٠، فإن حوالي ١٥% منهم استمروا في العمل بنفس القطاع، بينما أكثر من نصفهم حوالي ٥٩% منهم إنتقلوا للعمل بالقطاع الخاص خلال الربع الرابع من نفس العام. مما سبق يتبين ما للقطاع الخاص من أهمية في إستيعاب العاملين من جميع القطاعات الأخرى.

الخدمات ظلوا يعملون بنفس النشاط خلال نفس الفترة. وهذا ما يعكس مدى الاستقرار في العمل بقطاعي الزراعة والخدمات. كذلك تعكس بيانات الجدول أن حوالي ٣٧% ممن كانوا يعملون في النشاط الصناعي خلال الربع الثالث أصبحوا يعملون بقطاع الخدمات خلال الربع الرابع من عام ٢٠١٠.

جدول رقم (١٤) مصفوفة الانتقال للمشتغلين بالنشاط الاقتصادي خلال الربعين الأول والثاني لعام ٢٠١١

النشاط الاقتصادي للربع الأول ٢٠١١	النشاط الاقتصادي للربع الثاني ٢٠١١ %			
	الزراعة	الصناعة	الخدمات	الإجمالي
الزراعة	٥٧,٢١	١٤,٣٨	٢٨,٤١	١٠٠
الصناعة	٢٠,٤٧	٣٢,٦١	٤٦,٩٢	١٠٠
الخدمات	١٩,٥٦	٢٥,٥٠	٥٤,٩٤	١٠٠
الإجمالي	٣١,٨٧	٢٣,٦٣	٤٤,٥٠	١٠٠

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء.

أوضحت البيانات أن (٥٧%) فقط من العاملين بالنشاط الزراعي خلال الربع الأول من عام ٢٠١١ ظلوا يعملون بالزراعة خلال الربع الثاني من عام ٢٠١١، وأن حوالي ٣٢% فقط ممن كانوا يعملون في النشاط الصناعي استمروا في العمل بنفس النشاط، ويرجع هذا الانخفاض نتيجة الظروف التي شهدتها مصر في أعقاب ثورة ٢٥ يناير وما ترتب عليها من أحداث نتج عنها تباطؤ في الأنشطة الاقتصادية بشكل عام خلال تلك الفترة كذلك فإن حوالي ٥٤% ممن يعملون في قطاع الخدمات ظلوا يعملون بنفس النشاط خلال نفس الفترة. وهذا ما يعكس عدم الاستقرار في العمل في القطاعات الثلاثة.

كذلك تعكس بيانات الجدول أن حوالي ٤٦% ممن كانوا يعملون في النشاط الصناعي خلال الربع الأول أصبحوا يعملون بقطاع الخدمات خلال الربع الثاني من عام ٢٠١١.

جدول رقم (١٥) مصفوفة الانتقال للمشتغلين طبقاً لنوع القطاع خلال الربعين الأول والثاني لعام ٢٠١٠

القطاع للربع الأول ٢٠١٠	القطاع للربع الثاني ٢٠١٠ %				
	حكومي	عام	خاص	استثماري	أخرى وتشمل التعاونيات
حكومي	٦٤,٧٧	٣,٤٧	٣٠,٤٧	٠,٨٦	٠,٤٣
عام	٢١,٧٧	٣٥,٨٦	٣٧,٩٨	٣,٤١	٠,٩٨
خاص	١١,١٧	١,٧٨	٨٥,٥٥	١,٢٣	٠,٢٧
استثماري	١٨,٣١	١١,٢٥	٥٠,٩٧	١٩,٤٦	٠
أخرى وتشمل (التعاونيات والأجنبية)	٣٣,١٥	١,٩٢	٥٢,٧٩	١,٧١	١٠,٤٤
الإجمالي	٢٤,٨٦	٣,٦٤	٦٩,٥٩	١,٥٤	٠,٣٧

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء.

فيما يتعلق بالانتقال العاملين من قطاع إلى آخر خلال فترة الربعين الأول والثاني من عام ٢٠١٠. توضح بيانات الجدول أن الغالبية العظمى ٨٥% من العاملين بالقطاع الخاص خلال الربع الأول عام ٢٠١٠ استمروا يعملون في نفس القطاع خلال الربع الثاني من نفس العام. تليها نسبة العاملين

أوضحت بيانات الجدول السابق أن حوالي ٦٥% من المشتغلين مقابل أجر خلال الربع الأول من عام ٢٠١١ إستمروا يعملون مقابل أجر خلال الربع الثاني من نفس العام، وهذا يعكس مدى عدم الإستقرار في العمل بين العاملين بأجر. ويرجع هذا الإنخفاض نتيجة الظروف التي شهدتها مصر في أعقاب ثورة ٢٥ يناير وما ترتب عليها من أحداث نتج عنها تباطؤ في الأنشطة الاقتصادية بشكل عام خلال تلك الفترة، وهناك حوالي ٣١% من العاملين بأجر أصبحوا إما يعملون لحسابهم أو يعملون لدى الأسرة خلال نفس الفترة، وذلك قد يكون نتيجة لبلوغ سن المعاش والقيام بأعمال خاص بهم أو نتيجة لفقد الوظائف نتيجة للظروف التي شهدتها مصر خلال تلك الفترة. وبالنسبة لمن يعملون لحساب أنفسهم إما كأصحاب أعمال أو لحساب أنفسهم، نجد أن حوالي ٥٤% منهم إنتقلوا إلى العمل مقابل أجر خلال فترة الدراسة، وهذا قد يكون بين المشتغلين الذين تعثرت أعمالهم الخاصة فتم اللجوء إلى العمل لدى الغير مقابل أجر. ويوضح الجدول أيضاً، أن حوالي ٢١% ممن كانوا يعملون لدى الأسرة بدون مقابل خلال فترة الدراسة أصبحوا يعملون إما لدى غير مقابل أجر أو لحسابهم الخاص.

جدول رقم (١٢) مصفوفة الإنتقال للمشتغلين بالنشاط الاقتصادي خلال الربعين الأول والثاني لعام ٢٠١٠

النشاط لاقتصادي للربع الثاني ٢٠١٠%				النشاط الاقتصادي للربع الأول ٢٠١٠
الإجمالي	الخدمات	الصناعة	الزراعة	الزراعة
١٠٠	١٦,٩٣	١١,٠٦	٧٢	الزراعة
١٠٠	٢٥,٨	٦٢,١	١٢,١	الصناعة
١٠٠	٧٤,٨٦	١٤,٩	١٠,٢٥	الخدمات
١٠٠	٤٦,٣٢	٢٦,٧٤	٢٦,٩٤	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء. وبالنسبة للتغيرات التي طرأت على المشتغلين طبقاً للنشاط الاقتصادي خلال فترة الربعين المشار إليهم، أوضحت البيانات أن الغالبية العظمى (٧٢%) من العاملين بالنشاط الزراعي خلال الربع الأول ظلوا يعملون بالزراعة خلال الربع الثاني من عام ٢٠١٠، وأن حوالي ٦٢% ممن كانوا يعملون في النشاط الصناعي إستمروا في العمل بنفس النشاط، كذلك فإن حوالي ٧٤% ممن يعملون في قطاع الخدمات ظلوا يعملون بنفس النشاط خلال نفس الفترة. وهذا ما يعكس مدى الاستقرار في العمل بقطاعي الزراعة والخدمات.

جدول رقم (١٣) مصفوفة الإنتقال للمشتغلين بالنشاط الاقتصادي خلال الربعين الثالث والرابع لعام ٢٠١٠

النشاط لاقتصادي للربع الرابع ٢٠١٠%				النشاط الاقتصادي للربع الثالث ٢٠١٠
الإجمالي	الخدمات	الصناعة	الزراعة	الزراعة
١٠٠	٢٤,٤٢	١٤,٦٧	٦٠,٩٠	الزراعة
١٠٠	٣٧,٥٧	٤٨,٠١	١٤,٤٢	الصناعة
١٠٠	٦٢,٩٨	٢٢,١٧	١٤,٨٥	الخدمات
١٠٠	٤٤,٩٩	٢٦,٧٩	٢٨,٢٢	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء. أوضحت البيانات أن (٦٠%) فقط من العاملين بالنشاط الزراعي خلال الربع الثالث ظلوا يعملون بالزراعة خلال الربع الرابع من عام ٢٠١٠، وأن حوالي ٤٨% فقط ممن كانوا يعملون في النشاط الصناعي إستمروا في العمل بنفس النشاط، كذلك فإن حوالي ٦٢% ممن يعملون في قطاع

فيما يتعلق بالتغيرات التي طرأت على المشتغلين طبقاً لحالتهم العملية خلال فترة الدراسة، أوضحت بيانات الجدول السابق أن حوالي ٧٩% من المشتغلين مقابل أجر خلال الربع الأول من عام ٢٠١٠ إستمروا يعملون مقابل أجر خلال الربع الثاني من نفس العام، وهذا يعكس مدى الإستقرار في العمل بين العاملين بأجر. وهناك حوالي ١٧% من العاملين بأجر أصبحوا إما يعملون لحسابهم أو يعملون لدى الأسرة خلال نفس الفترة، وذلك قد يكون نتيجة لبلوغ سن المعاش والقيام بأعمال خاص بهم. وبالنسبة لمن يعملون لحساب أنفسهم إما كأصحاب أعمال أو لحساب أنفسهم، نجد أن حوالي ٣٨% منهم إنتقلوا إلى العمل مقابل أجر خلال فترة الدراسة، وهذا قد يكون بين المشتغلين الذين تعثرت أعمالهم الخاصة فتم اللجوء إلى العمل لدى الغير مقابل أجر. ويوضح الجدول أيضاً، أن حوالي ١٩% ممن كانوا يعملون لدى الأسرة بدون مقابل خلال فترة الدراسة أصبحوا يعملون إما لدى الغير مقابل أجر أو لحسابهم الخاص.

جدول رقم (١٠) مصفوفة الإنتقال للمشتغلين طبقاً للحالة العملية خلال الربعين الثالث والرابع لعام ٢٠١٠

الحالة العملية للربع الرابع ٢٠١٠ %				الحالة العملية للربع الثالث ٢٠١٠
يعمل بأجر	يعمل لحسابه	يعمل لدى الأسرة بدون أجر	الإجمالي	
٧٣,٩١	٢٤,٠	٢,٠٩	١٠٠	يعمل بأجر
٤٨,٠٤	٤٧,٨٥	٤,١١	١٠٠	يعمل لحسابه
٢٤,٠	١٥,١٨	٦٠,٨٢	١٠٠	يعمل لدى الأسرة بدون أجر
٦١,٩٩	٣١,١٤	٦,٨٧	١٠٠	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء. أوضحت بيانات الجدول السابق أن حوالي ٧٣% من المشتغلين مقابل أجر خلال الربع الثالث من عام ٢٠١٠ إستمروا يعملون مقابل أجر خلال الربع الرابع من نفس العام، وهذا يعكس مدى الإستقرار في العمل بين العاملين بأجر. وهناك حوالي ٢٤% من العاملين بأجر أصبحوا إما يعملون لحسابهم أو يعملون لدى الأسرة خلال نفس الفترة، وذلك قد يكون نتيجة لبلوغ سن المعاش والقيام بأعمال خاصة بهم. وبالنسبة لمن يعملون لحساب أنفسهم إما كأصحاب أعمال أو لحساب أنفسهم، نجد أن حوالي ٤٨% منهم إنتقلوا إلى العمل مقابل أجر خلال فترة الدراسة، وهذا قد يكون بين المشتغلين الذين تعثرت أعمالهم الخاصة فتم اللجوء إلى العمل لدى الغير مقابل أجر. ويوضح الجدول أيضاً، أن حوالي ٢٤% ممن كانوا يعملون لدى الأسرة بدون مقابل خلال فترة الدراسة أصبحوا يعملون إما لدى غير مقابل أجر أو لحسابهم الخاص.

جدول رقم (١١) مصفوفة الإنتقال للمشتغلين طبقاً للحالة العملية خلال الربعين الأول والثاني لعام ٢٠١١

الحالة العملية للربع الثاني ٢٠١١ %				الحالة العملية للربع الأول ٢٠١١
يعمل بأجر	يعمل لحسابه	يعمل لدى الأسرة بدون أجر	الإجمالي	
٦٥,٠٥	٣١,٥١	٣,٤٤	١٠٠	يعمل بأجر
٥٤,٦٣	٤١,٠٨	٤,٢٩	١٠٠	يعمل لحسابه
٢١,٦٣	١٨,٩٤	٥٩,٤٣	١٠٠	يعمل لدى الأسرة بدون أجر
٥٨,٤١	٣٣,٩٩	٧,٦٠	١٠٠	الإجمالي

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزي للتعبئة العامة والإحصاء.

ويعرض الجدول السابق مصفوفة الانتقال للأفراد داخل القوى البشرية طبقاً للموقف من العمل خلال الربعين الثالث والرابع لعام ٢٠١٠. بتحليل بيانات الجدول نجد أن حوالى ٧٣% من الأفراد المشتغلين خلال الربع الثالث من عام ٢٠١٠ إستمروا يعملون خلال الربع الرابع من نفس العام. وكذلك فإن حوالى ٣٠% من المتعطلين خلال الربع الثالث أصبحوا مشتغلين فى الربع الرابع. ويتضح أيضاً من بيانات الجدول أن حوالى ٤٨% من المتعطلين و ٢٤% فقط من المشتغلين خلال الربع الثالث قد خرجوا من سوق العمل خلال الربع الرابع لعام ٢٠١٠ وقد يرجع ذلك لعدة أسباب، من أهمها إما بلوغ سن المعاش لبعض أفراد عينة البحث ،أو لرجوع الطلبة للدراسة الذين كانوا يعملون خلال (الربع الثالث) أى فى فترة أجازة الصيف.

جدول رقم (٨) مصفوفة الانتقال للأفراد طبقاً للموقف من العمل خلال الربعين الأول والثانى لعام ٢٠١١

الموقف من العمل للربع الثانى ٢٠١١ %				الموقف من العمل للربع الأول ٢٠١١
مشتغل	متعطّل	خارج قوة العمل	الإجمالى	
٥٩,٩١	٦,٢٦	٣١,٨٣	١٠٠	مشتغل
٣٩,٤٧	١٢,٦٠	٤٧,٩٣	١٠٠	متعطّل
٢٧,٩٠	٨,٣٩	٦٣,٧١	١٠٠	خارج قوة العمل
٤١,٤٧	٧,٧٨	٥٠,٧٥	١٠٠	الإجمالى

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزى للتعبئة العامة والإحصاء.

ويعرض الجدول السابق مصفوفة الانتقال للأفراد داخل القوى البشرية طبقاً للموقف من العمل خلال الربعين الأول والثانى لعام ٢٠١١. بتحليل بيانات الجدول نجد أن حوالى ٥٩% من الأفراد المشتغلين خلال الربع الأول من عام ٢٠١١ إستمروا يعملون خلال الربع الثانى من نفس العام ويرجع هذا الإنخفاض نتيجة الظروف التى شهدتها مصر فى أعقاب ثورة ٢٥ يناير وما ترتب عليها من أحداث نتج عنها تباطؤ فى الأنشطة الإقتصادية بشكل عام خلال تلك الفترة. وكذلك فإن حوالى ٣٩% من المتعطلين خلال الربع الأول أصبحوا مشتغلين فى الربع الثانى. ويتضح أيضاً من بيانات الجدول أن حوالى ٤٨% من المتعطلين و ٣١% فقط من المشتغلين خلال الربع الأول قد خرجوا من سوق العمل خلال الربع الثانى لعام ٢٠١١ وقد يرجع ذلك لعدة أسباب، من أهمها إما بلوغ سن المعاش لبعض أفراد عينة البحث ،أو لرجوع الطلبة للدراسة الذين كانوا يعملون خلال (الربع الأول) أى فى فترة أجازة نصف العام.

جدول رقم (٩) مصفوفة الانتقال للمشتغلين طبقاً للحالة العملية خلال الربعين الأول والثانى لعام ٢٠١٠

الحالة العملية للربع الثانى ٢٠١٠ %				الحالة العملية للربع الأول ٢٠١٠
يعمل بأجر	يعمل لحسابه	يعمل لدى الاسرة بدون أجر	الإجمالى	
٧٩,٧٤	١٧,٨٤	٢,٤٢	١٠٠	يعمل بأجر
٣٨,٤٣	٥٦,٧٨	٤,٧٩	١٠٠	يعمل لحسابه
١٩,٤٧	١٥,١٤	٦٥,٣٩	١٠٠	يعمل لدى الاسرة بدون أجر
٦٣,٤٢	٢٨,٩٣	٧,٦٥	١٠٠	الإجمالى

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزى للتعبئة العامة والإحصاء.

٣,٤ دراسة وتحليل اتجاهات سوق العمل وذلك بتحليل ومقارنة البيانات المقطعية التى جمعت خلال الفترة من ٢٠١٠ - ٢٠١١ باستخدام الأسلوب التتبعى لأفراد الأسر المعيشية:

يستعرض هذا الجزء من الدراسة نتائج تتبع مجموعة من أفراد عينة بحث القوى العاملة فى سن العمل (١٥-٦٤ سنة) خلال الأربع الأول (يناير- مارس) والثانى (أبريل- يونيه) والثالث (يوليو - سبتمبر) والرابع (أكتوبر - ديسمبر) لعام ٢٠١٠ والإربع الأول (يناير - مارس) والثانى (أبريل - يونيه) والثالث (يوليو - ديسمبر) لعام ٢٠١١.

جدول رقم (٦) مصفوفة الانتقال للأفراد طبقاً للموقف من العمل خلال الربعين الأول والثانى لعام ٢٠١٠

الموقف من العمل للربع الثانى ٢٠١٠ %				الموقف من العمل للربع الأول ٢٠١٠
الإجمالى	خارج قوة العمل	متعطّل	مشتغل	
١٠٠	١٩,٦٢	٢,١١	٧٨,٢٧	مشتغل
١٠٠	٣٨,٦٦	٢٥,٤٣	٣٥,٩١	متعطّل
١٠٠	٧٣,١٥	٤,٦٢	٢٢,٢٣	خارج قوة العمل
١٠٠	٤٦,٤٥	٤,٢٥	٤٩,٣	الإجمالى

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزى للتعبئة العامة والإحصاء.

ويعرض الجدول السابق مصفوفة الانتقال للأفراد داخل القوى البشرية طبقاً للموقف من العمل خلال الربعين الأول والثانى لعام ٢٠١٠. بتحليل بيانات الجدول نجد أن حوالى ٧٨% من الأفراد المشتغلين خلال الربع الأول من عام ٢٠١٠ إستمروا يعملون خلال الربع الثانى من نفس العام. وكذلك فإن حوالى ٣٥% من المتعطّلين خلال الربع الأول أصبحوا مشتغلين فى الربع الثانى. ويتضح أيضاً من بيانات الجدول أن حوالى ٣٨% من المتعطّلين و ١٩% فقط من المشتغلين خلال الربع الأول قد خرجوا من سوق العمل خلال الربع الثانى لعام ٢٠١٠ وقد يرجع ذلك لعدة أسباب، من أهمها إما بلوغ سن المعاش لبعض أفراد عينة البحث، أو لرجوع الطلبة للدراسة الذين كانوا يعملون خلال (الربع الأول) أى فى فترة أجازة نصف العام.

جدول رقم (٧) مصفوفة الانتقال للأفراد طبقاً للموقف من العمل خلال الربعين الثالث والرابع لعام ٢٠١٠

الموقف من العمل للربع الرابع ٢٠١٠ %				الموقف من العمل للربع الثالث ٢٠١٠
الإجمالى	خارج قوة العمل	متعطّل	مشتغل	
١٠٠	٢٤,٢	٢,٢٦	٧٣,٥٥	مشتغل
١٠٠	٤٨,٤٦	٢٠,٧٩	٣٠,٧٦	متعطّل
١٠٠	٦٦,٥٩	٤,٨٢	٢٨,٥٩	خارج قوة العمل
١٠٠	٤٦,٠٤	٤,١٥	٤٩,٨١	الإجمالى

المصدر: إعداد الباحث من واقع البيانات الخام لبحث القوى العاملة - الجهاز المركزى للتعبئة العامة والإحصاء.

جدول رقم (٥) معدلات المساهمة في النشاط الإقتصادي (١٥ سنه فأكثر) طبقاً لفئات السن، الحالة التعليمية والأقاليم الجغرافية والنوع

(%)

٢٠١١			٢٠١٠			الخصائص
النوع						
جملة	إناث	ذكور	جملة	إناث	ذكور	
فئات السن						
١٧,١	٦,٧	٢٦,٦	٢٠,١	٨,٧	٣٠,٩	١٩ - ١٥
٥١,٥	٢٨,٤	٧٠,٩	٤٩,٩	٢٩,١	٦٩,١	٢٤- ٢٠
٦١,٢	٢٧,٧	٩٧,٢	٦٢,٣	٢٩,٣	٩٦,٥	٢٩ - ٢٥
٦١,٦	٢٧	٩٨,٦	٦٢,٢	٢٧,٧	٩٨,٣	٣٩ - ٣٠
٦٤,٥	٣٢,٢	٩٧,٧	٦٣,٤	٣٠,٢	٩٧,٣	٤٩ - ٤٠
٥٧,٨	٢٤,٧	٩٠,٢	٥٧,٥	٢٥,١	٨٩,٧	٥٩ - ٥٠
٣٠,٦	٨,٥	٥١,٧	٣٠,٩	٨,٥	٥٢,٧	٦٤ - ٦٠
١٢,٢	٢,٢	٢١,٤	١٣,٣	٣,٣	٢٢,٩	+ ٦٥
٤٨,٨	٢٢,٥	٧٤,٦	٤٩,٣	٢٣,١	٧٥,١	الإجمالي
الحالة التعليمية						
٣٩,٦	١٦,٧	٧٩,٥	٤٠,٠	١٧,٤	٧٩,٨	أمية
٥٨,٩	١٤,٣	٨٣,٤	٥٩,٢	١٥,٠	٨٣,٦	يقرأ ويكتب وشهادة محو الأمية
٢٦,٧	٥,١	٤٣,٧	٢٨,٢	٥,٢	٤٦,٤	مؤهل أقل من المتوسط
١٤,٥	٣,٨	٢٣	١٥,٠	٤,٥	٢٣,٨	ثانوية عامة و ثانوية أهلية
٦٤,٥	٣٠,٥	٩٢,١	٦٣,٨	٣٠,٦	٩١,١	مؤهل متوسط فني
٧٢,٥	٤٨,٨	٩٠,١	٧٣,٥	٥١,٨	٨٩,٩	مؤهل فوق المتوسط وأقل من الجامعي
٧٧,٧	٦١,٥	٩٠,٤	٧٨,٠	٦٢,٧	٨٩,٨	مؤهل جامعي ومؤهل فوق الجامعي
٤٨,٨	٢٢,٥	٧٤,٦	٤٩,٣	٢٣,١	٧٥,١	الإجمالي
الأقاليم الجغرافية						
٤٤,٧	٢٠,٣	٦٩,٢	٤٥,٥	٢١,٠	٧٠,٢	المحافظات الحضرية
٤٨,٣	٢٤,٥	٧٢,٤	٤٨,٠	٢٣,٤	٧٣,٠	حضر الوجه البحري
٥٢	٢٥,٦	٧٧,٦	٥١,٧	٢٤,٩	٧٧,٩	ريف الوجه البحري
٤٧,٧	٢٢,٥	٧٢,٣	٤٨,٢	٢٣,٤	٧٢,٨	حضر الوجه القبلي
٤٨,٢	١٨,٩	٧٦,٢	٤٩,١	٢١,٢	٧٦,٥	ريف الوجه القبلي
٥٧,١	٣٢,٧	٧٩	٥٨,٤	٣٣,٨	٨١,٤	حضر محافظات الحدود
٥٣,٩	٢٦,٨	٧٩,٩	٥٥,٥	٢٩,٤	٨٢,٨	ريف محافظات الحدود
٤٨,٨	٢٢,٥	٧٤,٦	٤٩,٣	٢٣,١	٧٥,١	إجمالي الجمهورية

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء. نشرة القوى العاملة السنوية لعام ٢٠١٠، ٢٠١١

يوضح الجدول السابق مدى تأثر سوق العمل المصرى بالأحداث التى شهدتها مصر والتى أثرت بشكل كبير على الأداء الإقتصادى وكثرة الإضرابات والإعتصامات والمطالبة بحقوق العمال مما أدى إلى تباطؤ الإداء الإقتصادى وغلق كثير من المصانع والشركات ، حيث حقق معدل البطالة تزايداً مستمراً خلال الفترة من عام (٢٠٠٧ - ٢٠١١) حيث بلغ عام ٢٠١١ (١٢,٠ %) بينما كان ٨,٨ % عام ٢٠٠٧ .

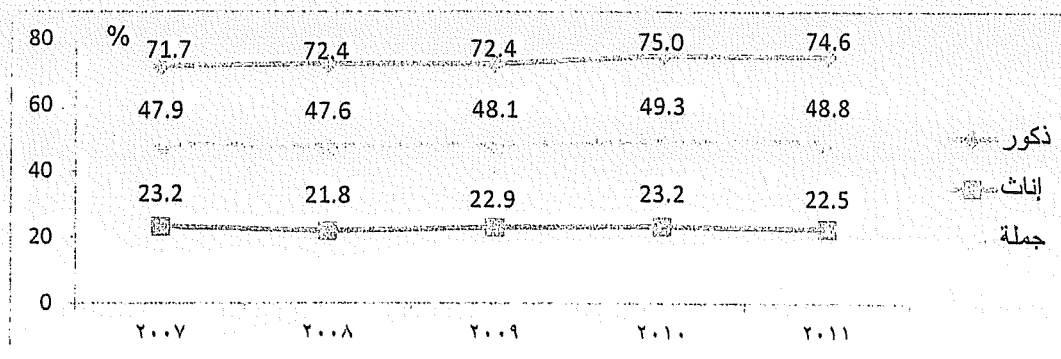
٢,٤ حالة سوق العمل المصرى:

جدول رقم (٤) تطور معدل المساهمة فى النشاط الإقتصادى للأفراد (١٥ سنة فأكثر) طبقاً للنوع للأعوام (٢٠٠٧ إلى ٢٠١١)

السنوات	معدل المساهمة فى النشاط الإقتصادى (%)		
	ذكور	إناث	جملة
٢٠٠٧	٧١,٧	٢٣,٢	٤٧,٩
٢٠٠٨	٧٢,٤	٢١,٨	٤٧,٦
٢٠٠٩	٧٢,٣	٢٢,٩	٤٨,١
٢٠١٠	٧٥	٢٣,٢	٤٩,٣
٢٠١١	٧٤,٦	٢٢,٥	٤٨,٨

المصدر: الجهاز المركزى للتعبئة العامة والإحصاء نشرة القوى العاملة السنوية لعام ٢٠١١

يقصد بمعدل المساهمة فى النشاط الإقتصادى نسبة الأفراد داخل قوة العمل إلى إجمالى السكان (١٥ سنة فأكثر) ويتضح لنا من الجدول السابق إنخفاض معدلات المساهمة خلال عام ٢٠١١ عنها فى عام ٢٠١٠ حيث بلغ معدل المساهمة لإجمالى الجمهورية فى عام ٢٠١١ ٤٨,٨ % بينما بلغ ٤٩,٣ % عام ٢٠١٠ وكان بين الذكور ٧٤,٦ % فى عام ٢٠١١ وبلغ ٧٥,٠ % فى عام ٢٠١٠ بينما بلغ بين الإناث ٢٢,٥ % فى عام ٢٠١١ بينما كان ٢٣,٢ % عام ٢٠١٠ .



شكل رقم (٤) تطور معدل المساهمة فى النشاط الإقتصادى للأفراد (١٥ سنة فأكثر) طبقاً للنوع للأعوام (٢٠٠٧ إلى ٢٠١١)

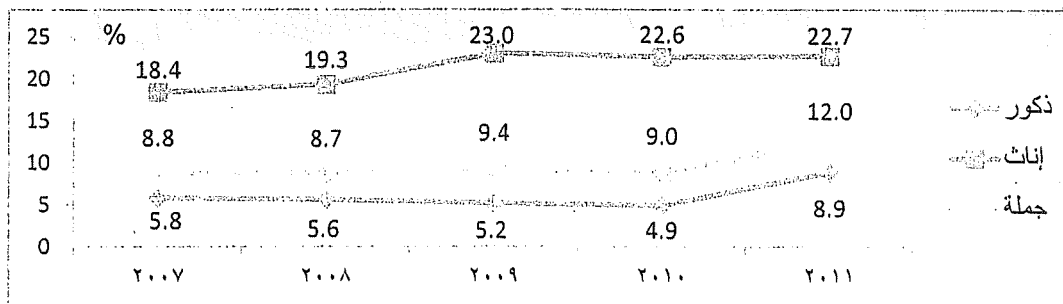
يوضح الجدول السابق قوة العمل والمشتغلين (١٥ سنه فأكثر) والمتعطلين (١٥ - ٦٤ سنه) ومعدل البطالة طبقاً للربع فى ٢٠١١ حيث أظهرت النتائج زيادة قوة العمل خلال الربع الرابع عن الربع الأول من نفس العام بزيادة قدرها ٤٤٦ ألف فرد بنسبه قدرها ١,٧% وبالنسبه للمشتغلين يوضح الجدول إنخفاض فى أعداد المشتغلين خلال الربع الأول عن الربع الرابع من عام ٢٠١٠ بنقص قدره ٦٥٣ ألف مشتغل بنسبه قدرها ٢,٧% ، العام ويرجع هذا الإنخفاض نتيجة الظروف التى شهدتها مصر فى أعقاب ثورة ٢٥ يناير وما ترتب عليها من أحداث نتج عنها تباطؤ فى الأنشطة الإقتصادية بشكل عام خلال تلك الفترة ، وبالنسبة لمعدل البطالة يوضح الجدول إرتفاع معدل البطالة خلال الربع الرابع عن ما كان عليه خلال الربع الأول حيث بلغ معدل البطالة ١٢,٤% بينما كان ١١,٩ خلال الربع الأول من نفس العام وبلغ ٨,٩% عن نفس الربع من عام ٢٠١٠ وهذه البيانات تعكس مدى تأثر سوق العمل المصرى بالأحداث التى شهدتها مصر والتى أثرت بشكل كبير على الأداء الإقتصادى وكثرة الإضرابات والإعتصامات والمطالبة بحقوق العمال مما أدى إلى تباطؤ الإداء الإقتصادى وغلق كثير من المصانع والشركات.

جدول رقم (٣) تقدير قوة العمل والمشتغلين (١٥ سنه فأكثر) والمتعطلين (١٥ - ٦٤ سنه) ومعدل البطالة طبقاً للنوع خلال الفترة (٢٠٠٧ - ٢٠١١)

الأعداد بالآلاف

البيان	النوع	٢٠٠٧	٢٠٠٨	٢٠٠٩	٢٠١٠	٢٠١١
قوة العمل	ذكور	١٨٥١١	١٩١٢٠	١٩٤١٠	٢٠١٤٠	٢٠٥٤١
	إناث	٥٧٤٠	٥٥٣١	٥٩٤٣	٦٠٤٠	٥٩٨٨
	جملة	٢٤٢٥١	٢٤٦٥١	٢٥٣٥٣	٢٦١٨٠	٢٦٥٢٩
المشتغلين	ذكور	١٧٤٣٣	١٨٠٤٢	١٨٣٩٧	١٩١٥٣	١٨٧١٩
	إناث	٤٦٨٢	٤٤٦٦	٤٥٧٨	٤٦٧٦	٤٦٢٧
	جملة	٢٢١١٥	٢٢٥٠٨	٢٢٩٧٥	٢٣٨٢٩	٢٣٣٤٦
المتعطلين	ذكور	١٠٧٨	١٠٧٨	١٠١٣	٩٨٧	١٨٢٢
	إناث	١٠٥٧	١٠٦٦	١٣٦٥	١٣٦٣	١٣٦١
	جملة	٢١٣٥	٢١٤٤	٢٣٧٨	٢٣٥٠	٣١٨٣
معدل البطالة (%)	ذكور	٥,٨	٥,٦	٥,٢	٤,٩	٨,٩
	إناث	١٨,٤	١٩,٣	٢٣,٠	٢٢,٦	٢٢,٧
	جملة	٨,٨	٨,٧	٩,٤	٩,٠	١٢,٠

المصدر: الجهاز المركزى للتعبئة العامة والإحصاء. نشرة القوى العاملة السنوية لعام ٢٠١١



شكل رقم (٣) معدل البطالة طبقاً للنوع خلال الفترة (٢٠٠٧ - ٢٠١١)

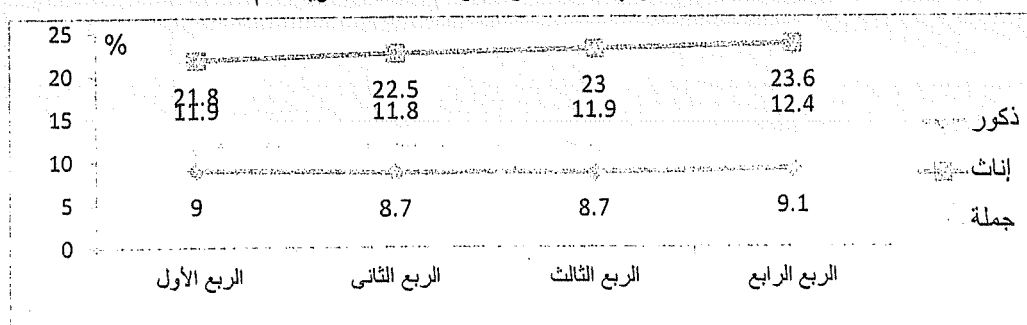
يوضح الجدول السابق قوة العمل والمشتغلين (١٥ سنة فأكثر) والمتعطلين (١٥ - ٦٤ سنة) ومعدل البطالة طبقاً للربع في ٢٠١٠ حيث أظهرت النتائج زيادة قوة العمل خلال الربع الرابع عن الربع الأول من نفس العام بزيادة قدرها ٦ آلاف فرد بنسبة قدرها ٠,٢% وبالنسبة للمشتغلين يوضح الجدول زيادة في المشتغلين خلال الربع الرابع عن الربع الأول بزيادة قدرها ٦٣ ألف مشتغل بنسبة قدرها ٠,٣% ولكن بالنسبة للمشتغلين الإناث فنجد إنخفاض بمقدار ٢٢٣ مشتغلة بنسبة قدرها ٤,٦% وبالنسبة للمتعطلين يوضح الجدول إنخفاض في المتعطلين خلال الربع الرابع عن الربع الأول بنقص قدره ٥٨ ألف متعطّل بنسبة قدرها ٢,٤% وبالنسبة لمعدل البطالة يوضح الجدول إنخفاض معدل البطالة خلال الربع الرابع عن ما كان عليه خلال الربع الأول حيث بلغ معدل البطالة ٨,٩% بينما كان ٩,١٢ خلال الربع الأول من نفس العام ولكن بالنسبة للإناث فقد ارتفع معدل البطالة خلال الربع الرابع حيث بلغ ٢٢,٧٦ بينما كان ٢١,٩٧.

جدول رقم (٢) تقدير قوة العمل والمشتغلين (١٥ سنة فأكثر) والمتعطلين (١٥ - ٦٤ سنة) ومعدل البطالة طبقاً للنوع والربع في ٢٠١١

الأعداد بالآلاف

البيان	النوع	الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع
قوة العمل	ذكور	٢٠٣٦٥	٢٠٣٩١	٢٠٦٩٢	٢٠٧٣٧
	إناث	٥٩٧٣	٥٩٥١	٥٩٧٢	٦٠٤٧
	جملة	٢٦٣٣٨	٢٦٣٤٢	٢٦٦٦٤	٢٦٧٨٤
المشتغلين	ذكور	١٨٥٣٨	١٨٦١٩	١٨٩٠١	١٨٨٤٢
	إناث	٤٦٧١	٤٦١٥	٤٦٠٠	٤٦٢٠
	جملة	٢٣٢٠٩	٢٣٢٣٤	٢٣٥٠١	٢٣٤٦٢
المتعطلين	ذكور	١٨٢٧	١٧٧٢	١٧٩١	١٨٩٥
	إناث	١٣٠٢	١٣٣٦	١٣٧٢	١٤٢٦
	جملة	٣١٢٩	٣١٠٨	٣١٦٣	٣٣٢١
معدل البطالة (%)	ذكور	٩	٨,٧	٨,٧	٩,١
	إناث	٢١,٨	٢٢,٥	٢٣	٢٣,٦
	جملة	١١,٩	١١,٨	١١,٩	١٢,٤

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء. نشرة القوى العاملة السنوية لعام ٢٠١١



شكل رقم (٢) معدل البطالة طبقاً للنوع والربع في عام ٢٠١١

٢. منهجية البحث

تتضمن الدراسة حالة سوق العمل المصرى من خلال بعض المؤشرات العامة، بدراسة وتحليل اتجاهات سوق العمل وذلك بتحليل ومقارنة البيانات القطاعية التى جمعت بين الفترتين (٢٠١٠ ، ٢٠١١) بإستخدام الأسلوب التتبعى لأفراد الأسر المعيشية ، وكذلك تطبيق نموذج الإنحدار اللوجستى ، وتهدف الدراسة إلى تطبيق النموذج على بيانات القوى العاملة لتحديد تأثير العوامل المتضمنة فى الدراسة وتشمل هذه العوامل : المستوى التعليمى ، الحالة الزوجية ، محل الإقامة ومدى تلقى الفرد للتدريب على مشاركة الفرد فى سوق العمل المصرى.

٣. مصادر البيانات

تم الإعتماد فى هذه الدراسة على بيانات بحث القوى العاملة الذى ينتجه الجهاز المركزى للتعبئة العامة والإحصاء بصفة مستمرة وذلك بالإستعانة بالبيانات السنوية لعام ٢٠١٠ والبيانات السنوية لعام ٢٠١١ وذلك للوقوف على الاختلافات التى حدثت فى تلك الفترة الفاصلة فى تاريخ مصر وهى ثورة ٢٥ يناير ٢٠١١ .

٤. تحليل البيانات

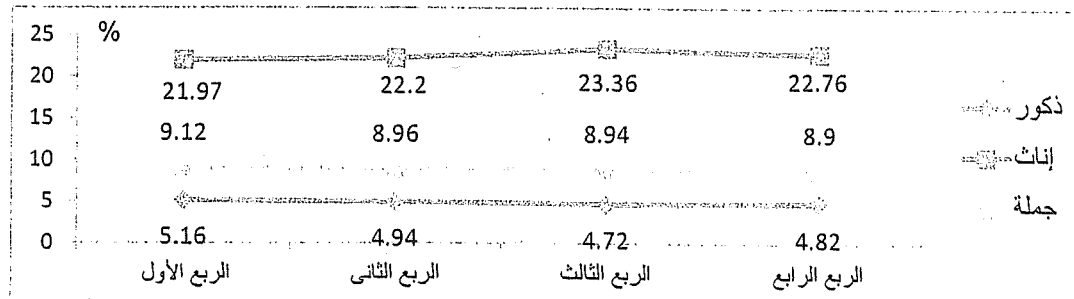
٤.١ مقدمة (بعض المؤشرات العامة):

جدول رقم (١) تقدير قوة العمل والمشتغلين (١٥ سنة فأكثر) والمتعطلين (١٥ - ٦٤ سنة) ومعدل البطالة طبقاً للنوع والربع فى ٢٠١٠

الأعداد بالآلاف

البيان	النوع	الربع الأول	الربع الثانى	الربع الثالث	الربع الرابع
قوة العمل	ذكور	٢٠٠١٣	٢٠٠٩٨	٢٠٢٣٣	٢٠٢٤٣
	إناث	٦١٧٣	٦١٠٣	٥٩٢٠	٥٩٤٩
	جملة	٢٦١٨٦	٢٦٢٠١	٢٦١٥٣	٢٦١٩٢
المشتغلين	ذكور	١٨٩٨١	١٩١٠٦	١٩٢٧٨	١٩٢٦٧
	إناث	٤٨١٨	٤٧٤٨	٤٥٣٧	٤٥٩٥
	جملة	٢٣٧٩٩	٢٣٨٥٤	٢٣٨١٥	٢٣٨٦٢
المتعطلين	ذكور	١٠٣٢	٩٩٣	٩٥٥	٩٧٦
	إناث	١٣٥٦	١٣٥٥	١٣٨٣	١٣٥٤
	جملة	٢٣٨٨	٢٣٤٨	٢٣٣٨	٢٣٣٠
معدل البطالة (%)	ذكور	٥,١٦	٤,٩٤	٤,٧٢	٤,٨٢
	إناث	٢١,٩٧	٢٢,٢	٢٣,٣٦	٢٢,٧٦
	جملة	٩,١٢	٨,٩٦	٨,٩٤	٨,٩٠

المصدر: الجهاز المركزى للتعبئة العامة والإحصاء. نشرة القوى العاملة السنوية لعام ٢٠١٠



شكل رقم (١) معدل البطالة طبقاً للنوع والربع فى عام ٢٠١٠

ورقة بحثية عن تدفق العمالة في سوق العمل المصري بين الفترة ٢٠١٠-٢٠١١ (دراسة مقارنة)

*أحمد ماهر أمين عبدالغنى عمر

**نجلاء عبدالمنعم مراد

ملخص:

يمثل العنصر البشرى قوة العمل التى تمارس كافة الأنشطة الخدمية والإنتاجية والإستهلاكية فى المجتمع ، من هنا تأتى أهمية الإحصاءات والدراسات المتعلقة بنمو وتركيب وتوزيع القوى البشرية وقوة العمل بمختلف أنواعها وخصائصها ، و يحتاج مخطط القوى العاملة إضافة إلى إقتصاد العمل وتقنيات تخطيط القوى العاملة إلى معرفة، على قدر من الوضوح، بأسواق العمل، موضع التخطيط، وبيئتها والتحديات الآنية والمستقبلية التى تواجهها، وتعتبر بيانات السكان والقوى العاملة مجالاً هاماً من المجالات التنظيمية الإحصائية والتخطيطية التى استحوذت على إهتمام الحكومات والمنظمات الإقليمية والدولية. وذلك لأهمية العنصر البشرى فى تخطيط وتنفيذ عملية التنمية الإقتصادية والإجتماعية. مما يتطلب عمل تدابير ورسم سياسات تؤدي إلى توفير البيانات السكانية والإقتصادية والإجتماعية وتداخلاتها مع بذل جهد أكبر من أجل تحسين جودة هذه البيانات وتوفيرها بصورة منتظمة، وعلى ذلك فتهدف هذه الورقة إلى دراسة سوق العمل المصرى بين الفترتين (٢٠١٠ ، ٢٠١١) ودراسة التغيرات التى طرأت عليه خلال تلك الفترة وذلك بالإعتماد على بحث القوى العاملة الذى يصدره الجهاز المركزى للتعبئة العامة والإحصاء.

الكلمات الدالة:

سوق العمل المصرى
قوة العمل
المشتغلين
المتعطلين
البطالة
تدفق العمالة
تحليل الإنحدار اللوجستى

١. أهداف البحث

- أ. التعرف على سوق العمل المصرى بين الفترتين (٢٠١٠ ، ٢٠١١).
- ب. دراسة التغيرات التى طرأت على سوق العمل المصرى والتدفقات خلال فترات مختلفة وخاصة قبل الثورة وبعد الثورة.
- ج. تطبيق نموذج الإنحدار اللوجستى والإستفادة منه فى التعرف على فرص العمل لقوة العمل طبقاً للعديد من المتغيرات والإختلافات بين المحافظات وبين المناطق المختلفة.

*باحث إحصائى بالجهاز المركزى للتعبئة العامة والإحصاء.

**أستاذ مساعد بقسم الإحصاء التطبيقى والإقتصاد القياسى بمعهد الدراسات والبحوث الإحصائية

المراجع

- 1- CURRAN, DEGARMO & SVIOKLA. (2013). PwC's 5th Annual Digital IQ Survey, New York: Digital Conversations, and the C-suite, Price waterhouse Coopers LLP.
- 2- GALLOWAY& GUTHRIE. (2010). Digital IQ Index: Public Sector. New York: L2ThinkTank for digital innovation.
- 3- NARDO, SALTELLI & TARANTOLA. (2005). Tools for Composite Indicators Building. Ispra. European Commission, Joint Research Center.
- 4- OECD (2005). "Handbook on Constructing Composite Indicators: Methodology and User Guide", Organization for Economic Co-operation and Development.

- ٥- مركز المعلومات ودعم اتخاذ القرار (٢٠٠٦)، دليل تكوين المؤشرات المركبة، الإدارة العامة لجودة البيانات، مجلس الوزراء، القاهرة، ٢٠٠٦.
- ٦- المعهد العربي للتدريب والبحوث الإحصائية (٢٠٠٥). نحو خطة استراتيجية للعلاقة بين منتجي ومستخدمي البيانات، ندوة الإحصاء من أجل التنمية: حوار بين منتجي ومستخدمي البيانات، الدوحة ١٦-١٩ مايو.

خاتمة وتوصيات

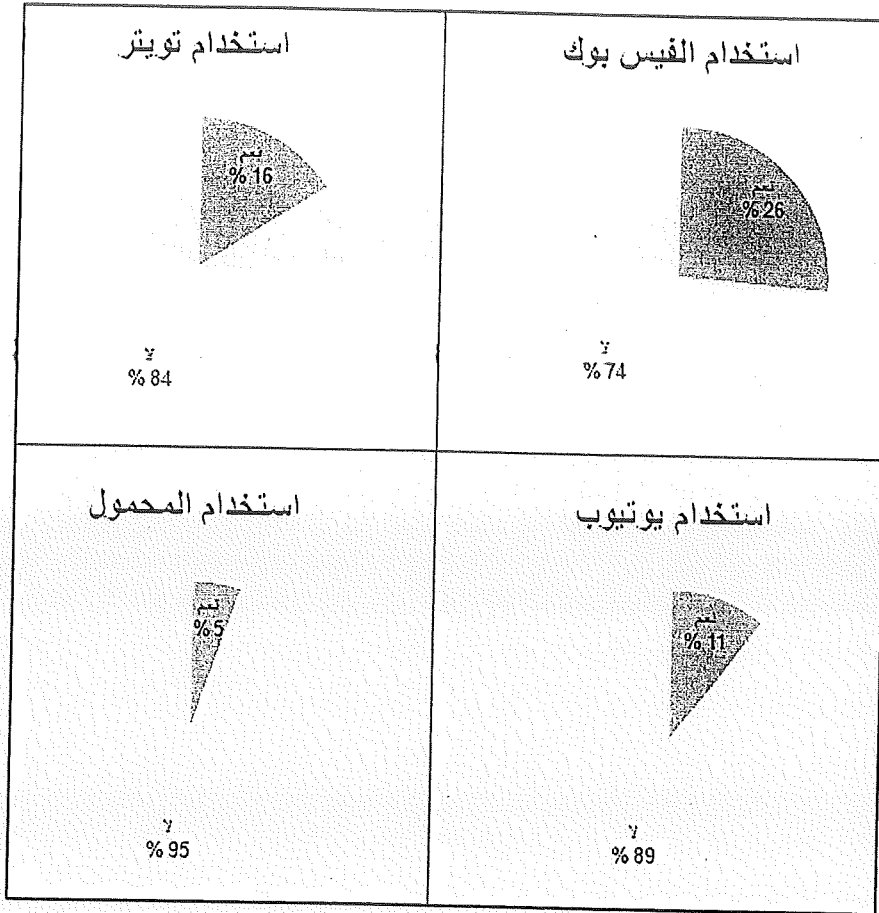
استهدف البحث محاولة تطوير مؤشر إحصائي يقيس الذكاء الرقمي للأجهزة الإحصائية الرسمية العربية، وقد اعتمد البحث في حساب المؤشر بصفة أولية على المؤشر المركب الذي طرحه "Scott Galloway" حول الذكاء الرقمي للشركات ومؤسسات القطاع العام في الولايات المتحدة، مع تطويره وإدخال تعديلات عليه لتلائم عمل الأجهزة الإحصائية الرسمية في العالم العربي. وتوصل البحث إلى محدودية الاستفادة من تكنولوجيا الاتصال الرقمي من قبل هذه الأجهزة، خاصة فيما يتعلق بالإعلام الجديد كالشبكات الاجتماعية وتطبيقات الهواتف الذكية؛ مما يشير إلى ضرورة العمل على تطوير عمل هذه الأجهزة لتيسير إتاحة البيانات لمستخدميها، والوصول إلى قاعدة أكبر من هؤلاء المستخدمين، مع تحقيق التواصل الفعال بين منتجي ومستخدمي البيانات لتعزيز دور الأجهزة الإحصائية الرسمية في رفع الوعي الإحصائي وإتاحة البيانات في عصر المعلومات. مع التأكيد على أن هذا البحث ما هو إلا محاولة أولية لتطوير مؤشر يقيس الكفاءة الرقمية للأجهزة الإحصائية العربية؛ ومن ثم فهناك حاجة لمزيد من التطوير.

التوصيات:

- ١- تكوين فريق عمل متخصص يعمل على تطوير المؤشر المركب للذكاء الرقمي للمؤسسات الإحصائية الرسمية، يضم إحصائيين ومتخصصين في تكنولوجيا المعلومات.
- ٢- عقد ورش عمل لمناقشة كيفية تفعيل استخدام وسائل التكنولوجيا الرقمية في الأجهزة الإحصائية الرسمية العربية لتعزيز ذكاءها الرقمي.
- ٣- عمل الأجهزة الإحصائية الرسمية العربية على تطوير بنيتها الرقمية والاستفادة من نتائج هذا المؤشر في تفادي أوجه القصور.
- ٤- تطبيق مؤشر الذكاء الرقمي على أجهزة الإحصاء الرسمية العربية سنوياً مع عرض النتائج وإجراء تقييم لها.
- ٥- إجراء مقارنات بين الأجهزة العربية وتلك المناظرة لها في الدول المتقدمة حول نتائج المؤشر مع ضرورة الاستفادة من التجارب الرائدة عالمياً في الذكاء الرقمي كجهاز الإحصاء الأسترالي على سبيل المثال.
- ٦- ضرورة الاستفادة من تجربة دولة الإمارات العربية والتعرف على أسباب تقدمها في نتائج مؤشر الذكاء الرقمي وتعميم النتائج على باقي أجهزة الإحصاء الرسمية في الدول العربية.

شكل رقم (٢)

نسب استخدام الأجهزة الإحصائية العربية للشبكات الاجتماعية وتطبيقات المحمول



من إعداد الباحثة.

على موقع ألكسا، كما يتوفر الموقع كنسخة محمولة على الهواتف الذكية، وكذلك إجابة استخدام الشبكات الاجتماعية والخدمات التفاعلية وسهولة وصول المواطنين إلى خدمات الموقع. ثم جاءت دولة فلسطين في المرتبة الثالثة، تليها دولة الكويت في المرتبة الرابعة، وجاءت السودان في المرتبة الأخيرة.

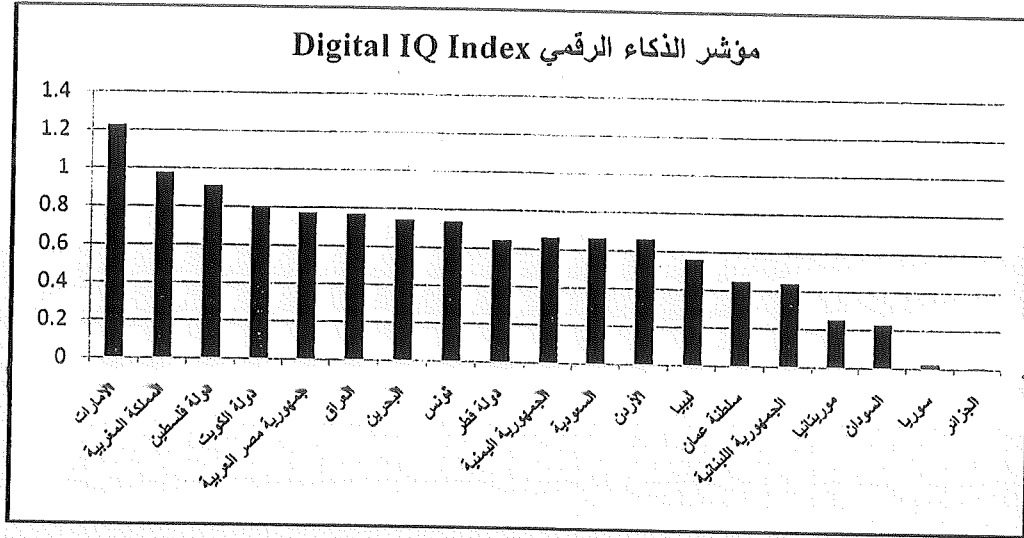
٣- احتلت مصر المركز الخامس في مؤشر الذكاء الرقمي؛ حيث حصلت على ٠,٧٧٥ من قيمة المؤشر؛ ويعود ذلك إلى الترتيب المتقدم الذي احتله موقع الجهاز الإحصائي على موقع ألكسا^٢ حيث يحتل المرتبة ٧١٦,٧٠٧ عالمياً، والمرتبة ٨٠٠١ على مستوى مصر، وتبلغ نسبة المصريين ٧٥% من إجمالي زوار الموقع، بينما يتوزع الباقي بين جنسيات أخرى.

وقد أظهرت إحصاءات موقع ألكسا أن الموقع الإلكتروني للجهاز المركزي للتعينة العامة والإحصاء المصري يزوره في المتوسط ٤ زوار يومياً، ويبلغ إجمالي عدد الدقائق التي يقضيها كل الزوار يومياً أقل من ٧ دقائق، وتزيد نسبة الإنلث بين الزوار، كما تزيد نسبة الحاصلين على مؤهلات عليا بين الزوار، ويتصفح أغلب الزوار الموقع من أماكن عملهم يليهم الزوار من المنازل، ولم يتم التوصل إلى أية إحصاءات مقارنة عبر الزمن لموقع الجهاز؛ نظراً لأن موقع ألكسا لا يُظهر إحصاءات تاريخية للمواقع التي يزيد ترتيبها عن ١٠٠,٠٠٠.

٤- من حيث المؤشرات الفرعية؛ أظهرت النتائج أن أغلب الأجهزة الإحصائية العربية تعتمد على الموقع الإلكتروني فقط، وأن هذه الأجهزة ما زالت بعيدة كل البعد عن الاستفادة من باقي الخدمات التفاعلية والشبكات الاجتماعية وتطبيقات المحمول؛ فهناك ٢٦% فقط من هذه الأجهزة تستخدم الفيس بوك، و ١٦% فقط تستخدم تويتر، و ١١% فقط تستخدم اليوتيوب، أما تطبيقات المحمول فلم تستخدمها إلا دولة واحدة فقط من ١٩ دولة، مما يدل على فقر هذه الأجهزة في الاستفادة من تكنولوجيا الاتصال التفاعلية، كما يوضح الشكل التالي:

(٢) تم الحصول على هذه الأرقام يوم ١٧ نوفمبر ٢٠١٣ من موقع ألكسا، وهذه الأرقام تتعرض للتغير اليومي، لذا لزم التنويه.

شكل رقم (١)
ترتيب الدول العربية وفق مؤشر الذكاء الرقمي للأجهزة الإحصائية الرسمية



من إعداد: الباحثة.

وتُشير بيانات الجدول والشكل السابقين إلى النتائج التالية:

- ١- احتلت الجزائر المرتبة الأخيرة في مؤشر الذكاء الرقمي للأجهزة الإحصائية الرسمية بين الدول العربية، وذلك بسبب أن الموقع الإلكتروني ما زال تحت الإنشاء، تليها سوريا نتيجة تعرض الموقع الإلكتروني للجهاز الإحصائي لمشاكل تقنية ربما بسبب أحداث الثورة التي تشهدها.
- ٢- باستبعاد الجزائر وسوريا فإن المواقع الإلكترونية للأجهزة الإحصائية الرسمية للدول العربية كانت بالترتيب التالي: جاءت الإمارات العربية في المركز الأول بقيمة ١,٢٢٥ في مؤشر الذكاء الرقمي؛ ويعود ذلك إلى الاحترافية العالية التي يبدو عليها الموقع الإلكتروني سواء من حيث التصميم البسيط الجذاب، واستخدام الخدمات التفاعلية والفلاشات المتحركة وتوفير معلومات هامة في الصفحة الرئيسية وسهولة التنقل في الموقع، إضافة إلى توافر الموقع باللغة الانجليزية والعربية، مع الاستخدام الأمثل للشبكات الاجتماعية كالفايس بوك وتويتر واليوتيوب، وكلها أدوات تسهل على المواطنين الحصول على البيانات بأشكال مختلفة، وبالتالي احتل الموقع الإلكتروني مرتبة متقدمة على موقع ألكسا، كما يشير عداد زوار الموقع إلى قيام ما يقارب مليون ونصف زائر بزيارة الموقع منذ إنشائه، فإذا أخذنا في الاعتبار أن العدد الكلي لسكان الإمارات لا يزيد عن ٩,٣ مليون نسمة في عام ٢٠١٢، أمكننا أن نستنتج مدى سهولة وصول المواطنين إلى خدمات الموقع.

وجاءت المملكة المغربية في المرتبة التالية في مؤشر الذكاء الرقمي؛ فعلى الرغم من عدم توفر الموقع الإلكتروني لجهاز الإحصاء المغربي إلا باللغة الفرنسية، إلا أن الموقع احتل مرتبة متقدمة

نتائج تطبيق مؤشر الذكاء الرقمي على أجهزة الإحصاء الرسمية العربية:
بعد حساب المؤشر المركب للذكاء الرقمي بكل مكوناته الفرعية، تم تطبيق هذا المؤشر على أجهزة الإحصاء الرسمية العربية، ويوضح كل من الجدول والرسم البياني التاليين ترتيب الدول وفق مؤشر الذكاء الرقمي للأجهزة الإحصائية العربية كما قامت الباحثة بتطبيقها:

جدول (٣)

نتائج مؤشر الذكاء الرقمي للأجهزة الإحصائية الرسمية في الدول العربية Digital IQ Index

م	الدولة	مؤشر الموقع الالكتروني	مؤشر التسويق الالكتروني	مؤشر الشبكات الاجتماعية	مؤشر المحمول	مؤشر الذكاء الرقمي
١	الإمارات العربية المتحدة	2.0008	0.7508	1	0	1.2252
٢	المملكة المغربية	1.7250	0.125	0.666	1	0.9776
٣	دولة فلسطين	1.600	0.907	0	0	0.9122
٤	دولة الكويت	1.7250	0.0494	0.666	0	0.8048
٥	جمهورية مصر العربية	1.4000	0.7169	0	0	0.775
٦	العراق	1.4250	0.6667	0	0	0.770
٧	البحرين	1.550	0.4104	0	0	0.743
٨	تونس	1.350	0.666	0	0	0.7400
٩	دولة قطر	1.35	0.012	0.666	0	0.6438
١٠	الجمهورية اليمنية	1.600	0.0796	0	0	0.6639
١١	السعودية	1.6	0.0768	0	0	0.6630
١٢	الأردن	1.400	0.339	0	0	0.6618
١٣	ليبيا	1.02	0.334	0.333	0	0.5604
١٤	سلطنة عمان	0.575	0.725	0	0	0.4476
١٥	الجمهورية اللبنانية	1.100	0.0002	0	0	0.4400
١٦	الجمهورية الإسلامية الموريتانية	0.6250	0.005	0	0	0.2517
١٧	السودان	0.575	0	0	0	0.2300
١٨	سوريا	1.25902E-05	0.0762	0	0	0.02289
١٩	الجزائر	1.11077E-06	0.0226	0	0	0.00679

المصدر: إعداد الباحثة.

المبحث الثاني

نتائج تطبيق المؤشر على الأجهزة الإحصائية الرسمية في العالم العربي

بعد استعراض المقصود بالذكاء الرقمي كأحد المؤشرات المركبة الحديثة وما يتضمنه من مؤشرات فرعية، قامت الدراسة بتطبيق المؤشر على أهم الأجهزة الإحصائية الرسمية في العالم العربي، وذلك وفق البيانات المتاحة حول هذه الأجهزة. لذا يعرض البحث أولاً لأهم أجهزة الإحصاء الرسمية العربية في ١٩ دولة، ثم بيان طريقة حساب مؤشر الذكاء الرقمي التي اتبعتها الباحثة، وأخيراً يشير البحث لأهم نتائج تطبيق مؤشر الذكاء الرقمي على أجهزة الإحصاء الرسمية العربية.

جدول (٢)

أهم الأجهزة الإحصائية الرسمية في العالم العربي ومواقعها الإلكترونية

م	الدولة	الجهاز الإحصائي	الموقع الإلكتروني
١	المملكة العربية السعودية	مصلحة الإحصاءات العامة والمعلومات	www.cdsi.gov.sa
٢	المملكة المغربية	المتدبيرة السامية للتخطيط	www.hcp.ma
٣	دولة البحرين	الجهاز المركزي للمعلومات	www.cio.gov.bh
٤	جمهورية مصر العربية	الجهاز المركزي للتعبئة العامة والإحصاء	www.capmas.gov.eg
٥	الإمارات العربية المتحدة	المركز الوطني للإحصاء	uaestatistics.gov.ae
٦	الجزائر	الديوان الوطني للإحصائيات (تحت الإنشاء)	www.ons.dz
٧	تونس	المعهد الوطني للإحصاء	www.ins.nat.tn
٨	الجمهورية اليمنية	الجهاز المركزي للإحصاء	www.Cso-yemen.org
٩	ليبيا	مصلحة الإحصاء والتعداد	www.bsc.ly
١٠	العراق	الجهاز المركزي للإحصاء وتكنولوجيا المعلومات	www.cosit.gov.iq
١١	الأردن	دائرة الإحصاء العامة	www.dos.gov.jo
١٢	دولة فلسطين	الجهاز المركزي للإحصاء	www.pcbs.gov.ps
١٣	دولة الكويت	الإدارة المركزية للإحصاء	www.csb.gov.kw
١٤	السودان	الجهاز المركزي للإحصاءات	www.cbs.gov.sd
١٥	الجمهورية الإسلامية الموريتانية	المكتب الوطني للإحصاء	www.ons.mr
١٦	سوريا	المكتب المركزي للإحصاء	www.cbssyr.org
١٧	سلطنة عمان	وزارة الاقتصاد الوطني	www.ncsi.gov.om
١٨	الجمهورية اللبنانية	إدارة الإحصاء المركزي	www.cas.gov.lb
١٩	دولة قطر	جهاز الإحصاء	www.qsa.gov.qa

المصدر: من إعداد الباحثة.

وقد تم إجراء خطوة تطبيع البيانات لكل المتغيرات محل الدراسة باستخدام طريقة إعادة القياس Re-scaling، وتعتمد هذه الطريقة على إيجاد القيمة المُطَبَّعة لكل مؤشر فرعي عن طريق طرح القيمة الصغرى للمؤشر من قيمة المؤشر، ثم قسمة الناتج على مدى المؤشر (الفرق بين القيمة العظمى والقيمة الصغرى)، وذلك لحساب القيمة المطلقة للمتغير (المؤشر الفرعي) (NARDO,2005) وفق المعادلة التالية:

$$\hat{X} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

أما طريقة تجميع المؤشر المركب النهائي فتم باستخدام طريقة التجميع الإضافي Additive aggregation methods (OECD, 2005) وفق المعادلة التالية:

$$\text{Digital IQ Index} = \sum W_i \hat{X}_i$$

حيث W_i هي الوزن الخاص بالمؤشر الفرعي i ، و $\hat{X}_i$ هي القيمة المُطَبَّعة للمؤشر الفرعي.

خريطة الموقع، ومدى توافر الموقع بعدة لغات، وغير ذلك مما قامت الباحثة بقياسه، ومن ثم يحصل موقع كل جهاز من الأجهزة الإحصائية العربية على درجة من ١ إلى ١٠.

أما التفاعل مع الجماهير فتم قياسه بترتيب الموقع الإلكتروني على موقع ألكسا Alexa الشهير^١، وكلما احتل الموقع ترتيباً متقدماً كلما حصل على درجة أكبر في هذا المؤشر، إضافة إلى قياس مدى توافر الخدمات التفاعلية ما بين الموقع والمتصفح من حيث إمكانية التواصل بين المتصفح والموقع، وتقديم تغذية راجعة إلى الموقع، وإمكانية قيام المتصفح بالتعليق أو تقديم استفسارات واستطلاعات الرأي والنشرات البريدية وغير ذلك من الخدمات التفاعلية، وأيضاً تم إعطاء كل موقع درجة من ١ إلى ١٠.

٢- مؤشر التسويق الإلكتروني Digital Marketing Index:

وتم قياسه بعدد مرات الظهور في محركات البحث (Google)، وتم إعطاء كل موقع درجة من ١ إلى ١٠، وأيضاً قيس مدى توافر بريد الكتروني للتواصل بين الموقع والمتصفح وإعطاء درجة (١) في حالة وجود البريد الإلكتروني ودرجة (صفر) في حالة عدم وجوده، وأخيراً استخدام أدوات تسويق الكتروني مبتكر من عدمه، وحصل كل موقع على درجة أيضاً ما بين (١) و(صفر) حسب استخدامه أساليب للتسويق الإلكتروني لمنتجات الجهاز الإحصائي من عدمه.

٣- مؤشر الشبكات الاجتماعية Social Media Index:

تم قياسه بدرجة استخدام الفيس بوك وتويتر واليوتيوب - المواقع الرسمية فقط- لتسويق خدمات الجهاز الإحصائي؛ بحيث يحصل كل جهاز على درجة تتراوح بين (الصفر) في حالة عدم استخدام إحدى الشبكات الاجتماعية، وبين (١) في حالة استخدام أي من هذه الشبكات، مع إعطاء ٥٠% للفيس بوك، و ٢٥% لكل من تويتر واليوتيوب.

٤- مؤشر المحمول Mobile Index:

وتم قياسه بمدى استخدام موقع الجهاز الإلكتروني لتطبيقات الهواتف الذكية وبرامج الآي باد، وتوافر نسخة محمولة يمكن للمتصفح من خلالها تصفح الموقع على هاتفه المحمول، وحصل موقع كل جهاز على درجة من (صفر) أو (١) في كل من هذين المتغيرين ذوي الأوزان المتساوية (مركز المعلومات ودعم اتخاذ القرار، ٢٠٠٦).

(١) ألكسا Alexa: هو موقع إلكتروني تابع لشركة أمازون، يقع مقره الرئيسي في ولاية كاليفورنيا بالولايات المتحدة، وهو متخصص في إحصائيات وترتيب مواقع الانترنت، وقد تم تأسيس موقع ألكسا سنة ١٩٩٦، وسُمي بهذا الاسم تكريماً لمكتبة الإسكندرية. ويحتل موقع جوجل المرتبة الأولى في ترتيب المواقع، يليه الفيس بوك ثم اليوتيوب، <http://www.alex.com/>.

ويمكن تلخيص هذه المؤشرات الفرعية في الجدول التالي:

جدول (١)

المؤشرات الفرعية التي يتكون منها مؤشر الذكاء الرقمي

م	المعيار	وزنه النسبي	ما يقيسه	المؤشرات الفرعية
١	الموقع الإلكتروني website	٤٠%	فعالية موقع المؤسسة الإلكتروني.	- الوظيفة والمحتوى: ٧٥% - التفاعل مع الجماهير: ٢٥% - ويقاس التفاعل مع الجماهير بمؤشرين: - ترتيب الموقع الإلكتروني على موقع ألكسا. - عدد الدقائق التي يقضيها المتصفح في الموقع.
٢	التسويق الإلكتروني Digital Marketing	٣٠%	فعالية جهود التسويق والتعريف بالعلامة التجارية.	- الظهور في محركات البحث. - البريد الإلكتروني. - التسويق الإلكتروني المبتكر.
٣	الشبكات الاجتماعية Social Media	٢٠%	كفاءة المؤسسة في استخدام الشبكات الاجتماعية.	- استخدام الفيس بوك: ٥٠% - استخدام تويتر: ٢٥% - استخدام اليوتيوب: ٢٥% - تطبيقات الهواتف الذكية.
٤	المحمول Mobile	١٠%	كفاءة استخدام المحمول.	- برامج أجهزة الآي باد. - توافر نسخة محمولة للموقع الإلكتروني.

المصدر: من إعداد الباحثة.

أما عن طريقة حساب مؤشر الذكاء الرقمي Digital IQ Index، فقد ذكرنا أن المؤشر يتكون من أربعة مؤشرات فرعية، ويوجد من بين هذه المؤشرات الفرعية ما يمكن قياسه بمقاييس رقمية محددة، ومنها ما يتم تقييمه بصورة تحكمية اعتماداً على رأي الباحث واستطلاع آراء الخبراء، وقد قامت الباحثة بحساب المؤشر للأجهزة الإحصائية العربية وفق المعطيات التالية (CURRAN,2013):

١- مؤشر الموقع الإلكتروني Website Index:

من حيث الوظيفة والمحتوى تم تقييم الموقع الإلكتروني بمدى شموليته للبيانات وإتاحتها للجمهور بسهولة، إلى جانب جاذبية الموقع واستخدام الفلاشات المتحركة والمتغيرة، وشمول الصفحة الرئيسية على أهم المعلومات الإحصائية عن الدولة، وسهولة التنقل داخل الموقع عبر

وحشد التأييد، والتسويق الإلكتروني وجمع التبرعات.

- مؤشر التفاعل مع الجماهير **Interactive Site**: ويمثل ٢٥% من فاعلية الموقع الإلكتروني، ويشمل التصميمات الجذابة والرسائل وسبل التواصل والتفاعل مع الزوار.

٢- مؤشر التسويق الإلكتروني **Digital Marketing**: ويشمل جهود التسويق، وإبراز العلامة التجارية عبر محركات البحث، ويتم ذلك من خلال تقنيات البحث، وعدد زيارات المستخدمين، والبريد الإلكتروني من حيث معدلات الإرسال، والمحتوى وتفعيل وسائل الإعلام الاجتماعية والدعاية والإعلان، وقياس الابتكار في وسائل التسويق الإلكتروني، سواء من خلال الإعلانات أو المسابقات وغيرها، وأخيراً من حيث تكرار الظهور في محركات البحث، ويمثل هذا المؤشر ٣٠% من إجمالي المعايير المركبة التي يتكون منها مؤشر الذكاء الرقمي.

٣- مؤشر الشبكات الاجتماعية **Social Media**: ويحتل المعيار الثالث نسبة ٢٠% من إجمالي المعايير ويختص بالشبكات الاجتماعية، ويقاس مدى حضور المؤسسة على مواقع الشبكات الاجتماعية، وعدد المعجبين والمتابعين ودرجة التأثير في المنابر الإعلامية الاجتماعية الأساسية. (CURRAN, 2013)

ويتم قياس كفاءة المؤسسة في استخدام الشبكات الاجتماعية من خلال ثلاث مقاييس:

- استخدام موقع الفيس بوك **Face book**: والذي يشغل ٥٠% من هذا المعيار، بقياس عدد المعجبين أو المتابعين، وعدد الإضافات والملاحظات والتطبيقات المستخدمة.

- استخدام موقع تويتر **Twitter**: الذي يأتي في المنزلة الثانية ليمثل ٢٥% من المعيار بقياس أعداد المتابعين وعدد الإدراجات ومعدلاتها.

- استخدام موقع اليوتيوب **YouTube**: والذي يتساوى مع تويتر في نفس النسبة (٢٥%)، ويقاس أعداد المشاهدين للفيديوهات المضافة من قبل المؤسسة على قنواتها، وعدد مرات التحميل، وعدد المقاطع المضافة، وعدد المسجلين في قنوات المؤسسة على اليوتيوب.

٤- مؤشر الهاتف المحمول **Mobile**: وضعت الدراسة نسبة ١٠% لمدى كفاءة استخدام المحمول، لقياس قدرة المؤسسة على استخدام تطبيقات المحمول وبرامج الهواتف وغيرها من الأجهزة النقالة في التسويق، ويتم قياس مدى توافر نسخة للموقع الإلكتروني متوافقة مع الجوال، ومدى استخدام تطبيقات الهواتف الذكية من حيث إتاحتها وشعبيتها وكفاءتها الوظيفية، وأخيراً مدى النجاح في استخدام تطبيقات أجهزة الآي باد في الدعاية والتسويق. (GALLOWAY, 2010)

ويعتبر "سكوت" هو مؤسس هذه المبادرة والأب الروحي لها، ويعمل "سكوت" أستاذاً مشاركاً بكلية إدارة الأعمال بجامعة نيويورك؛ حيث يقوم بتدريس إستراتيجيات العلامات التجارية، والتسويق الشبكي، وقد قام بتأسيس مركز أبحاث L2 للإبداع الرقمي، كما شارك في تأسيس شركة Firebrand، وهي شركة تنفيذية ناشطة استثمرت أكثر من مليار دولار أمريكي في مجال التسويق ووسائل الإعلام. (GALLOWAY, 2010)

ويقوم مركز أبحاث L2 للإبداع الرقمي بنشر تقارير سنوية لتصنيف العلامات التجارية في عدة مجالات تشمل صناعة السيارات، والماركات الفاخرة، وشركات السفر، والصيدلة، ومؤسسات الحكومة والقطاع العام، ووسائل الإعلام، وبطاقات الائتمان، ومتاجر البيع بالتجزئة، كما يقوم بنشر أبحاث ودراسات حول اتجاهات الصناعة، باستخدام التحليل والرسوم البيانية والإحصاءات وتعليقات الخبراء. (GALLOWAY, 2010)

ويعتبر مؤشر الذكاء الرقمي الذي وضعه "سكوت" هو الدراسة الأولى والأكبر من نوعها، التي تقدم معايير محددة لقياس كفاءة المؤسسة الرقمية؛ من حيث كفاءة وصول مرتادي الانترنت للموقع الإلكتروني للمؤسسة عبر محركات البحث، وقدرة المؤسسة على الاستفادة من التكنولوجيا الرقمية، وفاعلية الاستفادة من شبكات الانترنت الاجتماعية وشبكات المحمول.

وقامت الدراسة على فرضية أن التكنولوجيا الرقمية تلعب حالياً الدور الأكبر في تحديد أفكار الناس وتغيير اتجاهاتهم، وهو ما يجب أن تنتبه له جميع المؤسسات حكومية أو غير حكومية، ربحية أو غير ربحية؛ لذا فقد قامت الدراسة بتقييم مائة مؤسسة أمريكية من القطاع العام، وترتيبها من حيث مدى كفاءتها في استخدام التكنولوجيا الرقمية، كما هدفت لتوفير أداة قوية لتشخيص مواطن الضعف والقوة الرقمية في هذه المؤسسات لمساعدتها في تحديد أولوياتها لزيادة الاستثمار في مجال التكنولوجيا الرقمية. (GALLOWAY, 2010)

وقد جاءت وكالة الفضاء الأمريكية (ناسا) في مقدمة النتائج الخاصة بالمؤشر، تلاها البيت الأبيض، ثم منظمة بيثا لرعاية الحيوانات، ثم جيش الدفاع الأمريكي، ثم اللجنة الوطنية الديمقراطية، وصندوق الحياة البرية العالمي للولايات المتحدة، وكل هذه المؤسسات يمكن للأجهزة الإحصائية الرسمية العربية أن تستفيد من تجربتها الرائدة في كفاءة استخدام تكنولوجيا الاتصال والمعلومات الرقمية والتواجد بفاعلية في الفضاء الشبكي. (GALLOWAY, 2010)

المؤشرات الفرعية لمؤشر الذكاء الرقمي:

يتكون مؤشر الذكاء الرقمي من أربعة مؤشرات فرعية لقياس كفاءة المؤسسة في توظيف تكنولوجيا الاتصال الرقمي، وذلك على النحو التالي:

١- مؤشر الموقع الإلكتروني للمؤسسة website: ويقاس هذا المعيار فاعلية موقع المؤسسة الإلكتروني، وقد أعطاه "سكوت" وزناً نسبياً كبيراً بين باقي المؤشرات؛ حيث يبلغ وزنه ٤٠% من إجمالي معايير تقييم الذكاء الرقمي للمؤسسة، وتقاس فاعلية الموقع الإلكتروني بمقياسين أو مؤشرين فرعيين أساسيين:

- مؤشر الوظيفة والمحتوى Functionality & Content: وهذا المقياس يمثل ٧٥% من فاعلية الموقع الإلكتروني، ويمكن قياس مدى كفاءة الوظيفة والمحتوى من خلال التكنولوجيا المستخدمة في الموقع، وتقنيات البحث والتصفح للموقع، وخدمة العملاء، ومحتوى الموقع ودوره في التعليم والتوظيف والعضوية والشفافية والدعوة

المبحث الأول

الإطار النظري لمؤشر الذكاء الرقمي للأجهزة الإحصائية الرسمية

يُشير مفهوم الذكاء الرقمي Digital IQ إلى مدى فهم المؤسسات لقيمة التكنولوجيا الرقمية ودرجة اعتمادها عليها والاستفادة منها لتوسيع أنشطتها وأداء دورها على الوجه الأكمل، وتشمل التكنولوجيا الرقمية استخدام أدوات الانترنت الأساسية كالمواقع الإلكترونية وخدمات مواقع الشبكات الاجتماعية والبريد الإلكتروني وتطبيقات الهاتف المحمول وغيرها من الأدوات الرقمية (CURRAN, 2013)

ويُعد رفع معدل الذكاء الرقمي للمؤسسة من بين أهم الأهداف التي تسعى إليها المؤسسات المعاصرة؛ إذ تستهدف معظم المؤسسات تعزيز استخدام أساليب التكنولوجيا الرقمية وإيجاد قنوات لتلبية احتياجات العملاء، ومن بين هذه المؤسسات التي ينبغي عليها رفع معدل ذكاءها الرقمي الأجهزة الإحصائية الرسمية.

ويتمثل الهدف الرئيسي من وجود الأجهزة الإحصائية في تلبية احتياجات مستخدمي البيانات بكفاءة مستوياتهم؛ حيث يقوم الجهاز الإحصائي بجمع البيانات حول الظواهر المختلفة في الدولة، ومن ثم إتاحتها لمستخدمي البيانات؛ لذا فإن توسيع قاعدة مستخدمي البيانات يستلزم توسيع قاعدة البيانات المنتجة من جهة، وإيجاد علاقة تفاعلية تبادلية بين منتجي البيانات ومستخدميها من جهة أخرى. من هنا تأتي الحاجة لتبني الأساليب التكنولوجية الحديثة من قِبل هذه الأجهزة الإحصائية وتفعيل الاستفادة من تكنولوجيا الاتصال الرقمي، (المعهد العربي للتدريب والبحوث الإحصائية، ٢٠٠٥)، خاصة أنه في دولة مثل مصر بلغ عدد مستخدمي الانترنت ٣١,٢١ مليون نسمة في يونيو ٢٠١٢، من بينهم ٢,٤ مليون فرد فقط استخدموا الانترنت في الحصول على معلومات من الجهات الحكومية -وفقاً لبيانات الجهاز المركزي للتعبئة العامة والإحصاء-، مما يستوجب العمل على زيادة هذه النسبة بين مرتادي الانترنت.

وتُعد فكرة قياس الذكاء الرقمي للمؤسسات الرسمية أو الحكومية من الأفكار الحديثة، والتي لم يتم تناولها بالدراسة، وبالتالي لا يوجد مؤشر جاهز ومتخصص لقياس درجة الذكاء الرقمي للأجهزة الإحصائية الرسمية؛ لذا فإن البحث يعد محاولة أولى لم يسبق لها نظير من قبل، وذلك من خلال تقديم مقياس جديد يمكن من خلاله تقييم الذكاء الرقمي للمؤسسات الإحصائية ومدى كفاءتها في الاستفادة من التكنولوجيا الرقمية لتفعيل أدوارها ومهامها الأساسية.

ومع ذلك يمكن الاستعانة بأحد المؤشرات المركبة التي طرحها البروفيسور ستيرن جالاي سكوت Scott Galloway وآخرون لقياس مدى فاعلية المؤسسات الأمريكية في القطاع العام في استخدام وتوظيف التكنولوجيا الرقمية، وذلك كخطوة أولية، ثم محاولة تطوير المؤشر بما يتلائم مع عمل الأجهزة الإحصائية الرسمية.

المؤشر المركب للذكاء الرقمي:

في الثالث والعشرين من نوفمبر عام ٢٠١٠، قام مركز أبحاث L2 للإبداع الرقمي بجامعة جورج واشنطن الأمريكية بإصدار أول دليل إلكتروني لقياس مدى فاعلية المؤسسات في استخدام وتوظيف التكنولوجيا الرقمية، وقام المركز بإجراء دراسة على مائة مؤسسة أمريكية لقياس كفاءتها الرقمية، من خلال عدة معايير وضعها البروفيسور ستيرن جالاي سكوت وآخرون.

قياس الذكاء الرقمي للأجهزة الإحصائية الرسمية بالتطبيق على الدول العربية

أمل خيرى أمين محمد*

مقدمة:

كان لظهور شبكات الإنترنت الاجتماعية أكبر الأثر في تغيير سلوك المؤسسات الحكومية وغير الحكومية؛ حيث تعددت منصات الدعم والتوعية وتقديم المعلومات عبر أدوات إلكترونية بسيطة ساهمت في خفض التكاليف من ناحية، وتحقيق الانتشار الواسع للمؤسسات من ناحية أخرى، سواء كانت مؤسسات مجتمع مدني، ربحية أو غير ربحية، أو جمعيات خيرية، أو منظمات حكومية أو غير حكومية.

وتُعد الأجهزة الإحصائية الرسمية للدول من أهم المؤسسات التي ينبغي عليها الاستفادة من الفضاء الشبكي، ومواقع الشبكات الاجتماعية، وأدوات التدوين وتطبيقات المحمول، لما لدور هذه المؤسسات من أهمية في إتاحة البيانات ورفع الوعي الإحصائي بين المواطنين. ونظراً لتفاوت كفاءة المؤسسات بصفة عامة في الاستفادة من هذه التقنيات الحديثة؛ فقد أصبح من الضروري العمل على تطوير مؤشر إحصائي يمكن من خلاله اختبار ذكاء المؤسسة الرقمي، لقياس مدى فاعلية المؤسسة في الاستفادة من التكنولوجيا الرقمية في تحقيق أهدافها.

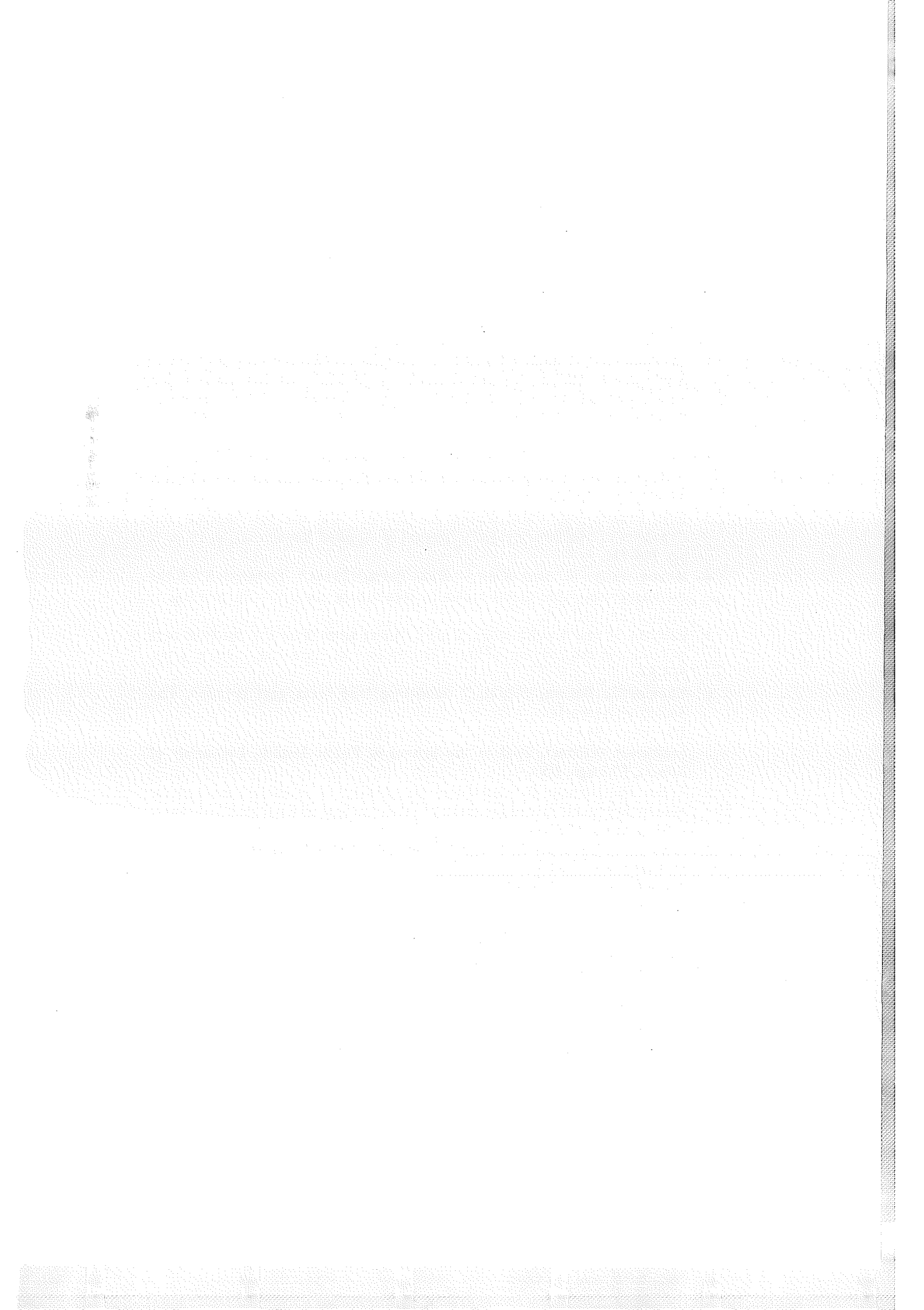
ويقدم هذا البحث محاولة لإيجاد منهجية جديدة لقياس الذكاء الرقمي للأجهزة الإحصائية الرسمية في العالم العربي، وتعتمد هذه المنهجية على أحد المؤشرات المركبة التي تم استحداثها لقياس الذكاء الرقمي للشركات والمؤسسات الخاصة، مع إدخال التعديلات اللازمة عليها لتتواءم مع طبيعة عمل الأجهزة الإحصائية الرسمية؛ بغرض قياس مدى فاعلية هذه الأجهزة في استخدام وتوظيف التكنولوجيا الرقمية.

ويعتمد البحث على الأسلوب الإحصائي الخاص بتكوين المؤشرات المركبة لتكوين المؤشر المركب الجديد لقياس الذكاء الرقمي للأجهزة الإحصائية العربية، إضافة إلى التحليل الوصفي لنتائج تطبيق المؤشر على أجهزة الإحصاء الرسمية العربية.

الكلمات الدالة:

الذكاء الرقمي – المؤشرات المركبة – التسويق الإلكتروني- الشبكات الاجتماعية.

*باحثة بالجهاز المركزي للتعبئة العامة والإحصاء – مصر، حاصلة على بكالوريوس الاقتصاد والعلوم السياسية جامعة القاهرة، وماجستير في الاقتصاد.



١. تقارير البنك الأهلي السعودي ، أعداد مختلفة.
٢. تقارير سوق التأمين السعودي، مؤسسة النقد السعودي، أعداد مختلفة.
٣. الكتاب الإحصائي السنوي لوزارة الصحة في المملكة العربية السعودية ، أعداد مختلفة.
٤. اللائحة التنفيذية (المعدلة) لنظام الضمان الصحي التعاوني بالمملكة العربية السعودية، المقر في جلسة رقم (٧٣) وتاريخ ١٤٣٠/٥/٨ هـ.

ثانياً- المراجع الاجنبية

1. Brumley R., Enguidanos S., Jamison P., Seitz R., Morgenstorn N., Saito S. Mellwane J., Hillary K. and Gonzalez J. (2007)," Increase satisfaction with care and lower costs: Results of a randomized trial of in- home Palliative care", *Journal of the American Geriatrics Society*, v. 55, n. 7, pp. 993-1000.
2. Chen Shen,(2013)," Determinations of healthcare decisions: Insurance, Utilization, and Expenditure", *The Review of Economics and Statistics*, v. 95, n.1, pp. 142-153.
3. Eric French and John Bailly J., (2004)," On the distribution and dynamics of healthcare costs", *Journal of Applied Economics*, v. 19, n. 6, pp. 705-721.
4. Huixia Judy W. and Xiao – Hua Zhou, (2010)," Estimation of the retransformed conditional mean in Healthcare studies", *Biometrika*, v. 97, n. 1, pp. 147-158.
5. Riley G. F. and Lubitz J. D., (2010), "Long term trends in medical care payments in the last year of life ", *Health Service Research*, v. 45, n. 2, pp. 565-576.
6. Thomas E. Getzen, (2007), " Modeling long term cost trends", *The Society of Actuaries*.
7. Tia Goss Sawhney, (2012), " Controlling indirect selection under healthcare reform", PhD.,in the School of Public Health, Univeristy of Illinois, Chicago, USA.
8. Tolley H., Michael D., and Pater D., (1989)," A statistical method for monitoring social insurance claims", *The Journal of Risk and Insurance*, v. 56, n. 4, pp. 670-685.
9. Xiao – Hua Zhou, Kevin T. and William M., (2001), " Regression Analysis of healthcare charges with heteroscedasticity", *Journal of the Royal Statistical Society Series C*, v. 50, n. 3, pp. 303-312.

٣- بدراسة العلاقة الاتجاهية بين مؤشرات نظم تكاليف الخدمات الصحية المقدمة والزمن في الفترة من ١٤١٥ - ١٤٣٢ من خلال اختبار أشكال الدوال المختلفة، أمكن تقدير نماذج للعلاقة الاتجاهية لمؤشرات نظم تكاليف الخدمات الصحية حيث تنوعت بين الدوال اللوغاريتمية والأسية و التربيعية والتكعيبية والانحدار الذاتي حيث تراوح معامل R- Square لهم بين ٤١,٤% و ٨٧% مما أمكن الاعتماد عليهم في تقدير العلاقة الاتجاهية واستخدامها في إيجاد القيم المتوقعة لها في السنوات القادمة بما يساعد في إعداد الموازنات وتقدير التزامات نظم التأمين الصحي بالمملكة العربية السعودية.

ثانياً- التوصيات :

- ١- الاستفادة من نتائج تقييم خدمات الرعايا الصحية المقدمة في إعادة النظر في نظم إدارة المستشفيات الحكومية مما يساعد على تطويرها و تشجيع القطاع الخاص على الاستثمار في مجال تقديم الخدمات الصحية حتى يمكن رفع مستوى جودة خدمات الرعايا الصحية المقدمة.
- ٢- تقدير الاحتياجات وأعداد الموازنات لخدمات الرعايا الصحية باستخدام الدوال الإحصائية التي تم التوصل إليها، بما يساعد في إعدادها على أساس علمي و التوجيه السليم لأموال موازنة الخدمات الصحية بما يعظم من المنافع المحققة منها.
- ٣- تفعيل دور الاشراف والرقابة على نشاط التأمين لشركات التأمين الصحي العاملة بالسوق السعودي، لضمان تحقيق معايير موحدة للممارسة وإدارة الخدمات الصحية ضمن برامج التأمين الصحي التعاوني لضمان عدم التباين في أسعار التأمين الصحي.
- ٤- تقوم الدولة بتقديم نظام شامل للتأمين الصحي مع العمل على تطوير وزيادة فعالية نظام للتأمين الصحي التعاوني من خلال تشجيع و إعطاء مزايا للشركات العاملة في سوق التأمين السعودي.

المراجع

اولاً- المراجع العربية

أ- البحوث العلمية

١. أحمد حسين ناصر الخليفي، (٢٠٠٥)، "دوافع التعامل مع شركات التأمين التعاوني وأثرها علي درجة الرضا لدى طالبي الخدمة الصحية التعاونية: دراسة تطبيقية على محافظة جدة بالمملكة العربية السعودية"، رسالة ماجستير كلية الاقتصاد والإدارة ، جامعة الملك عبد العزيز.
٢. تغريد خليل محمد اشتيوي (٢٠٠٦)، "التحديات المتوقعة والحلول المقترحة لتطبيق التأمين الصحي التعاوني في المملكة العربية السعودية - دراسة استكشافية بمدينة الرياض"، رسالة ماجستير ، كلية إدارة الأعمال ، جامعة الملك سعود.
٣. رأفت أحمد ابراهيم وعبدالله عمر النجار، (٢٠٠٨) " التحليل الكمي لمؤشرات قياس المستوى الصحي لتطبيق نظام التأمين الصحي بالمملكة العربية السعودية"، مجلة المصرية للدراسات التجارية جامعة المنصورة.
٤. صالح ناصر العمير، (٢٠٠٢)، " التأمين الصحي وأثره على الاقتصاد السعودي حتى عام ٢٠٢٠ م"، ندوة الرؤية المستقبلية للاقتصاد السعودي حتى عام ٢٠٢٠ م ، الرياض ، ١٩ - ٢٣ أكتوبر ٢٠٠٢ م .

ب- اللوائح والتقارير

جدول (١١)

القيم المتوقعة لمؤشرات نظم تكاليف الخدمات الصحية المقدمة في السنوات القادمة

م	المؤشر	سنة ١٤٣٤هـ	سنة ١٤٣٥هـ	سنة ١٤٣٦هـ
١	عدد الأسرة	3670	3912	4169
٢	متوسط وجبات المرضى لكل سرير في السنة	219.6159	222.2672	224.9505
٣	متوسط عدد الأفلام للمريض في السنة	1.150879	1.140828	1.131245
٤	الفحوص المخبرية في السنة	6146123	6284642	6423160
٥	عدد العمليات الجراحية العامة في السنة	9597	10618	11731
٦	عدد عمليات العيون في السنة	8453	9768	11191
٧	عدد عمليات الأنف والأذن في السنة	3031	3353	3701
٨	عدد عمليات الفك والأسنان في السنة	534	538	535
٩	عدد المنومون سنوياً	122109	125500	128891
١٠	معدل دوران السرير عامة في السنة	62.44	68.741	75.642
١١	معدل دوران السرير نساء في السنة	153.0251	158.4758	164.1206
١٢	متوسط مدة الإقامة عامة في السنة	4.823	3.979	2.923
١٣	متوسط مدة الإقامة نساء في السنة	2.392072	2.412923	2.429574
١٤	نسبة اشغال الأسرة عامة في السنة	0.530467	0.532547	0.534628
١٥	نسبة اشغال الأسرة نساء في السنة	0.837	0.89398	0.96364
١٦	عدد المراجعين للعيادات الخارجية في السنة	3670535	3686141	3686619

النتائج والتوصيات

أولاً - النتائج

- ١- عدم وجود فروق معنوي راجعة للعوامل الديموجرافية لمستقصي منهم في تقييمهم لخدمات الرعاية الصحية المقدمة ،حيث أن الاختلاف الوحيد الموجود كان يرجع إلى الحالة الاجتماعية في تقييمها للأداء الصحي الحكومي وهذا يعني وجود اتفاق كبير نسبياً بين مفردات الدراسة في آرائهم تجاه تقييم خدمات الرعاية الصحية مما يؤكد علي أهمية الأخذ بنتائج هذا التقييم.
- ٢- ومن أهم نتائج الدراسة الميدانية هو أنه على الرغم مما تم صرفه من موارد مالية وجهود بشرية نحو تحسين الخدمات الصحية في المملكة خلال العقود الماضية إلا أن مستوى جودة هذه الخدمات لم يرق إلى المستوى المطلوب ويرجع ذلك إلي:
الصعوبة في الحصول على فرص علاج في المستشفيات الحكومية حيث بلغت نسبة الاستجابة لها ٧٠% تقريباً وطول انتظار المواعيد عند العلاج في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٨٠,٣%. وطول انتظار صرف الوصفة الدوائية من الصيدلية في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٦٠,٥%. بالإضافة إلى تؤدي الوساطة والمحسوبية دوراً في حصول بعض المرضى على فرص علاج في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٧٩,٤٥%. وينعكس ذلك على انخفاض مستوى جودة الخدمات الصحية المقدمة.

جدول (١٠)

يوضح نماذج العلاقة الاتجاهية بين مؤشرات نظم تكاليف الخدمات الصحية المقدمة والزمن

م	المؤشر	نموذج الانحدار مع الزمن	قيمة F	معنوية F	R ²
١	عدد الأسرة	$y_1 = 1913.578 - 59.475t + 7.364t^2$	29.535	0.000	0.82
٢	متوسط وجبات المرضى لكل سرير في السنة	$y_2 = 172.756 e^{0.012t}$	12.861	0.003	0.479
٣	متوسط عدد الأفلام للمريض في السنة	$y_3 = 1.768 - 0.206 \ln(t)$	20.126	0.001	0.59
٤	الفحوص المخبرية في السنة	$y_4 = 3375749.25 + 138518.691t$	9.098	0.007	0.414
٥	عدد العمليات الجراحية العامة في السنة	$y_5 = 8462.466 - 861.675t + 49.72t^2$	8.964	0.004	0.58
٦	عدد عمليات العيون في السنة	$y_6 = 2835 - 613.5t + 40.42t^2 + 0.215t^3$	17.084	0.000	0.81
٧	عدد عمليات الأنف والأذن في السنة	$y_7 = 2347 - 239.4t + 13.68t^2$	17.125	0.000	0.724
٨	عدد عمليات الفك والأسنان في السنة	$y_{8t} = 696.5984 - 0.30225y_{8t-2}$	22.314	0.000	0.76
٩	عدد المنومون سنوياً	$y_9 = 54289 + 3391t$	36.130	0.000	0.72
١٠	معدل دوران السرير عامة في السنة	$y_{10} = 62.42 - 5.999t + 0.300t^2$	9.009	0.002	0.603
١١	معدل دوران السرير نساء في السنة	$y_{11} = 75.99 e^{0.035t}$	27.725	0.002	0.664
١٢	متوسط مدة الإقامة عامة في السنة	$y_{12} = 4.903 - 0.464t + 0.083t^2 - 0.003t^3$	7.001	0.006	0.663
١٣	متوسط مدة الإقامة نساء في السنة	$y_{13t} = 1.113587 + 0.550145 y_{13t-2}$	8.23	0.001	0.517
١٤	نسبة اشغال الأسرة عامة في السنة	$y_{14t} = 0.30417 + 0.4001 y_{14t-2}$	15.46	0.021	0.827
١٥	نسبة اشغال الأسرة نساء في السنة	$y_{15} = 0.00018^3 - 0.005t^2 + 0.035t + 0.697$	14.218	0.000	0.87
١٦	عدد المراجعين للعيادات الخارجية في السنة	$y_{16t} = 3117331.625 + 0.1550965y_{16t-2}$	8.627	0.001	0.613

يتضح لنا مما سبق أن:

أن معنوية اختبار F لجميع معادلات الانحدار المقدرة أقل من ٠,٠٥ مما يدل على معنوية معادلة الانحدار المقدرة كما بلغت نسبة تفسير الزمن للتغير في مؤشرات نظم تكاليف الخدمات الصحية المقدمة تتراوح ما بين ٤١,٤% و ٨٧% مما يؤكد على إمكانية الاعتماد على هذه المعادلات في التنبؤ بقيم مؤشرات نظم تكاليف الخدمات الصحية المقدمة في السنوات القادمة بما يساهم في مساعدة متخذ القرار في وضع تصور عن إجمالي الاحتياجات المطلوبة لبناء نظام للتأمين الصحي للمنطقة الشرقية بالمملكة العربية السعودية.

دراسة العلاقة الاتجاهية لمؤشرات نظم تكاليف الخدمات الصحية المقدمة

يتم دراسة العلاقة بين مؤشرات نظم تكاليف الخدمات الصحية المقدمة والزمن لنصل إلى معادلة الاتجاه العام لاستخدامها في التنبؤ مستقبلاً، والاعتماد على هذا التنبؤ في بالتكاليف المتوقعة للتأمين الطبي والتي تساهم في تحديد الأقساط و التعويضات المتوقعة لنظام التأمين الطبي المتمثلة في :

م	مؤشرات نظم تكاليف الخدمات الصحية	م	مؤشرات نظم تكاليف الخدمات الصحية	م	مؤشرات نظم تكاليف الخدمات الصحية
١	عدد الأسرة	٧	عدد عمليات الأنف والأذن في السنة	١٢	متوسط مدة الإقامة عامة في السنة
٢	متوسط وجبات المرضى لكل سرير في السنة	٨	عدد عمليات الفك والأسنان في السنة	١٣	متوسط مدة الإقامة نساء في السنة
٣	متوسط عدد الأفلام للمريض في السنة	٩	عدد المنومون سنوياً	١٤	نسبة اشغال الأسرة عامة في السنة
٤	الفحوص المخبرية في السنة	١٠	معدل دوران السرير عامة في السنة	١٥	نسبة اشغال الأسرة نساء في السنة
٥	عدد العمليات الجراحية العامة في السنة	١١	معدل دوران السرير نساء في السنة	١٦	عدد المراجعين للعيادات الخارجية في السنة
٦	عدد عمليات العيون في السنة				

بدراسة العلاقة الاتجاهية بين مؤشرات نظم تكاليف الخدمات الصحية المقدمة والزمن في الفترة من ١٤١٥ - ١٤٣٢ من خلال اختبار أشكال الدوال المختلفة والمتمثلة في : الخطية Linear والتربيعية Quadratic والتكعيبية Cubic واللوغارتمية Logarithmic والأسية Exponential والقوى Power والعكسية Inverse . فكانت النتائج تشير إلى أفضل نماذج للداول هي على الشكل التالي:

المستشفيات الحكومية بجودة ومواصفات صحية عالية بلغت نسبة الموافقة عليه ٥٩% وتمتاز المستشفيات الحكومية بطاقم طبي متميز قادر على فهم حاجات المرضى بلغت نسبة الموافقة عليه ٥٩%. فأهمية تدريب الكوادر البشرية والطبية منها على وجه الخصوص سوف تبرز بشكل ملحوظ عند البدء في تطبيق الضمان الصحي التعاوني، حيث أن التدريب المستمر للفريق الطبي سيرفع من كفاءتهم ومهاراتهم الطبية، في حين عدم التدريب سيؤدي مع مرور الوقت إلى ارتكاب أخطاء طبية فادحة قد تكون كبيرة النتائج مما سيؤدي إلى مشاكل إدارية ومالية بين أطباء القطاع ومستشفياتهم.

ومما سبق يتضح لنا وجود حاجة ضرورية و ملحة لنظام تأمين الصحي بالمنطقة الشرقية بالمملكة العربية السعودية

المبحث الثالث: تقدير القيم الإتجاهية مؤشرات نظم تكاليف الخدمات الصحية المقدمة

هناك العديد من الدراسات التي تناولت مؤشرات نظم تكاليف خدمات الرعاية الصحية وأثرها على نواحي عديدة والتي تم فيها استخدام نماذج إحصائية مختلفة منها:

دراسة Huixia وآخرين (٢٠١٠)، تم دراسة تكاليف الرعاية الصحية باستخدام نماذج الانحدار وذلك في حالة وجود تحويلات لدوال تكاليف الرعاية الصحية، حيث تم تقديم مقدر لتوزيع حر لمتوسط تكاليف الرعاية الصحية وذلك مع علاج مشكلة التجانس Heteroscedstic لإيجاد مقدرات جيدة لمعالم دوال تكاليف الرعاية الصحية. دراسة Eric and John (٢٠٠٤)، تم استخدام العمليات العشوائية stochastic process في تحديد توزيع تكاليف الرعاية الصحية وديناميكيته. ووجدوا أن توليد البيانات اللوغاريتمية لتكاليف الرعاية الصحية يمكن التعبير عنه من خلال دوال الانحدار الذاتي من الدرجة الأولى $AR(1)$.

دراسة Xiao وآخرين (٢٠٠١) وفيها تم اختبار تأثير تكاليف الدواء المستخدم وخصائص المرضى على تكاليف الرعاية الصحية المقدمة لكلاً من مرضى العيادات الخارجية والمرضى المنومون داخل المستشفى وذلك من خلال استخدام نماذج الانحدار مع الإخذ للتحويل اللوغاريتمية لتكاليف الرعاية الصحية وذلك في ظل عدم ثبات التباين Heteroscedstic. دراسة Tolley وآخرين (١٩٨٩)، قدمت طريقة إحصائية لعمل مراقبة على تكاليف الرعاية الصحية من خلال استخدام النماذج الإحصائية لمراقبة الجودة. حيث تم استخدام خرائط مراقبة الجودة في الحكم على تكاليف الرعاية الصحية ومراقبتها. دراسة Chen (٢٠١٣)، تم دراسة ثلاث عوامل متداخلة وهي التأمين والاستخدام والأنفاق كمحددات لقرارات الرعاية الصحية. وضع نموذج يعتبر أن التأمين متغير داخل endogenous variable ومتعلق بكلاً من الاستخدام والأنفاق ومنه توقع بأن تأمين الرعاية الصحية سيزيد من النفقات بنسبة ٤٨% من خلال الاعتماد على استبيان لنفقات الرعاية الصحية. دراسة Thomas (٢٠٠٧) ودراسة Riley وآخرين (٢٠١٠)، تم عرض كيفية تأثير الزيادة في تكاليف الخدمات الصحية على نظم الرعاية الصحية من حيث تأثيرها في كلاً من التزام منظمي خدمات الرعاية الصحية وأقساط النظام وكيفية التنبؤ بها.

كما عرض Tia (٢٠١٣)، كيف أن إجراء الاختيار الصحي غير المباشر لطالبي خدمات الرعاية الصحية يساعد منظمي خدمة الرعاية الصحية على التحكم في تكاليف الرعاية الصحية. ودراسة Brumley وآخرين (٢٠٠٧) بين فيها كيفية زيادة درجة رضا متلقي خدمات الرعاية الصحية وذلك عند مستويات أقل من تكاليف الرعاية الصحية المقدمة.

- الحكم على جودة خدمات الرعاية الصحية الحكومية المقدمة والوقوف على العوامل المؤثرة في رفع مستوى جودة هذه الخدمات.

١- الإداء الصحى الحكومى بلغت نسبة الموافقة عليه ٦٩% ويرجع ذلك إلى أن المؤسسات الصحية الحكومية يتم إدارتها بأسلوب الإدارة العامة الملتزم بالنظم الحكومية في إجراءاتها الإدارية والمالية وذلك لأنها في الأساس لا تهدف للربح بل تضطلع بمهام اجتماعية بحتة تتمثل في تقديم خدمات الصحة الوقائية وتوفير العلاج المجاني للمواطنين وتنفيذ حملات التوعية الصحية.

٢- هناك زيادة في الطلب على الخدمات الطبية وينعكس ذلك من خلال السؤال عند الشعور بوعكة صحية يختصر الذهاب إلى طبيب وأذهب مباشرة إلى الصيدلي بلغت نسبة الاستجابة ٥٥,٥٤%. وبالتالي أصبح من الضروري إتاحة الفرصة لنظام التأمين الطبي التعاوني للعمل بفعالية أكبر لمساندة الجهات الصحية المختلفة في توفير التمويل اللازم للتوسع في المنشآت الصحية لتلبية الاحتياجات المستقبلية للسكان في المملكة.

٣- الارتفاع في تكاليف تقديم الرعاية الصحية و يتضح ذلك من هناك ضرورة نحو قيام وزارة الصحة بتحديد أسعار ثابتة للخدمات الطبية في القطاع الخاص حيث بلغت نسبة الموافقة على ذلك ٨٧,٧٣%. و عدم مناسبة أسعار وتكاليف العلاج في القطاع الخاص للمرضى والمحتاجين للعلاج حيث بلغت نسبة الموافقة عليه ٤٧,١٨%. وأن المستشفيات و المستوصفات الخاصة تبالغ في تحميل المريض تكاليف مالية غير ضرورية بهدف تعظيم الربح المادي فقط حيث بلغت نسبة الموافقة عليه ٧٦,٤%.

٤- على الرغم مما تم صرفه من موارد مالية وجهود بشرية نحو تحسين الخدمات الصحية في المملكة خلال العقود الماضية إلا أن مستوى جودة هذه الخدمات لم يرق إلى المستوى المطلوب، حيث نجد أن هناك توافر للخدمات الطبية الحكومية المتكاملة حيث بلغت نسبة الاستجابة لها ٧١,٥٦%. كما تتوفر في المنطقة أكثر التخصصات الطبية حيث بلغت نسبة الاستجابة لها ٧٥,٣%. ولا يوجد اضطراب للسفر خارج المدينة أو المملكة لطلب العلاج حيث بلغت نسبة الاستجابة لهما ٥٩ و ٥٢%. إلا أن وجود صعوبة في الحصول على فرص علاج في المستشفيات الحكومية حيث بلغت نسبة الاستجابة لها ٧٠% تقريباً وطول انتظار المواعيد عند العلاج في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٨٠,٣%. وطول انتظار صرف الوصفة الدوائية من الصيدلية في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٦٠,٥%. بالإضافة إلى تؤدي الوساطة والمحسوبية دوراً في حصول بعض المرضى على فرص علاج في المستشفيات الحكومية حيث بلغت نسبة الموافقة عليه ٧٩,٤٥%. وينعكس ذلك على انخفاض مستوى جودة الخدمات الصحية المقدمة.

٥- إن النهوض بمستوى الخدمات الصحية في المملكة يتطلب بطبيعة الحال الإحلال الدائم لكافة الأجهزة الطبية واستخدام الأحدث منها لتقديم الخدمات العلاجية بمستويات راقية. ولا يمكن أن نصل إلى الطاقة التشغيلية القصوى لتلك الأجهزة إلا إذا كانت لدى المستشفيات والمستوصفات كوادر بشرية مدربة على تشغيل تلك الأجهزة بكفاءة، وهذا ما يفتقده معظم وحدات القطاع. وتعد عملية تدريب العمالة في المجال الصحي من الأنشطة التي لا توليها مستشفيات القطاع الخاص نفس الاهتمام الذي تقدمه المستشفيات الحكومية، وهذا يفرض ضرورة وضع قواعد تلزم مقدمي الخدمة الطبية بكافة قطاعاتها بوضع برامج لتدريب العاملين فيها وتجهيزهم لاستخدام أحدث التكنولوجيا المستخدمة في تقديم الخدمات العلاجية. ويتضح ذلك من يتصف الطاقم الطبي في المستشفيات الحكومية بمهنية عالية بلغت نسبة الموافقة عليه ٦٣,٨%. ويتصف الطاقم الإداري في المستشفيات الحكومية بمستوى عالٍ في الخدمة بلغت نسبة الموافقة عليه ٦٢%. وتمتاز منشآت

جدول (٨)

يوضح اجراء اختبار Kruskal Wallis لدراسة وجود اختلاف معنوي في تقييم الرعاية الصحية المقدمة في المنطقة الشرقية على حسب المستوى التعليمي

Test Statistics^{a,b}

العناصر	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Chi-Square	1.860	.650	.781	1.601	.402
Df	3	3	3	3	3
Asymp. Sig.	.602	.885	.854	.659	.940

a. Kruskal Wallis Test b. Grouping Variable: EDUCATION

يتضح لنا من الجدول السابق أن :

لا يوجد تأثير للمستوي التعليمي على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية دراسة الفرض :

" يوجد تأثير معنوي لمستوى الدخل الشهري في عملية تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

جدول (٩)

يوضح اجراء اختبار Kruskal Wallis لدراسة وجود اختلاف معنوي في تقييم الرعاية الصحية المقدمة في المنطقة الشرقية على حسب مستوى الدخل الشهري

Test Statistics^{a,b}

العناصر	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Chi-Square	1.860	.650	.781	1.601	.402
Df	3	3	3	3	3
Asymp. Sig.	.602	.885	.854	.659	.940

a. Kruskal Wallis Test

b. Grouping Variable: EDUCATION

يتضح لنا من الجدول السابق أن :

لا يوجد تأثير للمستوي الدخل الشهري على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية .

ومما سبق يتضح لنا عدم وجود فروق معنوي راجعة للعوامل الديموجرافية للمستقصي منهم في تقييمهم لخدمات الرعاية الصحية المقدمة ،حيث أن الاختلاف الوحيد الموجود كان يرجع إلى الحالة الاجتماعية في تقييمها للأداء الصحي الحكومي وهذا يعني أن هناك اتفاق كبير نسبياً بين مفردات الدراسة في ارائهم تجاه تقييم خدمات الرعاية الصحية مما يؤكد علي أهمية الأخذ بنتائج هذا التقييم.

جدول (٦)

يوضح اجراء اختبار Mann-Whitney U لدراسة وجود تأثير معنوي للحالة الاجتماعية علي تقييم الرعاية الصحية المقدمة في المنطقة الشرقية

Test Statistics^a

	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Mann-Whitney U	6772.500	8037.500	7294.500	7476.000	7140.500
Wilcoxon W	13213.500	18333.500	13735.500	13917.000	13581.500
Z	-2.231-	-.072-	-1.337-	-1.030-	-1.597-
Asymp. Sig. (2-tailed)	.026	.943	.181	.303	.110

a. Grouping Variable: SCOTIAL

توضح لنا من الجدول السابق أن :
يوجد تأثير معنوي للحالة الاجتماعية على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية وذلك من خلال الاختلاف في الأداء الصحي الحكومي المقدم من وزارة الصحة فقط.
دراسة الفرض :

" يوجد تأثير لنوع قطاع العمل عند تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

جدول (٧)

يوضح اجراء اختبار Kruskal Wallis لدراسة وجود اختلاف معنوي في تقييم الرعاية الصحية المقدمة في المنطقة الشرقية على حسب نوع قطاع العمل

Test Statistics^{a,b}

العناصر	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Chi-Square	.464	4.960	4.919	1.234	.731
Df	3	3	3	3	3
Asymp. Sig.	.927	.175	.178	.745	.866

a. Kruskal Wallis Test b. Grouping Variable: WORK

يتضح لنا من الجدول السابق أن :
لا يوجد تأثير لنوع قطاع العمل على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية.

دراسة الفرض: " لا يوجد اختلاف معنوي للمستوى التعليمي في عملية تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

دراسة الفرض: " يوجد تأثير معنوي لجنس الشخص على تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

جدول (٤)

يوضح اجراء اختبار Mann-Whitney U لدراسة وجود تأثير معنوي لجنس الشخص علي تقييم الرعاية الصحية المقدمة في المنطقة الشرقية

Test Statistics ^a					
	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Mann-Whitney U	7315.500	7753.000	7790.500	7955.000	7757.500
Wilcoxon W	18490.500	13531.000	18965.500	19130.000	18932.500
Z	-1.127	-.375	-.310	-.028	-.366
Asymp. Sig. (2-tailed)	.260	.707	.756	.977	.714

a. Grouping Variable: SEX

يتضح لنا من الجدول السابق أن :

لا يوجد تأثير لجنس الشخص على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية.

دراسة الفرض : " هناك اتفاق في تأثير جنسية الشخص عند تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

جدول (٥)

يوضح اجراء اختبار Mann-Whitney U لدراسة وجود تأثير معنوي لجنسية الشخص علي تقييم الرعاية الصحية المقدمة في المنطقة الشرقية

Test Statistics ^a					
	الاداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الاداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الاداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Mann-Whitney U	3917.000	3845.500	3930.500	3654.500	3884.000
Wilcoxon W	28227.000	4511.500	28240.500	4320.500	4550.000
Z	-.105	-.279	-.072	-.744	-.185
Asymp. Sig. (2-tailed)	.917	.780	.943	.457	.854

a. Grouping Variable: NATIONAL

يتضح لنا من الجدول السابق أن :

لا يوجد تأثير لجنسية الشخص على عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية.

دراسة الفرض : " هناك اتفاق في تأثير للحالة الاجتماعية عند تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية "

يتضح من الجدول السابق أن :

وجود صدق عالي لدى الأفراد المستقصى منهم حيث بلغ معامل الصدق للمتغيرات البحثية وإجمالي أسئلة الاستمارة ٨٧ % وكذلك ارتفاع قيمة معامل الثبات لسبيرمان بروان حيث بلغ ٠,٦٥٧ و نتائج اختبار كوكران ٢١ كما لجميع المتغيرات محل الدراسة حيث بلغت معنوية الاختبار لهم أقل من ٠,٠٥ مما يدل على وجود فروق معنوية في إجابات المستقصى منهم. وبالتالي يمكن الاعتماد على البيانات الواردة من قائمة الاستقصاء تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية.

دراسة اختباريه لفروض الدراسة

يتم اختبار فروض الدراسة التالية:

- أ- لا يوجد اختلاف معنوي في تقييم خدمات الرعاية الصحية المقدمة على حسب الفئة العمرية
- ب- يوجد تأثير معنوي لجنس الشخص على تقييم خدمات الرعاية الصحية المقدمة
- ت- هناك اتفاق في تأثير جنسية الشخص عند تقييم خدمات الرعاية الصحية المقدمة
- ث- هناك اتفاق في تأثير الحالة الاجتماعية عند تقييم خدمات الرعاية الصحية المقدمة
- ج- يوجد تأثير لنوع قطاع العمل عند تقييم خدمات الرعاية الصحية المقدمة
- ح- لا يوجد اختلاف معنوي للمستوى التعليمي في عملية تقييم خدمات الرعاية الصحية المقدمة
- خ- يوجد تأثير معنوي لمستوى الدخل الشهري على تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية

وحتى يمكن دراسة هذا الفروض لابد من دراسة المكونات الفرعية المتعلقة بمكونات تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية كما يتضح من الملحق :

دراسة الفرض : "لا يوجد اختلاف معنوي في تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية على حسب الفئة العمرية"

جدول (٣)

يوضح اجراء اختبار Kruskal Wallis لدراسة وجود اختلاف معنوي في تقييم الرعاية الصحية المقدمة في المنطقة الشرقية على حسب الفئة العمرية

Test Statistics^{a,b}

العناصر	الأداء الصحي الحكومي المقدم من وزارة الصحة	توافر الخدمات الصحية بالمنطقة	الأداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	الأداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	إجمالي تقييم خدمات الرعاية الصحية بالمنطقة الشرقية
Chi-Square	9.141	2.201	1.314	4.006	1.954
Df	4	4	4	4	4
Asymp. Sig.	.058	.699	.859	.405	.744

a. Kruskal Wallis Test b. Grouping Variable: AGE

يتضح لنا من الجدول السابق

لا يوجد تأثير للفئة العمرية علي عناصر تقييم الخدمات الصحية المقدمة بالمنطقة الشرقية.

أدوات التحليل الإحصائي :

خضعت البيانات البحث لخطه تحليل إحصائي متعددة المستويات تتفق معها ، وتهدف إلى تحقيق أهداف البحث والتحقق من صدق فروضها وقد تم تحليل هذه البيانات باستخدام برنامج : (Statistical package for social science) المعروف باسم (spss) .

وقد تطلبت طبيعة البيانات تحديد الأدوات الإحصائية اللازمة والملائمة والتي تتمثل في الآتي :

أ- المقاييس الوصفية Descriptive Measures.

تستخدم لوصف البيانات البحثية من حيث القيم التي تتركز فيها البيانات و المتمثلة في مقاييس النزعة المركزية و التي يمثلها الوسط الحسابي .بالأضافة الى مقاييس التشتت و التي تبين مدى تجانس البيانات و قربها أو بعدها عن الوسط الحسابي و المتمثلة في التباين و الانحراف المعياري.

ب- مقاييس ثبات وصدق استمارة الاستقصاء Reliability Analysis و التي تمثلت في كلا من

- معامل ألفا كرونباخ. Alph Coronpach لقياس الصدق في الاجابات .
- معامل الثبات سبيرمان بروان لقياس الثبات النسبي في الاجابات.
- اختبار Cochran Chi-square لقياس الفروق المعنوية في اجابات المستقصي منهم.
- مقاييس النزعة المركزية و التشتت والنسب.
- اختبار " Kruskal Wallis Test " و " Mann-Whitney U " لتحديد مدى وجود فروق معنوية بين عناصر تقييم الخدمات الرعاية الصحية.
- اختبار χ^2 لتقييم العلاقة بين المتغيرات محل الدراسة.
- معامل الارتباط Correlation Coefficient ومعامل ارتباط التوافق contingency coefficient لتحديد مدى الارتباط بين المتغيرات البحثية محل الدراسة والتي تعكس العلاقة بينهم .
- تحليل التباين ANOVA

قياس ثبات وصدق استمارة الاستبيان Reliability Analysis

تم دراسة ٢٥٦ استمارة استبيان من المقيمين بالمنطقة الشرقية بالمملكة العربية السعودية. وتتكون استمارة الاستبيان من ٣٣ سؤال مقسمة إلى عدة مجموعات كل مجموعة تعكس عنصر من عناصر تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية.

جدول (٢)

يوضح معامل الصدق و الثبات ونتائج اختبار كوكران كا^٢ لاستمارة الاستقصاء

المتغير	عدد الأسئلة	معامل الصدق ألفا كرونباخ	معامل الثبات سبيرمان بروان	اختبار كوكران كا ^٢	المعنوية
الأداء الصحي الحكومي المقدم من وزارة الصحة	٦	٠,٨٦	٠,٧٩٤	٣١٤,٣١	٠,٠٠٠
توافر الخدمات الصحية بالمنطقة	٧	٠,٨٢	٠,٦٨	٢٠٢,٨	٠,٠٠٠
الأداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية	١٣	٠,٩٧	٠,٧٨١	٣٢١,٠٨	٠,٠٠٠
الأداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة	٧	٠,٨٥٢	٠,٧١٢	٣١٥,١٥	٠,٠٠٠
إجمالي الاستقصاء	٣٣	٠,٨٧	٠,٦٥٧	١١٣٨,٢	٠,٠٠٠

العينة محل الدراسة وخصائصها

تم تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية من خلال استبيان تم توجيهه لمجموعة من الأفراد المقيمين بالمنطقة الشرقية بالمملكة العربية السعودية
تم حساب حجم العينة بالمعادلة الآتية :

هـ ق ك

= ن

٢_د

ن = حجم العينة

ق = نسبة عدد المفردات التي تتوفر فيها الخصائص محل الدراسة وتساوي (٥٠%)

ك = النسبة المتممة لها وتساوي (١ - ق)

هـ = الإحداثي السيني للمنحنى المعتدل الذي يقطع مساحة من الذيلين (وتساوي ١,٩٦ عند مستوى الثقة ٩٥ %)

د = الخطأ المسموح به (ويساوي في هذه الدراسة ٦ %)

ن = $\frac{2(1,96) \times 0,5 \times 0,5}{(0,06)^2} = 267$ مفردة

٢(٠,٠٦)

ورغباً من الباحث في ألا يقل حجم العينة عن هذا العدد فتم توزيع ٣٠٠ استمارة إستبيان فتم الرد على ٢٥٦ منها فقط أي أن نسبة الرد بلغت ٨٥,٣٣%. وبالتالي تمت الدراسة على ٢٥٦ مفردة وكانت عينة الدراسة موزعة وفقاً للخصائص الديموجرافية للعينة محل الدراسة كما في الجدول التالي :

- وتتمثل الخصائص الديموجرافية لمفردات العينة محل الدراسة في العمر بالسنوات و الجنس والجنسية و الحالة الاجتماعية و قطاع العمل و المستوى التعليمي و مستوى الدخل الشهري وكانت النتائج كما يلي:

جدول (١)

يوضح توزيع مفردات الدراسة على حسب الخصائص الديموجرافية

النسبة %	التكرار	الخصائص الديموجرافية
37.5	96	٢٠ -
20.7	53	٣٠ -
22.7	58	٤٠ -
16.8	43	٥٠ - ٦٠
2.3	6	أكثر من ٦٠
58	149	ذكر
42	107	أنثى
86	220	سعودي
14	36	غير سعودي
44	113	غير متزوج
56	143	متزوج
33	85	حكومي مدني
17	43	حكومي عسكري
38	97	قطاع خاص
12	31	أعمال حرة
15	39	متوسط فأقل
24	62	ثانوية عامة
59	150	جامعي او دبلوم
2	5	دراسات عليا
38	96	أقل من ٥٠٠٠
35	89	٥٠٠٠ - ١٠٠٠٠
18	45	١٠٠٠٠ - ١٥٠٠٠
10	26	أكثر من ١٥٠٠٠

موظفي الجهات الحكومية وموظفي شركات التأمين وموظفي المستشفيات ودراسة أثر الخصائص الشخصية لمفردات الدراسة من عمر ومؤهل علمي على إجابات المستقصي منهم. دراسة رافت أحمد و عبد الله النجار (٢٠٠٨)، تتناول الدراسة المؤشرات التي تعكس الوضع العام للحالة الصحية في المملكة والمؤشرات التي تعكس كفاءة أداء المستشفيات لاستخدامها كأساس لإرساء القواعد العلمية لتطبيق نظام التأمين الصحي بالمملكة. وقد تم استخدام نماذج الانحدار التدرجي Stepwise Regression في تقدير معالم المؤشرات التشغيلية للمستشفى.

المبحث الثاني: تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية

لقد تطورت الخدمات الصحية على مدى سنوات الخطط الخمسية التي نفذتها المملكة العربية السعودية، وتمكنت المملكة عن طريق المخصصات المالية الضخمة التي وفرتها من تطوير مستوى الخدمات الصحية حيث زادت من عدد المستشفيات والأسرة والمستوصفات والعاملين في القطاع الصحي. غير أن هناك بعض المعوقات التي حدت من بلوغ القطاع الصحي لمستوى الطموح. وبات ضرورياً دراسة تلك المشكلات والبحث عن حلول جذرية وواقعية يمكن الاستعانة بها في تطوير القطاع الصحي السعودي والوصول به إلى المستويات القياسية في الدول المتقدمة. وقد رأينا في هذا الفصل أن نلقي الضوء على الوضع الراهن للخدمات الصحية في المملكة وأهم معوقات تطوير القطاع الصحي.

يتم حالياً تقديم خدمات الرعاية الصحية في المملكة عن طريق وزارة الصحة بشكل أساسي، وبعض الجهات الحكومية الأخرى، بالإضافة إلى القطاع الخاص. وتضطلع وزارة الصحة بالنصيب الأكبر حيث تقدم ما يزيد عن ٦٠% من تلك الخدمات، تليها مستشفيات الجهات الحكومية (الجامعات والجهات الأمنية والعسكرية) حيث تقدم ما يعادل ٢٠% تقريباً من هذه الخدمات بينما تأتي مساهمة القطاع الخاص في حدود ١٧% تقريباً. ويرجع ذلك إلى أن ميزانية وزارة الصحة تعد الأعلى بين الجهات المشاركة في تقديم الخدمات الصحية في المملكة حيث بلغت ميزانية قطاع الخدمات الصحية والاجتماعية في ميزانية الدولة لعام ١٤٢٢هـ / ١٤٢٣هـ ما يقارب ٢٣ مليار ريال وبما يعادل ١٤,٦% من إجمالي الميزانية العامة للدولة والبالغة ١٥٧ مليار ريال سعودي، في حين يبلغ حجم استثمارات القطاع الخاص في المجال الصحي السعودي ٦ مليار ريال.

وصلت ميزانية الصحة في المملكة العربية السعودية إلى قيمة تتجاوز ٢٢ مليار دولار. وتزيد المملكة إنفاقها الصحي سنوياً لتغطية متطلبات سكانها، حيث جاءت آخر الزيادات من خلال الدعم المقدم لها في مارس ٢٠١١ مضافاً إلى ميزانية الصحة الحالية ١٦ مليار ريال (٤,٣ مليار دولار). بالإضافة إلى الموارد المالية والدعم المادي، فإن اهتمام المملكة في قطاع الصحة يتمثل بوجود مستشفيات مجهزة بأحدث وأعلى التقنيات في العالم. ووفقاً لمنظمة الصحة العالمية، فقد ساهم التقدم في نشر الأدوية والمستحضرات الصيدلانية في رفع متوسط العمر إلى ٧٣ عاماً في ٢٠٠٨ من ٦٧ عاماً في ١٩٨٨. وحسب توقعات تقرير صدر أخيراً عن البنك الأهلي التجاري سيكون هناك زيادة في ميزانية الصحة في المملكة بنسبة ٥٥% مقارنة بالمستويات الحالية لترتفع إلى ٩٦ مليار ريال بحلول عام ٢٠١٥، على أن ترتفع القدرة الاستيعابية للمستشفيات من ٥٥٩٣٢ إلى ٧٩٢٢٢ سريراً.

يتم تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية من خلال تقييم كلاً من:

- ١- الأداء الصحي الحكومي المقدم من وزارة الصحة
- ٢- توافر الخدمات الصحية بالمنطقة
- ٣- الأداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية
- ٤- الأداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة

الشركات ولم يحصل المستفيدون من العملاء على الخدمة المتعاقد عليها، وخسرت المستشفيات والمستوصفات جزءاً كبيراً من مستحقاتها.

وبدأ من منتصف التسعينات أدركت شركات التأمين الجادة في السوق أهمية الممارسة الكاملة لنشاط التأمين الطبي حيث طرحت الكثير من البرامج التي تديرها شركات التأمين إدارة ذاتية لكافة مراحل العملية التأمينية واستطاعت تلك الشركات بالتنسيق مع عدد كبير من المستشفيات والمراكز الطبية الخاصة من إعداد منظومة للتعامل مع عملاء التأمين باستخدام آلية الدفع المباشر لتلك المستشفيات، وأتاحت تلك الآلية للعميل الحصول على الخدمة الطبية التي يحتاجها من المستشفى أو المركز الطبي المتعاقد مع شركة التأمين التي تتولى بدورها عملية الدفع المباشر لمقدمي الخدمة الطبية.

بدأ تطبيق نظام التأمين الصحي التعاوني في منتصف عام ٢٠٠٥ م على العاملين في القطاع الخاص من غير السعوديين حيث تم التأمين على ما يقارب ٣,٣ مليون عامل في القطاع الخاص، وقد صاحب ذلك تأهيل ٢٤ مستشفى من مستشفيات وزارة الصحة لتقديم خدمات الضمان الصحي وفقاً للأسس التجارية المعمول بها في القطاع الخاص. وسيم في المرحلة اللاحقة تعميم هذا النظام على جميع المواطنين. وقد أعدت وزارة الصحة مشروعاً لتطوير النظام الصحي بالمملكة بمسمى (بلسم) ، ويهدف إلى تحويل جميع مستشفيات وزارة الصحة إلى كيانات مؤسسية تعمل بأسلوب القطاع الخاص وتتمتع بالشخصية الاعتبارية والذمة المالية المستقلة ولا تمول مباشرة من الخزينة العامة للدولة. ويمر تطبيق نظام الضمان الصحي التعاوني الإلزامي بمرحلتين وهما:

الأولى:الاقتصار على المقيمين الأجانب فقط والذين يقدر عددهم بحوالي ٦ ملايين مقيم، ويتوقع أن تبلغ قيمة الاشتراكات عند تطبيق المرحلة الأولى ٥ مليارات ريال .
ثانياً- اتساع النظام ليشمل المواطنين السعوديين أيضاً والذين يقدر عددهم في ذلك الوقت بحوالي ١٧ مليون نسمة، فالاشتراكات المحصلة بقيمة اليوم ستبلغ ١٤ مليار ريال، (٣,٧ مليار دولار)، وهذا يعني أن التطبيق الكامل لنظام الضمان الصحي التعاوني على السعوديين والمقيمين سيرفع حجم اشتراكات المخاطر والتأمين الطبي في المملكة إلى ٢١ مليار ريال، علماً بأن حجم سوق المخاطر والتأمين الطبي في المملكة يقدر عام ٢٠٠١ بحوالي مليار ريال، تبلغ حصة التعاونية للتأمين منها ما نسبته ٣٠ في المائة من السوق.

وتوجد العديد من الدراسات التي اهتمت بدراسة نظام التأمين الصحي بالمملكة العربية السعودية منها:

دراسة صالح العمير (٢٠٠٢)، قام بالتعرف على الوضع الراهن للتأمين الصحي التعاوني بالمملكة العربية السعودية مع تحديد رؤية مستقبلية للتأمين الصحي التعاوني وأثره على الاقتصاد السعودي حتى عام ٢٠٢٠ م من خلال التعرف على تطور اشتراكات التأمين الصحي التعاوني والاشتراكات التقديرية للتأمين الصحي والاعتماد في ذلك على نسبة اشتراكات التأمين الصحي إلى الناتج المحلي الاجمالي. وقد خلص إلى أن التأمين الصحي التعاوني سيزيد من الناتج المحلي الإجمالي بنسبة تتراوح بين ٥,٧% إلى ٦,٩% عام ٢٠٢٠ م .

دراسة أحمد حسين (٢٠٠٥)، تناول دوافع التعامل مع شركات التأمين الصحي التعاوني وأثرها على درجة الرضا لدى طالبي الخدمة الصحية وذلك من خلال دراسة تطبيقية على محافظة جدة وكان ومن أهم نتائجها: تأثر درجة رضا المستفيد من خدمة التأمين الصحي التعاوني بجوانب خدمة التأمين الصحي والمتمثلة في الخدمة والاعتمادية والجوانب المالية والأمان. ووجود ارتباط عكسي بين درجة احتياج الخدمات التأمينية الصحية للتطوير والتحسين وفقاً لأراء المسؤولين في شركات التأمين الصحي ودرجة رضا المسؤولين في الشركات المتعاقدة عن الخدمة التأمينية الصحية المتعاقدة عليها.

دراسة تغريد خليل (٢٠٠٦)، بينت أن هناك تحديات فنية وثقافية وإدارية ورقابية تواجه تطبيق نظام التأمين الصحي التعاوني بالمملكة والبحث عن حلول لها من خلال وجهة نظر

المبحث الأول: التأمين الصحي بالمملكة العربية السعودية

عرفت المملكة عدة أشكال للتأمين الصحي أهمها التأمينات الاجتماعية التي توفر الرعاية لحالات إصابات العمل للعاملين المشمولين بهذا التأمين الذي تتم ممارسته عن طريق المؤسسة العامة للتأمينات الاجتماعية، وهو يعد أقدم أنواع التأمين الصحي التي عرفتها المملكة العربية السعودية. والنوع الثاني من التأمين الصحي في المملكة هو التأمين الذي يتم من خلال التعاقد المباشر بين الشركات والمستشفيات لتوفير الرعاية الصحية لمنسوبيها أو قيام بعض المستشفيات بتوفير برامج للرعاية الصحية التي يحتاجها الفرد مقابل قسط يدفع للمستشفى مباشرة ولا يفضل ممارسو التأمين الصحي هذا النوع من العلاقة باعتبار أن المستشفى تكون في هذه الحالة خارج أي نظام رقابي. أما النوع الثالث الأكثر انتشاراً والأحدث في الوقت نفسه هو التأمين الصحي الخاص والذي يشمل التأمين التجاري والتعاوني الذي تتم ممارسته عن طريق شركات التأمين كطرف ثالث.

وقد بدأت ممارسة التأمين الصحي الخاص في المملكة العربية السعودية منذ فترة وجيزة لا تتعدى عشرون عاماً فقط فيما عدا حالات محدودة كانت موجودة قبل ذلك التاريخ، وذلك لأن القطاع الصحي الحكومي تولى عن الأفراد ولمدة طويلة عبء تكاليف الرعاية الصحية، لكن تغير هذا التوجه بعد أن أصبح هذا العبء يشكل عامل ضغط قوي على خلفية الزيادة الطبيعية في السكان من جهة والارتفاع الكبير في أعداد العمالة الوافدة من جهة ثانية فضلاً عن ارتفاع تكلفة العلاج الطبي، وما تؤدي إليه مجانية الرعاية الصحية من سوء استخدام الموارد الصحية. من هنا بدأ الاهتمام بالتأمين الصحي يتزايد أولاً لدى مؤسسات القطاع الخاص تحت تأثير عاملين اثنين هما رغبة هذه الشركات في تحويل عبء إدارة التعويضات الطبية إلى شركات تأمين متخصصة ومصصلحة هذه الشركات أن تتنبأ سلفاً بحجم التعويضات وتدرجها في حساباتها. وقد شجع على نمو هذا التوجه، أن شركات التأمين العاملة في المملكة أخذت تهتم بتقديم خدمات التأمين الصحي لعملائها من الشركات إدراكاً منها أن هذا النشاط يشكل رافداً قوياً لأعمالها ومصدراً هاماً من الأقساط التأمينية إضافة إلى رغبتها في تقديم خدمات تلبي احتياجاً ملحاً لدى عملائها الكبار وبجودة مقبولة لديهم.

في تلك الأثناء كان سوق التأمين الصحي يعاني من وجود عدد هائل من ممارسي خدمة التأمين الطبي حتى دون أن يكون لديهم ترخيص بالمزاولة أو خبرة فنية كافية للاكتتاب وكذلك لم يكن لديهم القدرة المالية أو الأنظمة المعلوماتية لإدارة محفظة التأمين الصحي. أكثر من ذلك، فقد مارس عدد من الأفراد خدمة التأمين الصحي دون أن يكون لديهم شركات معتمدة. كل ذلك على خلفية ضعف الوعي داخل المملكة بالتأمين عموماً وبالتأمين الصحي على وجه الخصوص. وباعتبار أن المستشفيات طرف أساسي في خدمة التأمين الصحي فقد كانت تعاني هي أيضاً من خلل في منظومة إدارة تعاقدات التأمين الطبي التي كانت ترد إليها من شركات التأمين. ومع دخول بعض شركات التأمين الكبرى كالتعاونية للتأمين لسوق التأمين الطبي السعودي اعتباراً من عام ١٩٩٠ تطور أسلوب ممارسة هذا التأمين، حيث تركز دور شركات التأمين على الاكتتاب في التأمين الطبي، بينما تم التعاقد مع طرف ثالث متخصص يملك أنظمة لإدارة التعويضات ومتعاقد مع شبكة من المستشفيات والمستوصفات لتقديم الخدمة الطبية نفسها.

لكن تلك المرحلة تمخضت عن عدد من السلبيات التي ظهرت نتيجة تطبيق نظام الإدارة الخارجية فيما يتعلق بسعر التأمين أو إدارة التعويضات بشكل نتج عنه زيادة حجم التعويضات بما يفوق كثيراً الاشتراكات المكتتبه في التأمين الطبي، كما نتجت عنه عدد من المشاكل سواء مع العملاء أو المستشفيات. واضطرت بعض شركات التأمين تحت وطأة المنافسة غير المتكافئة إلى طرح التأمين الطبي بأسعار منخفضة لا تتناسب مع طبيعة الأخطار المغطاة بغرض الحصول على النصيب الأكبر من السوق، لكن عندما تزايد معدل التعويضات اختفت تلك

ثانياً- المستشفيات الحكومية / الخاصة

من خلال الدراسة يتم الكشف على نظم تكاليف الرعاية الصحية المقدمة والعمل بقدر الإمكان على صياغتها في صور دوال رياضية بما يساعدنا من إمكانية التنبؤ بسلوكها المستقبلي. وكذلك بناء نموذج إحصائي يعكس المتغيرات المختلفة المؤثرة في تكاليف خدمات الرعاية الصحية مما يضع مخططاً للرعاية الصحية في موقف يمكنهم من ترشيد هذه التكاليف. وهذا بطبيعة الحال سيتيح للمسؤولين بعض المؤشرات التي تعمل على الرفع من ناتج مخرجات المستشفيات الحكومية والخاصة كما تساعد على تحسين معدلات الاستثمار والعائد للمستشفيات الخاصة بدون تحميل المرضى المزيد من الأعباء أو تتيح لها إمكانية تقديم خدمات الرعاية الصحية بأسعار أقل مما يخفف الأعباء المالية من على كاهل أفراد المجتمع بالمملكة.

ثالثاً- شركات التأمين الصحي والمؤسسات

يساهم البحث من خلال دراسته لواقع خدمات الرعاية الصحية المقدمة ومؤشرات نظم التكاليف والتي يمكن على أساسها وضع نظام للتأمين الصحي في المملكة العربية السعودية مع الأخذ في الاعتبار البعد الاجتماعي وكيفية تغطية فئات المجتمع منخفضة الدخل. وفي ضوء ذلك يمكن تقدير مستوى الأقساط لكل نوع من أنواع وثائق التأمين الطبي المقترحة والمزايا الممنوحة بموجبها بما يسهل من عمل شركات التأمين الطبي واعتمادها على نتائج دراسة علمية وعملية ويتم فيها مراعاة الأمور الاكتوارية الخاصة ببناء نظام التأمين الصحي. وكذلك الأمر بالنسبة لكافة مؤسسات الدولة التي سبستفيد منسوبها من هذا النظام من معرفة التكاليف التي ستتحملها.

منهج الدراسة:

لتحقيق هدف الدراسة يتم أولاً تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية من خلال تقييم كلاً من الأداء الصحي الحكومي المقدم من وزارة الصحة وتوافر الخدمات الصحية بالمنطقة و الأداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية والأداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة وذلك بالاعتماد استبيان تم توجيهه لمجموعة من الأفراد المقيمين بالمنطقة الشرقية بالمملكة العربية السعودية. دراسة القيم الاتجاهية مؤشرات نظم تكاليف الخدمات الصحية المقدمة من خلال دراسة العلاقة الاتجاهية لأهم مؤشرات تكاليف الخدمات الصحية مع الزمن وإيجاد دوال التكاليف المختلفة والتي تمكنا من التوقع لهذه التكاليف وفقاً لنموذج الإحصائي الذي يعبر عنها.

تبويب البحث

لتحقيق هدف البحث السابق الإشارة إليه، يتم تقسيم هذا البحث إلى مايلي:

المبحث الأول: التأمين الصحي بالمملكة العربية السعودية

المبحث الثاني: تقييم خدمات الرعاية الصحية المقدمة في المنطقة الشرقية

المبحث الثالث: تقدير القيم الاتجاهية لمؤشرات نظم تكاليف الخدمات الصحية المقدمة

خاتمة: النتائج والتوصيات.

للعلاج. المساهمة في دعم الاقتصاد الوطني من خلال زيادة فرص الاستثمار في القطاع الصحي وفروعه المختلفة وتشجيع المستثمرين على توظيف أموالهم في المشاريع الصحية.

لذلك تعتمد منهجية البحث على تقييم الوضع الحالي للرعاية الصحية المقدمة وذلك بدراسة أهم مؤشرات ملامح الوضع الصحي بالملكة ، وكذلك دراسة الموارد الصحية المتاحة من المرافق والقوى العاملة بوزارة الصحة والمراكز الصحية والمستشفيات والأسرة في وزارة الصحة وفي الجهات الحكومية. ودراسة الأنشطة والخدمات الصحية زيارات المراجعين والتنويم والعمليات الجراحية والولادات والمواليد وغيرها وذلك من أجل تقييم الرعاية الصحية الحكومية المقدمة من خلال دراسة تطورها ومقارنتها مع المؤشرات العالمية لأدائها وبالتالي تحديد مدى الاستفادة من الإمكانيات المتاحة والوقوف على حسن استخدام الموارد الصحية المتاحة الحكم على جودة الخدمات الصحية المقدمة. واستخدام الدوال الإحصائية المختلفة في وصف تطور تكاليف الخدمات الصحية المختلفة وكذلك استخدامها في تقدير معالم هذه الدوال ومن ثم تقدير تكاليف خدمات الرعاية الصحية المختلفة وذلك لعينة من مناطق المملكة المختلفة تعكس كافة الخصائص المتاحة لمجتمع الخدمات الصحية بالملكة.

أهمية الدراسة:

يتضح لنا مما سبق أنه مع تزايد الفجوة بين الموارد الصحية وتكاليف تقديم الخدمات الصحية تظهر الحاجة إلى وجود نظام للتأمين الصحي يعمل على تقديم الرعاية الصحية للمواطنين والمقيمين بالملكة بما يساهم في رفع المستوى الصحي للمجتمع السعودي مما ينعكس على تزايد معدلات التنمية الاقتصادية للمجتمع . ترجع أهمية هذه الدراسة إلى أن تقييم الرعاية الصحية الحكومية المقدمة قد يسفر عن بعض النتائج الهامة من حيث الإمكانيات وجودة الخدمة الصحية المقدمة والتي تساعد واضعي الخطط الاستراتيجية الصحية بالملكة والمساهمة في تطوير أداء هذه الخدمات. وكذلك دراسة نظم التكاليف المختلفة للرعاية الصحية المطبقة حالياً يؤدي إلى حسن الحكم على كفاءة هذه النظم واستخدامها في التنبؤ بما تكون عليه التكاليف المستقبلية. وبنعكس هذا في بناء نظام متكامل للتأمين الصحي يعمل على تحديد المستفيدين منه والمزايا الممنوحة لهم وتكاليف هذه الخدمات بما يسمح ويساعد في تقييم الأعباء التي تتحملها الدولة في سبيل التغطية الصحية للأفراد منخفضي الدخل.

تهدف الدراسة إلى:

- ١ - تقييم خدمات الرعاية الصحية المقدمة بالمنطقة الشرقية بالملكة العربية السعودية.
- ٢ - دراسة العوامل المختلفة المؤثرة في جودة خدمات الرعاية الصحية المقدمة بالمنطقة الشرقية .
- ٣ - دراسة القيم الاتجاهية لمؤشرات نظم تكاليف الخدمات الصحية المقدمة بما يساهم في التنبؤ بها مستقبلياً وبناء الخطط والاستراتيجيات المختلفة لمواجهتها.

الدراسة تفيد العديد من الجهات وهي :

أولاً- وزارة الصحة:

يمكن لها من خلال الدراسة الوقوف على مستوى خدمات الرعاية الصحية المقدمة لأفراد المجتمع السعودي ومدى كفايتها والأساليب التي تقدم بها بما يمكنها من معرفة أسباب الخلل أن وجد والعمل على إيجاد الحلول المناسبة لها. وكذلك معرفة هل الإمكانيات المتاحة من موارد صحية كافية أم غير كافية وكذلك ما مدى الاستفادة منها وتحقيق أقصى استخدام لها وعدم أهدرها مما يساعد في وضع خططها الاستراتيجية المستقبلية وتوجيه هذه الإمكانيات نحو الاستخدامات المطلوبة لرفع من كفاءة وجودة الخدمات الصحية المقدمة.

خدمات الرعاية الصحية وتقدير القيم الاتجاهية لمؤشرات تكاليفها باستخدام الأساليب الكمية

كمدخل للتأمين الصحي : دراسة تطبيقية على المنطقة الشرقية بالمملكة العربية السعودية

أسامة حنفي محمود حسن^١

المخلص

أصبح للتأمين الصحي أهمية متزايدة مع الارتفاع الكبير والمستمر في تكاليف الرعاية الصحية لما يحققه من أهداف اقتصادية واجتماعية. و للمساعدة في بناء نظام للتأمين الصحي بالمملكة العربية السعودية فلا بد أولاً تقييم خدمات الرعاية الصحية المقدمة من خلال تقييم كلاً من الأداء الصحي الحكومي المقدم من وزارة الصحة و توافر الخدمات الصحية بالمنطقة و الأداء الصحي للمستشفيات الحكومية ومراكز الرعاية الأولية والأداء الصحي لمستشفيات القطاع الخاص والمراكز والعيادات الخاصة وذلك بالاعتماد استبيان تم توجيهه لمجموعة من الأفراد المقيمين بالمنطقة الشرقية بالمملكة العربية السعودية. كما تم دراسة مؤشرات نظم تكاليف الخدمات الصحية المقدمة بالمنطقة الشرقية من خلال دراسة العلاقة الاتجاهية لأهم مؤشرات تكاليف الخدمات الصحية مع الزمن و إيجاد دوال التكاليف المختلفة والتي تمكنا من التوقع لهذه التكاليف وفقاً لنموذج الإحصائي الذي يعبر عنها.

مقدمة:

يشكل اطمئنان الإنسان على صحته وصحة أسرته أهم حافز له من أجل بذل المزيد من العطاء في العمل الذي يوكل إليه، كما أن قلقه على صحته وصحة أبنائه من شأنه أن يشكل أول عوامل الإعاقة عن تحقيق ما يتطلبه عمله من اهتمام وما يحتاج إليه من عناية. ويعتبر التأمين الصحي أحد أهم أنواع التأمين وأكثره ارتباطاً وحساسيةً بأفراد المجتمع ومنظّماته وذلك نتيجة لما له من أثر إيجابي في توفير الأمان و الادخار و حماية الأسرة و الوقاية من الأمراض ومساهمته في التنمية الاقتصادية .

كما يتم تعريف التأمين الصحي كما عرفته منظمة الصحة العالمية "بأنه وسيلة لدفع بعض أو كل تكلفة الرعاية الصحية ، فهو يحمي المؤمن عليهم من دفع التكلفة العالية للعلاج في أحوال المرض. وأساس التأمين الصحي هو قيام المستهلك بدفع اشتراك منتظم لمؤسسة إدارية تُعد هي المسؤولة عن إدارة تلك المدفوعات في نظام لدفع نفقات العلاج إلى مقدمي الخدمة الصحية".

كما تتبع أهمية التأمين الصحي من عدة عوامل اجتماعية واقتصادية منها:

الارتفاع الكبير والمستمر في تكاليف الرعاية الطبية الحديثة أصبح فوق القدرة المالية لغالبية الأفراد والأسر والتي يصل تأثيرها في بعض الحالات المرضية الى لجوء الفرد لبيع كل ما يملك والتدين في سبيل العلاج. عدم امكانية التنبؤ بحدوث المرض او كلفته على مستوى الفرد والأسرة مما يجعل من الصعب اتخاذ الاجراءات الضرورية نحو موازنة الدخل لتغطية هذه التكاليف. مساهمة التأمين الصحي في تحقيق الاستقرار النفسي والاجتماعي والاقتصادي لفئات العمال والموظفين والتخلص من مظاهر القلق الممكن حدوثه نتيجة لعدم توافر الاساليب المناسبة

^١ الأستاذ المساعد بقسم الرياضيات والإحصاء والتأمين - أكاديمية السادات للعلوم الإدارية

المراجع

- 1- إبراهيم،منى دراسة تحليلية لإقتصاديات نظم التأمينات الإجتماعية والمعاشات رسالة ماجستير كلية التجارة و ادارة الاعمال - جامعة حلوان 2007
- 2- الصياد،محمد ، دور التأمينات الإجتماعية فى شبكة الأمان الإجتماعي ، المؤتمر القومى للتنمية الإجتماعية ، وزارة التأمينات والشئون الإجتماعية ، القاهرة ، 17-19 ديسمبر 2002 .
- 3- الصياد،محمد و الوزيرى، ليلي حالات إستحقاق المعاش فى قانون التأمين الإجتماعي الصادر بالقانون رقم 79 لسنة 1975 ، معهد التأمينات ، المؤسسة الثقافية العالية ، الإتحاد العام لنقابات عمال مصر.
- 4- خيرى، أمنية الاستدامة المالية لنظام التأمينات الاجتماعية فى مصررسالة ماجستير كلية التجارة جامعة عين شمس 2008.
- 5- محمد ، علي واقع التأمين الإجتماعي فى مصر ، مؤتمر التأمينات الإجتماعية بين الواقع والمأمول ، جامعة الأزهر ، 13-15 أكتوبر ، 2002 ،
- 6- محمود، السيد ، نشأة وتطور التأمينات الإجتماعية ، مؤتمر التأمينات الإجتماعية بين الواقع والمأمول ، جامعة الأزهر ، 13-15 أكتوبر ، 2002.

النتائج والتوصيات

النتائج

توضح بعض المؤشرات الإحصائية زيادة متوسط عدد الشكاوى الواردة الهيئة القومية للتأمينات الاجتماعية بقطاعيها (العام والخاص) و (الحكومي) ويرجع ذلك الى :-

- 1- عدم نشر الوعي التأميني
- 2- عدم إصدار أو متابعة تنفيذ التعليمات والقرارات التي من شأنها الحد من شكاوى أصحاب المعاشات وضم المدد
- 3- عدم التواصل بين السجل المدنى والهيئة القومية للتأمينات الاجتماعية
- ويرى الطالب أنه قد يكون لهذه التوصيات بعض النتائج الإيجابية وتتلخص فيما يلي :

التوصيات

- 1- توحيد الرقم التأميني للمواطن حيث يستلزم التسجيل أول مرة تسجيل الرقم القومي حتى نضمن عدم تكرار الرقم التأميني لنفس الشخص للحد من شكاوى الرقم التأميني والتغطية التأمينية حيث اتضح من مصفوفة معاملات الارتباط ان الارتباط بينهما طردي قوي
- 2- توحيد رقم المنشأة لنفس العين حتى نتلافى ظاهرة المنشآت والمديونيات الرهمية والحد من شكاوى التغطية التأمينية ومعاشات اصحاب الأعمال حيث اتضح من مصفوفة معاملات الارتباط ان الارتباط بينهما طردي قوي
- 3- تعديل القوانين المختلفة للتأمينات الاجتماعية لتشمل جميعها التأمين المرضي و إصابات العمل مع إختلاف نسبة الإشتراك لتشمل عدد من الشرائح وإختلاف معدلات العائد منها للحد من ظاهرة التهرب التأميني و الشكاوى المترتبة عليها
- 4- إنشاء إدارة مستقلة لتقييم الأداء تتبع لرئيس الهيئة مباشرة يكون الهدف منها تجميع البيانات الخاصة بالشكاوى وتحليلها إحصائيا ووضع نموذج للتنبؤ بعدد الشكاوى الواردة في فترات زمنية مستقبلية ليساعد ذلك في تقييم أداء العاملين
- 5- إمداد المختصين بنتائج التحليل الإحصائي عن عدد الشكاوى وإرتباط بعض الأقسام ببعضها البعض (التغطية التأمينية و أصحاب المعاشات - التغطية التأمينية والرقم التأميني) لإصدار قرارات وتعليمات تساعد على الحد من هذه الشكاوى ومقارنة الشكاوى الواردة بالشكاوى المتوقعة طبقا للنموذج المقترح.
- 6- نشر الوعي التأميني عن طريق الإعلام والمؤتمرات الدورية والندوات في أماكن العمل
- 7- تفعيل الدور الرقابي والمتابعة لتنفيذ التعليمات التي من شأنها الحد من عدد الشكاوى.

بمقارنة عدة نماذج تم اختيار هذا النموذج حيث اتضح

Box-Pierce (Ljung-Box)

P-Value 0.294 > 0.05

SARIMA (0,1,1) (1,1,0)₁₂

$$(1-\beta) (1-\beta^{12}) (1-\beta^{12}) Y_t = C + (1-\theta\beta) e_t$$

$$(1-\beta) (1+1.0021\beta^{12}) (1-\beta^{12}) Y_t = 31.876 + (1-0.9223 \beta) e_t$$

$$Y_t + 0.0021\beta^{12} Y_t - 1.0021\beta^{24} Y_t - \beta Y_t - 0.0021\beta^{13} Y_t + 1.0021\beta^{25} Y_t =$$

$$32.876 - 0.9223\beta$$

MODEL

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	-0.1039	0.2550	-0.41	0.688
SAR	12	-1.0021	0.0440	-22.76	0.000
MA	1	0.9223	0.1928	4.78	0.000
Constant		31.876	7.906	4.03	0.001

Differencing: 1 regular, 1 seasonal of order 12
 Number of observations: Original series 36, after differencing 23

Residuals: SS = 3491570 (backforecasts excluded)

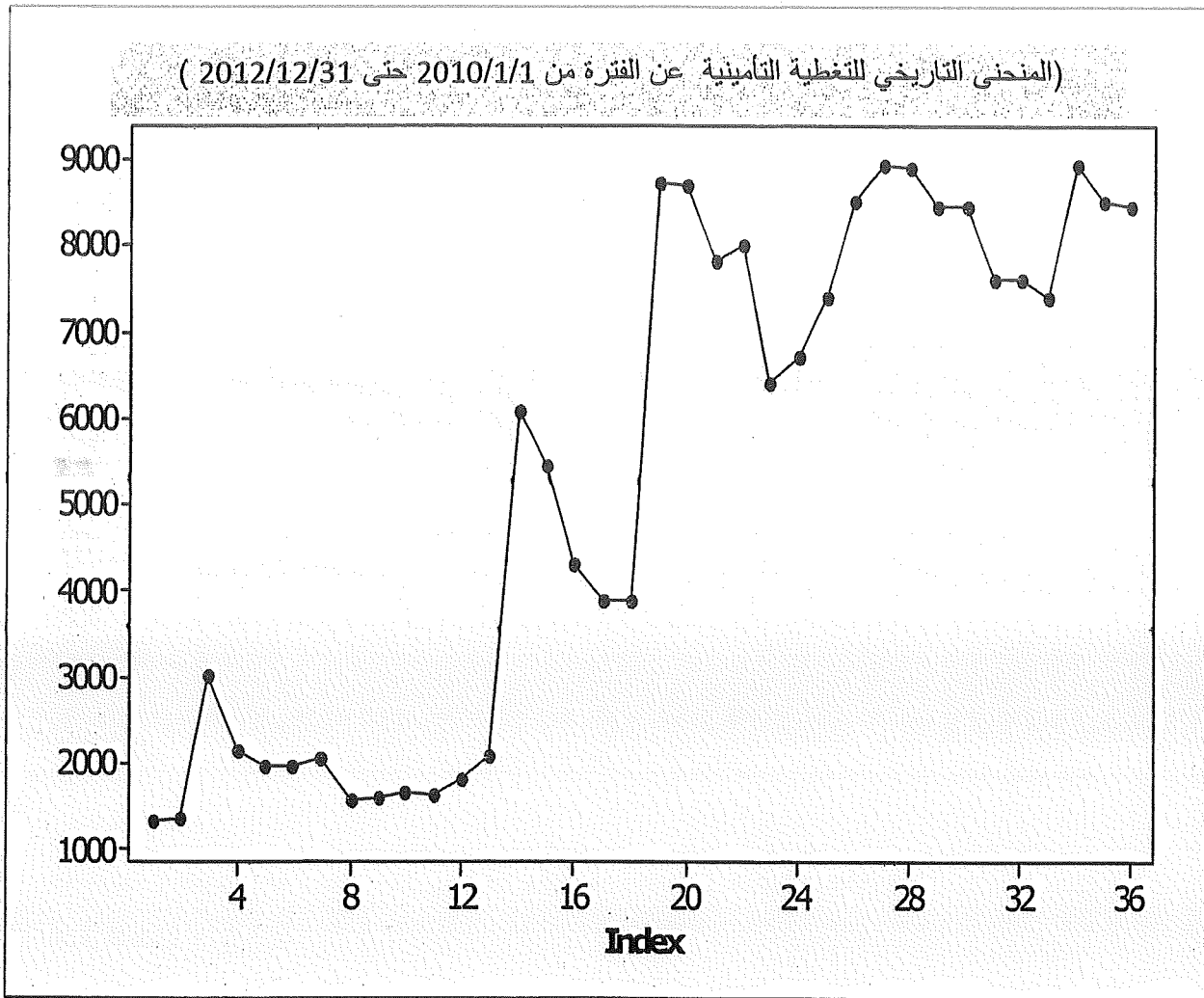
MS = 183767 DF = 19

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	9.6	*	*	*
DF	8	*	*	*
P-Value	0.294	*	*	*

Forecasts from period 35

Period	Forecast	95 Percent Limits		Actual
		Lower	Upper	
36	8416.21	7575.83	9256.60	8465.00
37	8748.26	7907.59	9588.93	



شكل رقم (1)

يستخدم هذا الاختبار للمقارنة بين متوسطي الرقم التأميني بالقطاع العام والخاص والقطاع الحكومي و نستخدم هنا احصائية T للاختبار و المطلوب هنا اختبار وجود فرق معنوي بين المتوسطات عند مستوى دلالة 5%

اختبار التجانس

$$H_0 : \sigma_1 = \sigma_2 , H_1 : \sigma_1 \neq \sigma_2$$

وحيث ان $P.VALUE 0.00 < 0.05$ اذن نرفض الفرض العدمي و يوجد فروق معنوية و نقبل الفرض البديل القائل ان المجتمعين غير متجانسين لذا نهتم بالصف الثاني

نلاحظ ان الاختبار يجري في حالتين
الفروض

$$H_0 : \mu_1 = \mu_2 , H_1 : \mu_1 \neq \mu_2$$

الأولى : نفترض تساوي تباين المجتمعين المأخوذ منها العينة $\mu_1 = \mu_2$ متوسط عدد شكاوى الرقم التأميني بالقطاع الخاص = متوسط عدد شكاوى الرقم التأميني بالقطاع الحكومي

الثانية : نفترض عدم تساوي التباينات $\mu_1 \neq \mu_2$ متوسط عدد شكاوى الرقم التأميني بالقطاع الخاص $\neq$ متوسط عدد شكاوى الرقم التأميني بالقطاع الحكومي

و نلاحظ انه في الحالة الأولى $P.Value 0.00 < 0.025$ اذن نرفض الفرض العدمي القائل ان المتوسطات متساوية اي انه توجد فروق معنوية بين المتوسطين اي نقبل الفرض البديل القائل ان متوسطي المجتمعين غير متساويين

يستخدم هذا الاختبار للمقارنة بين متوسطي شكاوى واستفسارات اصحاب المعاشات بالقطاع العام والخاص والقطاع الحكومي و نستخدم هنا احصائية T للاختبار و المطلوب هنا اختبار وجود فرق معنوي بين المتوسطات عند مستوى دلالة 5%

اختبار التجانس

$$H_0 : \sigma_1 = \sigma_2 , H_1 : \sigma_1 \neq \sigma_2$$

وحيث ان $P.VALUE 0.00 < 0.05$ اذن نرفض الفرض العدمي و يوجد فروق معنوية و نقبل الفرض البديل القائل ان المجتمعين غير متجانسين لذا نهتم بالصف الثاني

نلاحظ ان الاختبار يجري في حالتين
الفروض

$$H_0 : \mu_1 = \mu_2 , H_1 : \mu_1 \neq \mu_2$$

الأولى : نفترض تساوي تباين المجتمعين المأخوذ منها العينة $\mu_1 = \mu_2$ متوسط شكاوى واستفسارات اصحاب المعاشات بالقطاع الخاص = متوسط عدد شكاوى واستفسارات اصحاب المعاشات بالقطاع الحكومي

الثانية : نفترض عدم تساوي التباينات $\mu_1 \neq \mu_2$ متوسط عدد شكاوى واستفسارات اصحاب المعاشات بالقطاع الخاص $\neq$ متوسط عدد شكاوى واستفسارات اصحاب المعاشات بالقطاع الحكومي

و نلاحظ انه في الحالة الأولى $P.Value 0.00 < 0.025$ اذن نرفض الفرض العدمي القائل ان المتوسطات متساوية اي انه توجد فروق معنوية بين المتوسطين اي نقبل الفرض البديل القائل ان متوسطي المجتمعين غير متساويين

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
التغطية التأمينية	Equal variances assumed	154.998	.000	8.262	70	.000	4195.000
	Equal variances not assumed			8.262	38.064	.000	4195.000
الرقم التأميني	Equal variances assumed	49.503	.000	9.142	70	.000	12198.44
	Equal variances not assumed			9.142	46.686	.000	12198.44
اصحاب المعاشات	Equal variances assumed	94.988	.000	11.297	70	.000	6186.361
	Equal variances not assumed			11.297	36.806	.000	6186.361

جدول رقم (5)

يستخدم هذا الاختبار للمقارنة بين متوسطي التغطية التأمينية بالقطاع العام والخاص والقطاع الحكومي و نستخدم هنا احصائية T للاختبار و المطلوب هنا اختبار وجود فرق معنوي بين المتوسطات عند مستوى دلالة 5%

اختبار التجانس

$$H_0 : \sigma_1 = \sigma_2 , H_1 : \sigma_1 \neq \sigma_2$$

وحيث ان $P.V.A.L.U.E < 0.05$ $0.00 < P.V.A.L.U.E$ اذن نرفض الفرض العدمي و يوجد فروق معنوية و نقبل الفرض البديل القائل ان المجتمعين غير متجانسين لذا نهتم بالصف الثاني

نلاحظ ان الاختبار يجري في حالتين

الفروض

$$H_0 : \mu_1 = \mu_2 , H_1 : \mu_1 \neq \mu_2$$

الأولى : نفترض تساوي تباين المجتمعين المأخوذ منها العينة $\mu_1 = \mu_2$ متوسط عدد شكاوى التغطية التأمينية بالقطاع الخاص = متوسط عدد شكاوى التغطية التأمينية بالقطاع الحكومي

الثانية : نفترض عدم تساوي التباينات $\mu_1 \neq \mu_2$ متوسط عدد شكاوى التغطية التأمينية بالقطاع الخاص $\neq$ متوسط عدد شكاوى التغطية التأمينية بالقطاع الحكومي

و نلاحظ انه في الحالة الأولى $P.Value < 0.025$ $0.00 < P.Value$ اذن نرفض الفرض العدمي القائل ان المتوسطات متساوية اي انه توجد فروق معنوية بين المتوسطين اي نقبل الفرض البديل القائل ان متوسطي المجتمعين غير متساويين

● بالنسبة للارتباط بين الرقم التأميني بالقطاع الخاص الرقم التأميني بالقطاع الحكومي

- هو ارتباط طردي قوي و ان المتغيرين لهم نفس السلوك و الاتجاه مما يدل على انه لا تقوم الجهات الادارية بالحكومة او المواطنين بتسجيل الارقام التأمينية لكل شخص و هذا بشكل عام يوضح عدم التنسيق بين الهيئة القومية للتأمينات الاجتماعية و السجل المدني من جهة اخرى حيث انه من الممكن ان يتم الربط بينهم بإمداد السجل المدني للتأمينات الاجتماعية بأسماء المواليد الجدد بصفة دورية مما يقلل او يمنع عدم تسجيل رقم تأميني لكل مواطن او تسجيل اكثر من رقم للمواطن مما يؤثر بالسلب في حصوله على مستحقاته

● بالنسبة للارتباط بين اصحاب المعاشات بالقطاع الخاص و اصحاب المعاشات بالقطاع الحكومي

- هو ارتباط طردي قوي و ان المتغيرين لهم نفس السلوك و الاتجاه و هذا يوضح ان هناك نفس الاجراءات المعقدة او التأخير في صرف المعاش حيث ان القائمين على الصرف مرتبطين بقوانين و تعليمات بها جزء من الروتين و البيروقراطية و هذا يتطلب اصدار تعليمات او تحديث تعليمات ضم المدد بالقطاع الخاص و تحديث البيانات اولا بأول للقطاع الحكومي قبل بلوغ سن المعاش بفترة كافية حتى يتسنى للعاملين عند بلوغهم سن استحقاق المعاش انهاء كافة اجراءات الصرف في حدود شهر على الأكثر

● بالنسبة للارتباط بين التغطية التأمينية بالقطاع الحكومي و الرقم التأميني بالقطاع الحكومي

- هو ارتباط طردي ضعيف حيث ان هناك كثير ممن لهم ارقام تأمينية مسجل لهم بيانات تأمينية بالقطاع الحكومي و لكنها غير محدثة (لم يتم تسجيل الوظيفة و الاجر الحاليين)

• بالنسبة للارتباط بين التغطية التأمينية بالقطاع الخاص و الرقم التأميني بالقطاع الخاص

- هو ارتباط طردي قوي و ان المتغيرين لهم نفس السلوك و الاتجاه و هذا دليل ان المواطنين الغير مسجل لهم ارقام تأمينية غير مؤمن عليهم او غير مسجل لهم بيانات تغطية تفيد اماكن و تواريخ و اجر التحاقهم بالمنشآت الخاصة و هذا يؤدي الى ضياع حقوق العاملين .

• بالنسبة للارتباط بين التغطية التأمينية بالقطاع الخاص واصحاب المعاشات بالقطاع الخاص

- هو ارتباط طردي قوي و ان المتغيرين لهم نفس السلوك و الاتجاه حيث انه ليس كل من له تغطية تأمينية يستحق معاش عند بلوغه سن استحقاق المعاش و ايضا هناك كثير ممن لهم مدد تأمينية متعددة يتضررون من عدم الصرف بسبب عدم ضم هذه المدد لآخر مكتب تأمينات مؤمن على العامل به و هذا يدل على ان هناك قصور في تعليمات او في تطبيق تعليمات ضم المدد

• بالنسبة للارتباط بين التغطية التأمينية بالقطاع الخاص و التغطية التأمينية بالقطاع الحكومي

- هو ارتباط طردي متوسط و هذا دليل ان القطاع الحكومي رغم التزامه بالتأمين على جميع العاملين لديه الا انه يسلك نفس سلوك القطاع الخاص في عدم تسجيل بيانات العاملين طرفهم و لكن مع الفرق ان القطاع الحكومي يقوم في كثير من الاحيان بتسجيل بيانات العاملين و تحديث بياناتهم عند خروجهم على المعاش بموجب قرار استلام و نهاية الخدمة اما في القطاع الخاص فيعد عدم تسجيل بيانات العاملين بموجب استمارة بداية الخدمة خلال اسبوعين من الالتحاق بالعمل يعتبر بمثابة تهرب تأميني مما يؤثر على ضياع حقوق العاملين

• بالنسبة للارتباط بين الرقم التأميني بالقطاع الخاص واصحاب المعاشات بالقطاع الخاص

- هو ارتباط طردي قوي و ان المتغيرين لهم نفس السلوك و الاتجاه حيث انه كثير ممن بلغوا سن استحقاق المعاش ليس لديهم ارقام تأمينية و بالتالي ليس لديهم تغطية تأمينية (مدد تأمينية مسجلة على الحاسب الخاص بالهيئة) و من ثم لا يستحقون معاش و لكن يتم ان ذاك تسجيل و طباعة رقم تأميني يؤكد عدم وجود مدد تأمينية تعطيه الحق في المعاش لحصولهم على اعانة الضمان الاجتماعي

مصفوفة معاملات الارتباط

	تغطية تأمينية قطاع خاص	رقم تأميني قطاع خاص	اصحاب معاشات قطاع خاص	تغطية تأمينية قطاع حكومي	رقم تأميني قطاع حكومي	اصحاب معاشات قطاع حكومي
تغطية تأمينية قطاع خاص	1	.856**	.821**	.486**	.867**	.895**
	36	.000	.000	.003	.000	.000
		36	36	36	36	36
رقم تأميني قطاع خاص		1	.991**	.481**	.966**	.955**
		36	.000	.003	.000	.000
			36	36	36	36
اصحاب معاشات قطاع خاص			1	.437**	.967**	.954**
			36	.008	.000	.000
				36	36	36
تغطية تأمينية قطاع حكومي				1	.319	.421*
				36	.058	.010
					36	36
رقم تأميني قطاع حكومي					1	.974**
					36	.000
						36
اصحاب معاشات قطاع حكومي						1
						36
						36

جدول رقم (4)

معامل بيرسون للارتباط عند $\alpha = 5\%$

قيمة معامل بيرسون r توضح اتجاه العلاقة و قوتها

يتضح من الجدول (3)

المقارنة بين القطاع العام و الخاص و القطاع الحكومي

أولا / القطاع العام و الخاص

• التغطية التأمينية

يبلغ متوسط شكاوى و استفسارات التغطية التأمينية تقريبا 5371 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 8932 و ادنى مستوى وصل الى 1334 شكوى و استفسار

• الرقم التأميني

يبلغ متوسط شكاوى و استفسارات الرقم التأميني تقريبا 18364 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 31558 و ادنى مستوى وصل الى 8011 شكوى و استفسار

• اصحاب المعاشات

يبلغ متوسط شكاوى و استفسارات اصحاب المعاشات تقريبا 7208 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 13749 و ادنى مستوى وصل الى 2120 شكوى و استفسار

ثانيا/ القطاع الحكومي

• التغطية التأمينية

يبلغ متوسط شكاوى و استفسارات التغطية التأمينية تقريبا 1176 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 2894 و ادنى مستوى وصل الى 356 شكوى و استفسار

• الرقم التأميني

يبلغ متوسط شكاوى و استفسارات الرقم التأميني تقريبا 6166 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 10278 و ادنى مستوى وصل الى 400 شكوى و استفسار

• اصحاب المعاشات

يبلغ متوسط شكاوى و استفسارات اصحاب المعاشات تقريبا 1021 شكوى و استفسار و اعلى مستوى للشكاوى و الاستفسارات وصل الى 1569 و ادنى مستوى وصل الى 20 شكوى و استفسار

المقارنة بين معدلات الشكاوى الشهرية
(القطاع العام و الخاص) و (القطاع الحكومي)

id	Variables			
	التغطية التأمينية	الرقم التأميني	اصحاب المعاشات	
قطاع خاص	Mean	<u>5371.36</u>	<u>18364.92</u>	<u>7208.19</u>
	N	36	36	36
	Std. Deviation	2981.659	7395.689	3244.189
	Maximum	8932	31558	13749
	Minimum	1334	8011	2120
قطاع حكومي	Mean	<u>1176.36</u>	<u>6166.47</u>	<u>1021.83</u>
	N	36	36	36
	Std. Deviation	624.404	3066.085	521.200
	Maximum	2894	10278	1569
	Minimum	356	400	20
Total	Mean	3273.86	12265.69	4115.01
	N	72	72	72
	Std. Deviation	3006.028	8325.961	3876.171
	Maximum	8932	31558	13749
	Minimum	356	400	20

(جدول 3)

الاحصاء الوصفي

الانحراف المعياري	المتوسط	الحد الأقصى	الحد الأدنى	عدد المشاهدات	
3006.028	3273.86	8932	356	72	التغطية التأمينية
8325.961	12265.69	31558	400	72	الرقم التأميني
3876.171	4115.01	13749	20	72	اصحاب المعاشات
				72	Valid N (listwise)

(جدول 2)

يتضح من الجدول (2)

● التغطية التأمينية

- ان متوسط عدد العاملين الذين يعملون في منشآت سواء قطاع عام وخاص او قطاع حكومي الذين يشكون من عدم وجود تغطية تأمينية لهم او لم يحصلوا عليها يصل الى حوالي 3273 عامل
- هذا و قد وصل اعلى مستوى لعدد الاشخاص الذين يشكون من عدم وجود تغطية تأمينية لهم الى 8932 شخص خلال شهر مارس 2012 و ادنى مستوى الى 365 شخص خلال شهر يناير 2010

● الرقم التأميني

- ان متوسط عدد المواطنين الذين لم يتم تسجيل رقم تأميني لهم او لم يحصلوا عليه يصل الى حوالي 12265 مواطن
- هذا و قد وصل اعلى مستوى لعدد الاشخاص الذين يشكون من عدم تسجيل رقم تأميني لهم او لم يحصلوا عليه الى 31558 شخص خلال شهر فبراير 2011 و ادنى مستوى الى 400 شخص خلال شهر يناير 2010

● اصحاب المعاشات

- ان متوسط عدد العاملين الذين تقدموا للحصول على معاش و لم يحصلوا عليه لأي سبب ما او يتضررون من احتساب قيمة المعاش يصل الى حوالي 4115 عامل
- هذا و قد وصل اعلى مستوى لعدد العاملين الذين تقدموا للحصول على معاش و لم يحصلوا عليه لأي سبب ما او يتضررون من احتساب قيمة المعاش الى 13749 شخص خلال شهر فبراير 2011 و ادنى مستوى الى 20 شخص خلال شهر يناير 2010

دراسات سابقة

- ابراهيم، منى دراسة تحليلية لإقتصاديات نظم التأمينات الإجتماعية والمعاشات

رسالة ماجستير كلية التجارة و ادارة الاعمال - جامعة حلوان 2007 (1)

تهدف هذه الدراسة إلى تقييم الوضع الحالي لنظام التأمينات الإجتماعية في مصر مقارنة بنظام التأمينات الإجتماعية في دولتي شيلي والسويد وتطرح تساؤل عن التحديات التي تواجه هذا النظام ، هل تقتصر فقط علي التوظيف غير الكفاء للمال الإحتياطي للنظام؟. وتتوصل في ضوء الأسس النظرية التي تحكم نظم التأمينات الإجتماعية ، وما رسمه الأدب الإقتصادي من ضوابط تحكم تلك النظم إلي أنه علي الرغم من حدة التوظيف غير الكفاء للمال الإحتياطي للنظام إلا أنه لا يمثل التحدي الوحيد الذي يواجه النظام فهناك تحديات أخرى . لذا فإن هذه الدراسة ترسم سياسة واضحة للتغلب علي التحديات التي تواجه نظام التأمينات الإجتماعية في مصر بإقتراح نظام تأميني أكثر كفاءة في ضوء تجربتي دولتي المقارنة ، وفي ضوء ما شهدته نظم التأمينات الإجتماعية من تحولات علي مستوي العالم .

- خيرى ، أمنية الاستدامة المالية لنظام التأمينات الاجتماعية فى مصر رسالة ماجستير كلية التجارة جامعة

عين شمس 2008.(4)

و قد قامت باستعراض وعلى مدار أربعة فصول تحوى بداخلها على عشرة مباحث ، كيفية تحقيق الإستدامة المالية لنظام التأمينات الإجتماعية المطبق فى مصر التى تمكن النظام من أداء التزاماته تجاه المستفيدين من النظام بشكل يكفله لهم القانون، وقد تم تناول ذلك بدراسة طبيعة النظم المختلفة للمعاشات والتأمينات وكذلك الإشارة الى تجارب بعض الدول ، فضلا على التعرف على طبيعة نظام المعاشات والتأمينات المطبق فى مصر وتقييم هذا النظام من أجل إبراز النواحي الإيجابية للنظام ودعمها وتلافى النواحي السلبية التى يفرزها النظام .

إن الحكومات المتعاقبة فى جمهورية مصر العربية جعلت منها دولة متميزة سواء من حيث طول فترة تطبيق النظام والتي تعدت الخمسون عاما ، أو من حيث التوسع فى مد المظلة التأمينية للمواطنين كونهم مشتركين أو مستفيدين من النظام ، الأمر الذى تطلب معه تطبيق نظام جديد للمعاشات وهو ما طرحته حكومتنا المصرية فى الآونة الأخيرة مستعينة بالخبرة الدولية فى هذا المجال وبما يتلاءم مع خصوصية الاقتصاد المصرى.

المبحث الثالث

التحليل الإحصائي

في هذا الجزء سوف نقوم بالدراسة الإحصائية للبيانات الخاصة بالشكاوى و الاستفسارات الواردة بالهيئة القومية للتأمين الاجتماعي بقطاعيها (القطاع العام و الخاص - القطاع الحكومي) و ذلك خلال الفترة الزمنية من 2010/1/1 حتى 2012/1/31 و قد تمثلت متغيرات الدراسة في :-

X1 و هي تمثل شكاوى التغطية التأمينية (عدد الأشخاص العاملين الذين ليس لهم اي بيانات تأمينية تخص الوظيفة الحالية و الوظائف السابقة و تواريخ الالتحاق في كل وظيفة و الأجر الخاص بكل وظيفة و تدرجه)

X2 و هي تمثل شكاوى الرقم التأميني (عدد المواطنين الذين لم يتم تسجيل رقم تأميني لهم و هو مثل الرقم القومي يتم تسجيل كافة البيانات الشخصية عليه لمعرفة هل الشخص مؤمن عليه ام لا و لابد من وجود رقم تأميني لكل مواطن حيث انه يتم تسجيل التغطية التأمينية على هذا الرقم و لكل مواطن رقم تأميني واحد فقط)

X3 و هي تمثل شكاوى اصحاب المعاشات او الذين تقدموا للحصول على معاش و لم يحصلوا عليه لأي سبب ما او ينضرون من احتساب المعاش

2/1/2 القانون رقم 108 لسنة 1976

ويقتصر هذا القانون علي تغطية ثلاثة أخطار فقط هي الشيخوخة ، والعجز ، والوفاة ،

وتنظم أحكام هذا القانون التأمين الإجتماعي علي أصحاب الأعمال

3/1/2 القانون رقم 50 لسنة 1978

وتنظم أحكام هذا القانون نظام التأمين الإجتماعي علي العاملين المصريين في الخارج من

غير الخاضعين لأحكام قانون التأمين الإجتماعي رقم 79 لسنة 1975 ، وتتمثل الفئات التي

يشملها في «عاملين المرتبطين بعقود شخصية ، والعاملين لحساب أنفسهم .

ويغطي هذا القانون أخطار الشيخوخة ، والعجز ، والوفاة ، بالإضافة إلي معاش إصابة العمل

في حالتها إنتهاء العمل للعجز الكامل ، أو الوفاة ، وتجدر الإشارة هنا إلي أن الإشتراك في

هذا النظام يعد إشتراكا إجتماعيا (٦)

4/1/2 القانون رقم 112 لسنة 1980

ينصب هذا القانون المعروف باسم قانون التأمين الشامل علي العمالة غير المنتظمة التي

لا تشملها بالتغطية قوانين أخرى للتأمين الإجتماعي ، بالإضافة إلي أولئك الذين لا يوجد لهم

مكان محدد للعمل ، كما يمنح معاشا يسمى معاش السادات بالإضافة إلي معاش التأمين

الشامل. (٥)

وبموجب هذا القانون إمتدت الحماية التأمينية لصغار المشتغلين لحساب أنفسهم ويمثلون

(12) فئة ، مع إمتداد الحماية التأمينية لكل من بلغ سن الخامسة والستين ، أو وقعت وفاته ،

أو ثبت عجزه الكامل ولم يستحق معاشا وفقا لقوانين التأمين الإجتماعي الأخرى قبل

1980/7/1 ، ويغطي هذا القانون ثلاثة أخطار فقط هي الشيخوخة ، والعجز ، والوفاة وتجدر

الإشارة هنا إلي وجود قانون للتأمين ينتفع به أفراد القوات المسلحة وحدهم هو القانون رقم

90 لسنة 1975.

(٦) السيد محمود السيد غانم ، مرجع سبق ذكره .،

(٥) محمد، علي مرجع سبق ذكره

1/1/2 القانون رقم 79 لسنة 1975

تنظم أحكام هذا القانون نظام التأمين الإجتماعي لفئة العاملين بأجر لدي الغير بالقطاع الحكومي وقطاع الأعمال العام والخاص ، وتتكون الفئات الخاضعة لهذا القانون من العاملين المدنيين بالجهاز الإداري للدولة والهيئات العامة وقطاع الأعمال العام ، والعاملين الخاضعين لأحكام قانون العمل وعمال الخدمة المنزلية.

وتوفر أحكام هذا القانون تقديم التأمين ضد ستة من الأخطار الإجتماعية وهي الشيخوخة والعجز ، والوفاة ، والمرض ، وإصابة العمل ، والبطالة ، وتضمن هذا القانون إنشاء صندوق التأمينات للعاملين بالحكومة وصندوق التأمينات للعاملين بقطاعي الأعمال العام والخاص .

معدلات الاشتراكات التي يدفعها كل من العامل وصاحب العمل

وفقا للقانون رقم 79 لسنة 1975

جدول رقم (1)

نوع التأمين	أجر أساسي ⁽⁶⁾			أجر متغير ⁽⁷⁾		
	صاحب العمل			صاحب العمل		
	مؤمن عليه	قطاع خاص	قطاع عام	مؤمن عليه	قطاع خاص	قطاع عام
الشيخوخة والعجز والوفاة	15%	15%	15%	10%	15%	15%
مكافأة	2%	2%	2%	3%	-	-
إصابات العمل	1%	2%	3%	-	3%	2%
المرض	3%	3%	4%	1%	3%	4%
البطالة	-	2%	2%	-	2%	-

المصدر

محمد حامد الصياد ، ليلي محمد الوزيري ، التمويل ومجال التطبيق في قانون التأمين الإجتماعي الصادر بالقانون

رقم 79 لسنة 1975 ، معهد التأمينات ، المؤسسة الثقافية العمالية ، الاتحاد العام لنقابات عمال مصر ، ص 14

(١) يختلف مفهوم الأجر الأساسي باختلاف العلاقة التي تربط العامل بصاحب العمل ، بالنسبة لمن تربطهم بصاحب العمل علاقة لائحية يشير الأجر الأساسي إلى الأجر المنصوص عليه في الجداول المرفقة بنظم التوظيف ، وبالنسبة لمن تربطهم بصاحب العمل علاقة تعاقدية يقصد بالأجر الأساسي الأجر المحدد بعقد العمل كأجر أساسي وما يطرأ عليه من علاوات مستبعدة منه العناصر التي تعتبر جزءا من الأجر المتغير .

(٢) يقصد بالأجر المتغير كل ما يحصل عليه المؤمن عليه من مقابل نقدي بالإضافة إلى الأجر الأساسي أيا كانت قواعد حساب هذا المقابل سواء كان محسوباً بالمدة أو بالانتاج وسواء كان مقابل جهد بذل خلال ساعات العمل الأصلية أو خلال ساعات إضافية ، وكذلك تلك الأجزاء من الأجر التي تمنح لمواجهة الزيادة في تكاليف وأعباء المعيشة وتلك التي يقصد بها التعويض عن الجهود غير العادية.

نحو الإتساع من حيث التغطية، ونوع الحماية التي إقتصرت في بادئ الأمر علي تغطية أخطار الشيخوخة ، والعجز ، والوفاة. (6)

فقد إمتدت تلك الحماية لتستكمل الدولة كافة أنواع خدمات التأمين الإجتماعي من خلال القانون رقم 50 لسنة 1963 للعاملين بالقطاع الحكومي والقانون رقم 63 لسنة 1964 لكافة فئة العاملين لحساب الغير.

ثم يأتي القانون رقم 79 لسنة 1975 وما أدخل عليه من تعديلات نتيجة لما أسفر عنه التطبيق العملي له من ضرورات تستوجب هذا التعديل ليمثل بداية التوحيد لكافة التشريعات لتخضع فئة العاملين بالدولة لقانون واحد ايا كان القطاع الذي يعملون به ، ليشكل بذلك بداية الطفرة الحقيقية لنظام التأمينات الإجتماعية. (3)

1/2 الإطار القانوني لنظام التأمينات الإجتماعية المصري:

تمتلك مصر نظاما شاملا للضمان الإجتماعي يجمع بين جوانبه مدخلا الضمان الإجتماعي ، حيث يقوم المدخل الأول علي الإشتراكات المسبقة ممثلا في نظام التأمين الإجتماعي ، ويقوم المدخل الثاني علي تقديم الإعانات المالية ممثلا في معاشات الضمان الإجتماعي .

ونتيجة لكون هذه الدراسة تنصب فقط علي نظام التأمينات الإجتماعية ، ولا تتطرق إلي دراسة نظام الضمان الإجتماعي ككل فمن الضرورة عرض التشريعات التي تقتصر فقط علي التأمين الإجتماعي في مصر ، وذلك لتحديد الإطار القانوني الذي يحكم فقط نظام التأمين الإجتماعي محور هذه الدراسة .

وتحكم منظومة التأمين الإجتماعي في مصر في الوقت الراهن والتي تخضع للإشراف المباشر للهيئة القومية للتأمين الإجتماعي أربعة قوانين رئيسية صدرت تباعا منذ عام 1975، وتتمثل في :

(6) محمود، السيد ، نشأة و تطور التأمينات الاجتماعية ، مؤتمر التأمينات الاجتماعية بين الواقع و المأمول ،جامعة الأهر ، 13-15 أكتوبر 2002 ،

(3) الصياد ، محمد و الوزيري ، ليلي ، مرجع سبق ذكره ،

المبحث الثاني

الإطار العام لنظام التأمينات الاجتماعية

في جمهورية مصر العربية

الحماية من الخطر الاجتماعي وسيلة ليست بالغريبة علي المجتمع المصري ، فهي تضرب بجذورها في التاريخ المصري ، لتبدأ معه ، وتسير في طريقه ، لتتأثر ، وترتبط بما يؤل إليه الوضع السياسي المصري بدءا من تبعية مصر للدولة الإسلامية فقد تأثر مفهوم الحماية من الخطر الاجتماعي بالمفاهيم الإسلامية للرعاية الاجتماعية والتي أهمها الزكاة. (2)

ثم يمضي الحال لتبدأ مصر أول تعارف لها علي نظم المعاشات ولكن علي نطاق ضيق في منتصف القرن التاسع عشر وذلك بصدر أول قانون للمعاشات المدنية في 26 ديسمبر عام 1854 والمعروف آنذاك باسم قانون سعيد باشا. (5)

وقد تلا هذا القانون مجموعة من القوانين الأخرى إقتصرت جميعها علي معالجة أخطار معينة معالجة قاصرة وعاجزة، لتمثل تلك الحقبة من تاريخ مصر التي صدرت خلالها هذه القوانين إرهاصات لقوانين التأمينات الاجتماعية.

ثم تبدأ مصر في تطبيق نظم التأمينات الاجتماعية بمفهومها العلمي مع بداية ثورة يوليو عام 1952 بصدر القانون رقم 316 لسنة 1952 لإنشاء نظاما للتأمين ، وآخر للدخار لموظفي الحكومة ، ثم صدر القانون رقم 419 لسنة 1955 لإنشاء أول نظام للتأمين والدخار للعاملين بالقطاع الخاص. (3)

وقد تلا هذان القانونان مجموعة من القوانين تشكلت خلالها السمة الرئيسية التي ميزت التأمينات الاجتماعية بعد ثورة يوليو عام 1952 والممثلة في إتجاه نظام التأمينات الاجتماعية

(2) الصياد ، محمد دور التأمينات الاجتماعية في شبكة الأمان الاجتماعي ، ملحق 2 ، المؤتمر القومي للتنمية الاجتماعية ، وزارة التأمينات والشؤون الاجتماعية ، القاهرة ، 17-19 ديسمبر ، 2000 ،

(3) محمد ، علي واقع التأمين الاجتماعي في مصر ، مؤتمر التأمينات الاجتماعية بين الواقع والمأمول ، جامعة الأزهر ، 13-15 أكتوبر ، 2002 ، .

(3) الصياد ، محمد و الوزيري ، ليلي التمويل ومجال التطبيق في قانون التأمين الاجتماعي الصادر بالقانون رقم 79 لسنة 1975 ، معهد التأمينات ، المؤسسة الثقافية العمالية ، الاتحاد العام لنقابات عمال مصر ،

ويمكن على سبيل الحصر تناول بعض من هذه المخاطر التى يجب على النظام أن

يغطيها ألا وهى :

1- انخفاض الدخل أو انقطاعه

2- الوفاة

3- البطالة

4- العجز

5- الشيخوخة

6- المرض

7- عدم التكيف الإجتماعى

مما سبق نلاحظ أن

جميع المخاطر السالف ذكرها تعد من المخاطر الإجتماعية فى طبيعتها فالكل معرض لها أو لأحدها ولا يمكن للفرد دفعها أو تحمل أثارها بمفرده ومن هنا تتبع أهمية النظام التأمينى وتشريعاته التى تمنح للأفراد تغطية هذه المخاطر فى حالة حدوثها .

المبحث الأول

مفهوم التأمين الاجتماعى

يمكن تعريف نظام التأمينات الاجتماعية كآلاتى :

" التأمينات الاجتماعية هى نظام اجتماعى قانونى يهدف لتحقيق الأمن الاجتماعى والسياسى لأفراد المجتمع فى حالة تعرضهم لأحد المخاطر المهنية أو الاجتماعية وذلك عن طريق استخدام مجموعة من الأساليب الفنية الخاصة بهذا النظام والتي تعمل على إعادة توزيع الدخل القومى بشكل أكثر عدالة ."

والأفراد المشتركين فى نظام التأمين الاجتماعى لهم الحق فى الاستفادة من المزايا التى منحها لهم القانون هم وذويهم من بعدهم نظرا لاشتراكهم طيلة فترة عملهم فى هذا النظام .

الهدف من التأمينات الاجتماعية

توفير الحياة الكريمة والأمان الاجتماعى من أهم ما يمنحه النظام من مزايا للأفراد المشتركين فيه ولتحقيق هذه المزايا فلا بد من أن تشمل الجوانب الآتية :

أ - الجانب الإقتصادى

ب- الجانب الاجتماعى

ج- الجانب السياسى

شمول التغطية التأمينية

المقصود من شمول التغطية التأمينية هو أن يكون كل فرد من أفراد المجتمع مؤمنا عليه سواء كان هذا التأمين بصورة مباشرة أو بصورة غير مباشرة بمعنى أنه سيكون عند وقت وقوع الخطر المؤمن ضده صاحب معاش أو مستحقا عن صاحب معاش.

بعد جانب الإلزام هاما فى التغطية التأمينية مما أدى ذلك الى زيادة شمول التغطية بمعنى أنه يجب أن يكون هناك إلزاما لكل مواطن يعمل أو قادر على العمل فى أن يصبح خاضعا للتأمين الاجتماعى مهما كان عمره (فيما بين السن المسموح به للعمل و سن التقاعد) أو عمله أو مكان تواجده.

إن شمول التغطية التأمينية يمنح الأفراد شعورا بالاطمئنان على الحاضر ويزول معه الخوف والقلق من المستقبل.

شمول الأخطار

المقصود بشمول الأخطار هو أن يشمل النظام التأمينى كافة المخاطر التى يتعرض لها

أفراد المجتمع سواء للشخص المؤمن عليه أو لأحد أفراد أسرته وأن يكفل النظام لمموليه الحماية من هذه المخاطر.

نموذج احصائي للتنبؤ بعدد الشكاوى الواردة بالهيئة القومية للتأمينات الاجتماعية

محمد سيد احمد

أحمد أمين الشيخ^١

الملخص

تهدف هذه الدراسة إلى تقييم الوضع الحالي لعدد الشكاوى الواردة إلى الهيئة القومية للتأمينات الاجتماعية بقطاعيها (العام و الخاص) و (الحكومي) و طرح تساؤلات عن كيفية وضع نموذج لقياس مستوى تحسين الخدمة التأمينية وكيفية الحد من عدد الشكاوى الواردة في الفترات الزمنية المستقبلية

الكلمات الدالة :-

معامل ارتباط بيرسون - السلاسل الزمنية - التأمينات الاجتماعية

المقدمة

نشأت فكرة التأمين الإجتماعي لحماية الإنسان ضد ما يسمى بالخطر الإجتماعي ، فالإنسان هو الكائن الوحيد الذي يعيش الزمن بكافة أبعاده ، الماضي يخبراته وتجاريه للاستفادة منه ، والحاضر بالمعيشة فإن كان مثمرأ نال منه ، وإن كان شراً حاول تجنبه وتفاديه ولكن المستقبل مجهول بالنسبة له والإنسان بطبيعته يخشى المجهول فيحاول أن يعمل جاهداً على توفير الوسائل التي تحقق له ما يسمى بالحماية الإجتماعية .

الحماية الإجتماعية هي ما تمنحه نظم التأمين الإجتماعي لأفراد المجتمع، ويقصد بالحماية الإجتماعية السياسات والبرامج التي تحد من الفقر وتعزز قدرة الأفراد على حماية أنفسهم ضد المخاطر التي يسببها انقطاع أو فقدان الدخل

مشكلة الدراسة

في ضوء إختلاف الإتجاهات ما بين مؤيد ومعارض تجاه ما طُرح من إقتراحات تتعلق بقضية التأمينات الاجتماعية في مصر ، والتساؤلات التي أثارها تلك القضية ، وفي ظل ما تشهده نظم التأمينات الاجتماعية في مختلف دول العالم من تحول ، فإن الأمر يقتضي تحليلاً واضحاً يستند إلى أساس نظري و احصائي من خلال إجراء دراسة كيفية تطوير الاداء الخدمي و ايجاد حلول للشكاوى الواردة و التقليل منها حتى يتسنى تقديم الخدمات التأمينية بسهولة و يسر و دون اعاقه او اطالة للوقت و ذلك عن طريق دراسة و تحليل الشكاوى الواردة و الوقوف على اسبابها و اصدار القرارات و التعليمات التي تحد من تكرار هذه الشكاوى و التنبؤ بعددها في السنوات المقبلة و ذلك حتى يكون لدينا معيار لتقييم الاداء في كافة افرع التأمينات (معاشات - اشتراكات - تغطية تأمينية - رقم تأميني)

الهدف من الدراسة

على الرغم من وجود بيانات مجمعة عن عدد الشكاوى و اقسامها فإننا لم نضع معيار لقياس مستوى تحسين الخدمات التأمينية لذلك نههدف بوضع بيان بعدد الشكاوى المتوقعه في كافة الأقسام في السنوات المقبلة و ذلك حتى يكون لدينا معيار لتقييم الاداء في كافة افرع التأمينات (معاشات - اشتراكات - تغطية تأمينية - رقم تأميني) و عرض البيانات الاحصائية امام المتخصصين في مجال التأمينات لاصدار القرارات و التعليمات التي تحد من تكرار هذه الشكاوى

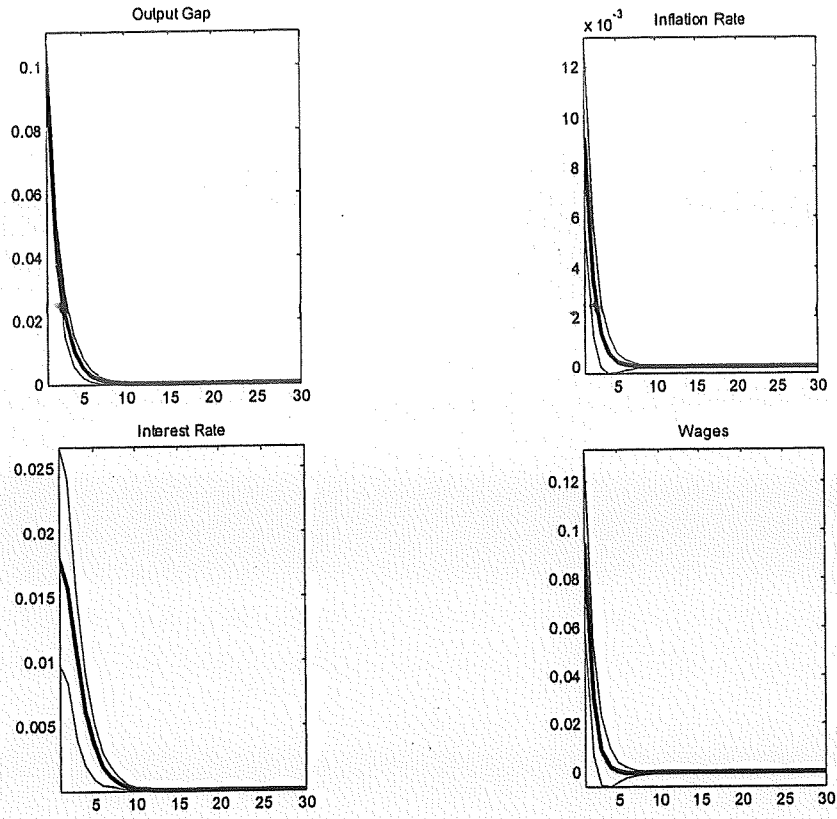
المقاييس الاحصائية المستخدمة

تم عرض المشكلة محل الدراسة و هي وضع نموذج للتنبؤ بعدد الشكاوى الواردة في الفترات المقبلة و الهدف منها تقييم الاداء في كافة افرع التأمينات (معاشات - اشتراكات - تغطية تأمينية - رقم تأميني -) باستخدام الجداول الاحصائية برنامج SPSS18 و برنامج MINITAB16

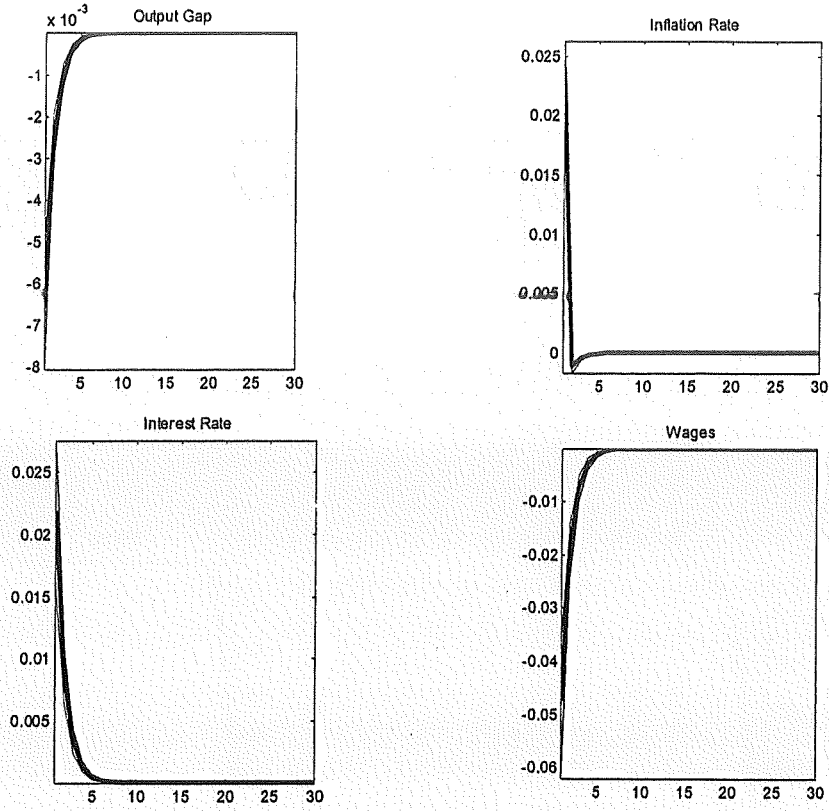
١ رئيس قسم الاحصاء التطبيقي - معهد الدراسات و البحوث الاحصائية - جامعة القاهرة

٢ عضو الادارة العامة للمتابعة و الرقابة - الهيئة القومية للتأمينات الاجتماعية

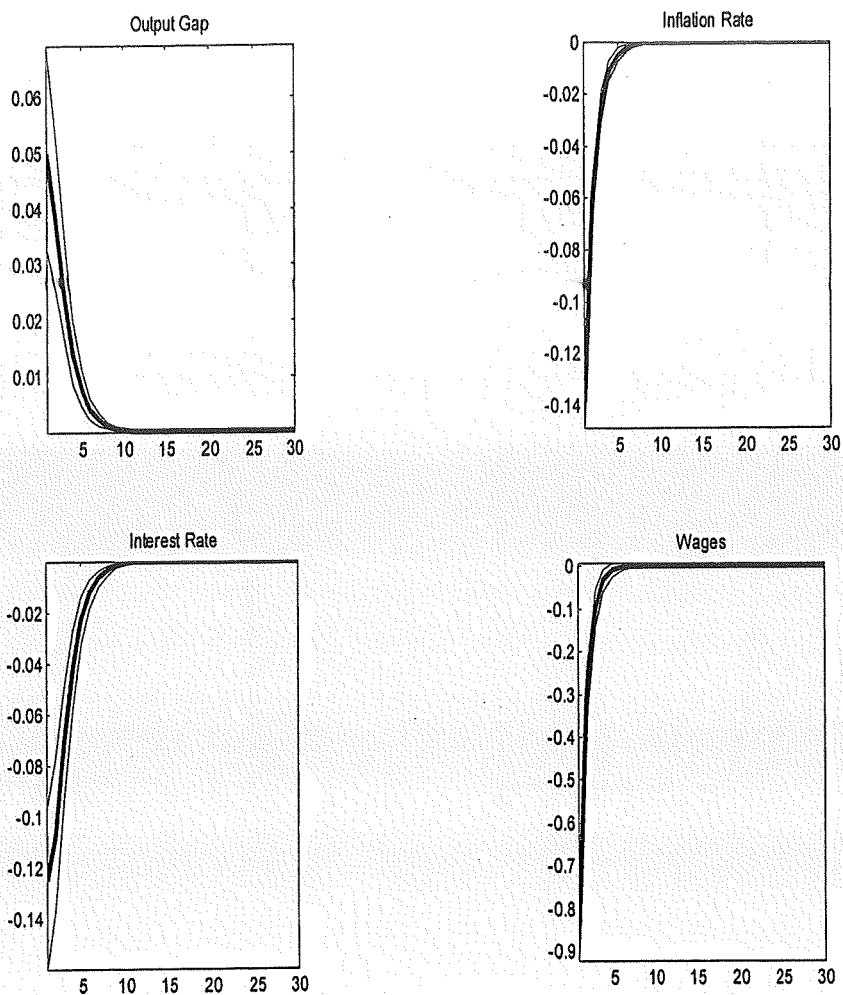
شكل رقم (٨)
دوال استجابة المتغيرات المشاهدة لصدمة تفضيل المستهلك



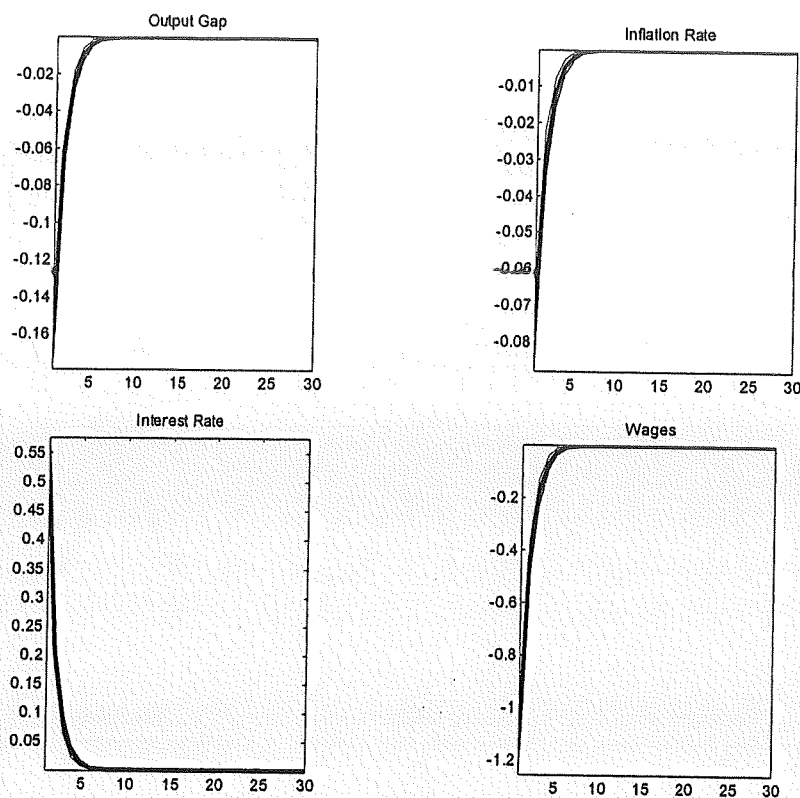
شكل رقم (٧)
دوال استجابة المتغيرات المُشاهدة لصدمة هامش الربح



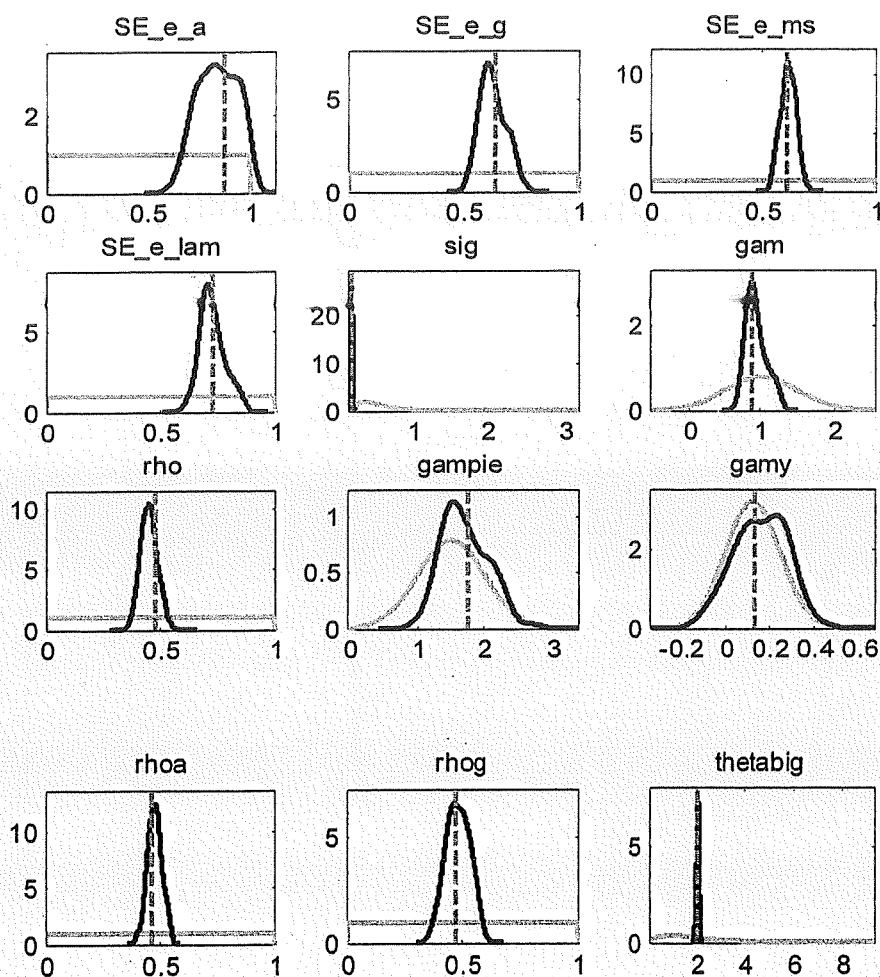
شكل رقم (٦)
دوال استجابة المتغيرات المُشاهدة للصدمة التكنولوجية



شكل رقم (٥)
دوال استجابة المتغيرات المشاهدة للصدمة النقدية

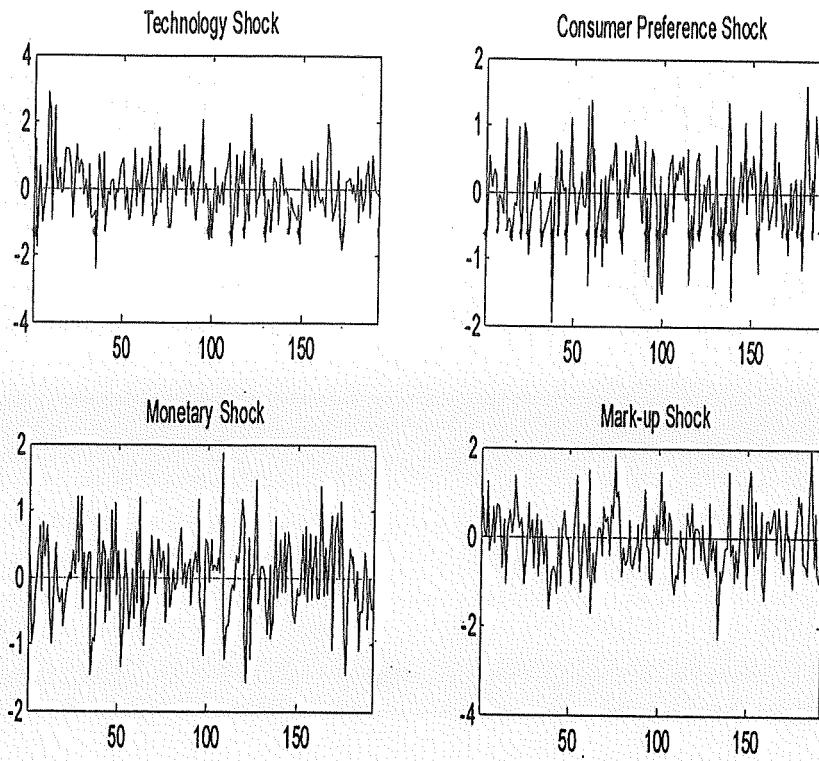


شكل رقم (٤)
التوزيعات الاحتمالية القبلية والبعدية لمعاملات النموذج

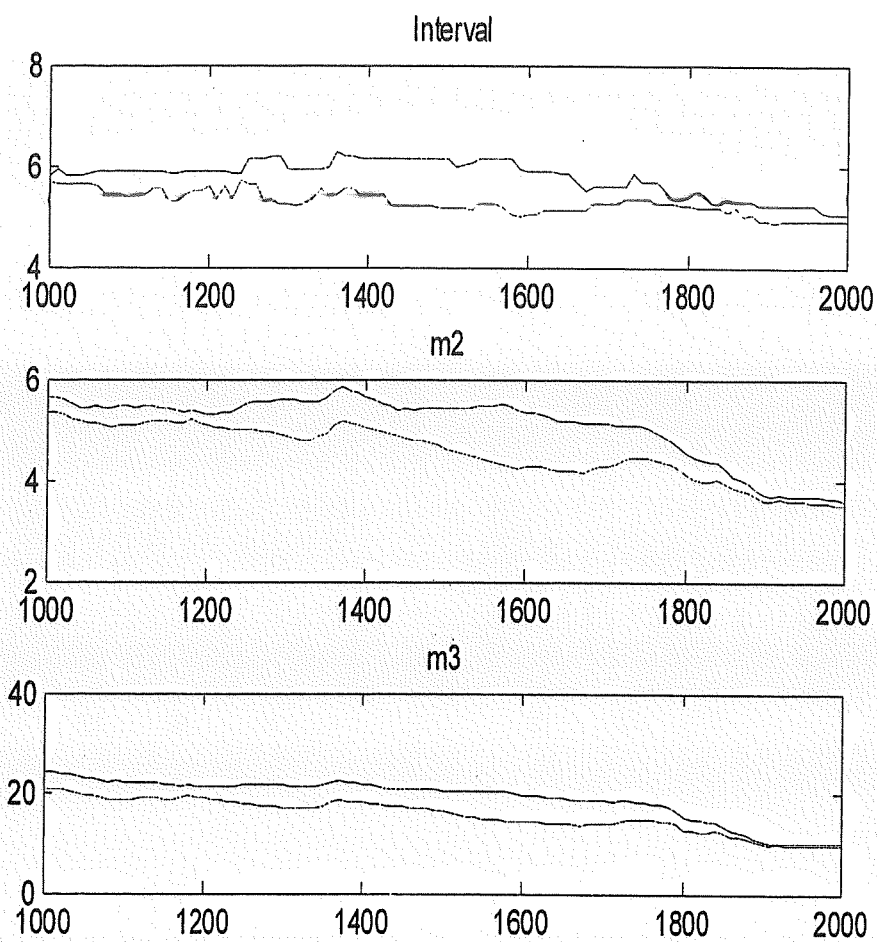


■ التوزيعات الاحتمالية القبلية
■ التوزيعات الاحتمالية البعدية
■ المنوال

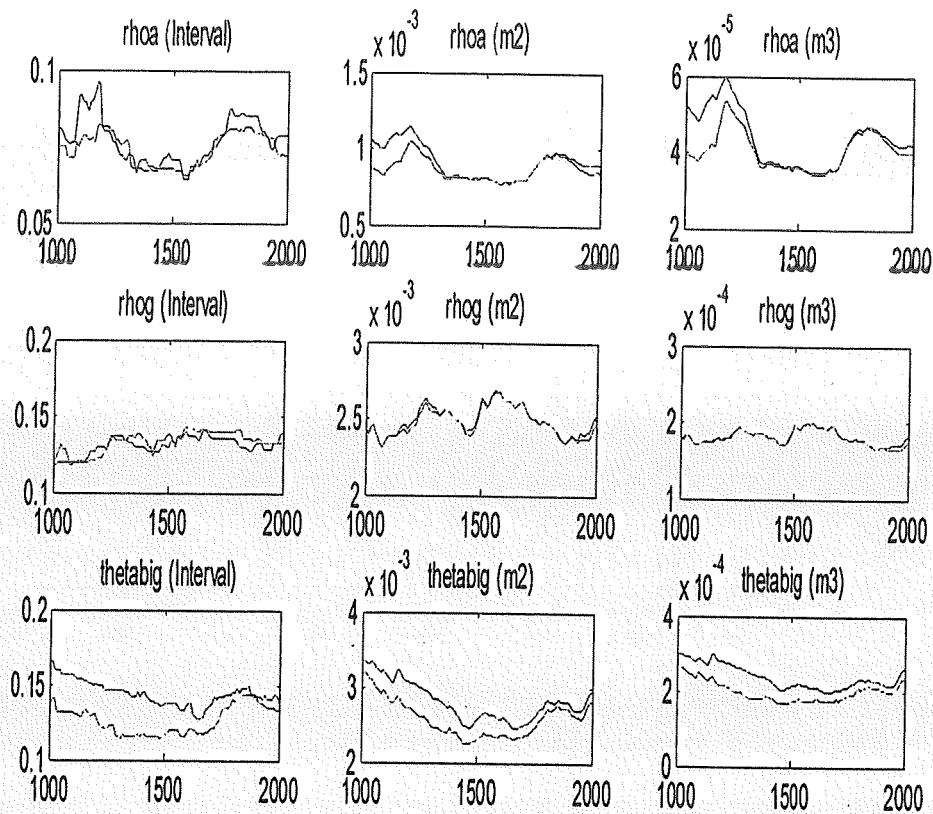
شكل رقم (٣)
"اختبار الصدمات المُمهدة" باستخدام MCMC



شكل رقم (٢)
"اختبار الفحص المتعدد" لمعطيات النموذج باستخدام MCMC

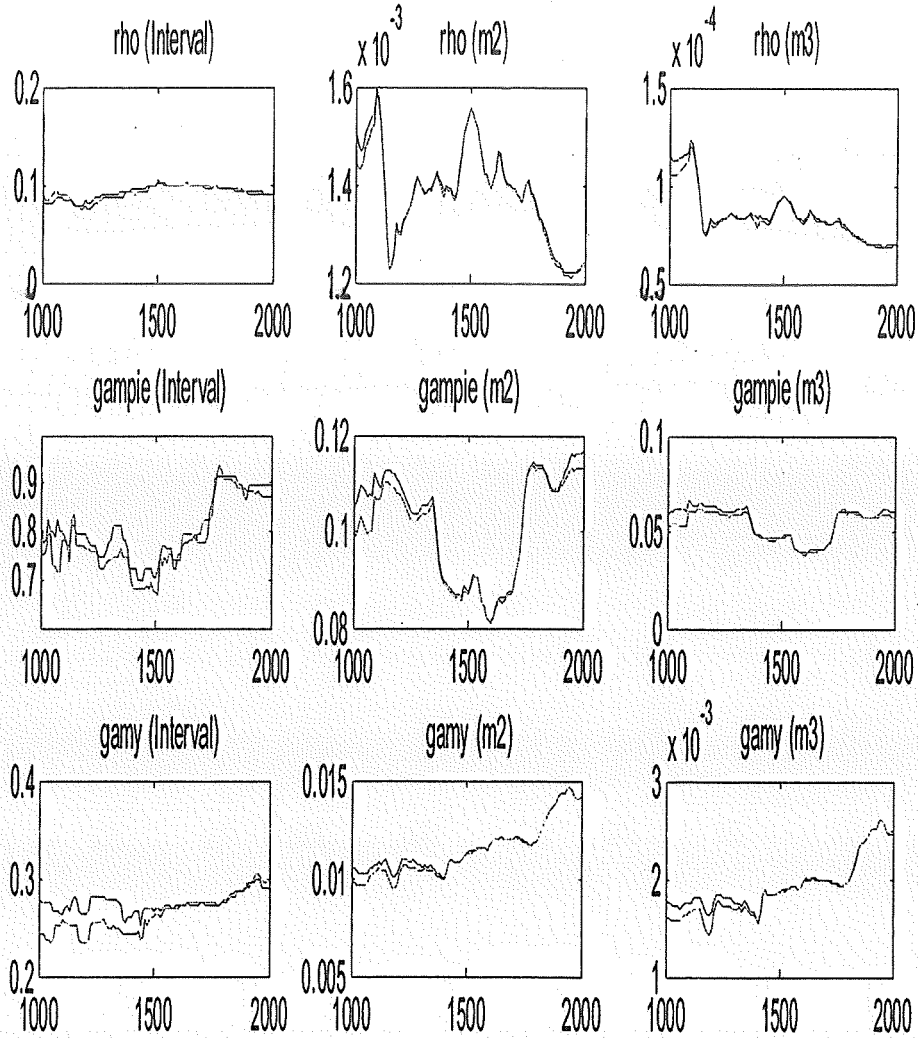


شكل رقم (٤-١)
"اختبار الفحص الأحادي" لمعاملات النموذج باستخدام MCMC

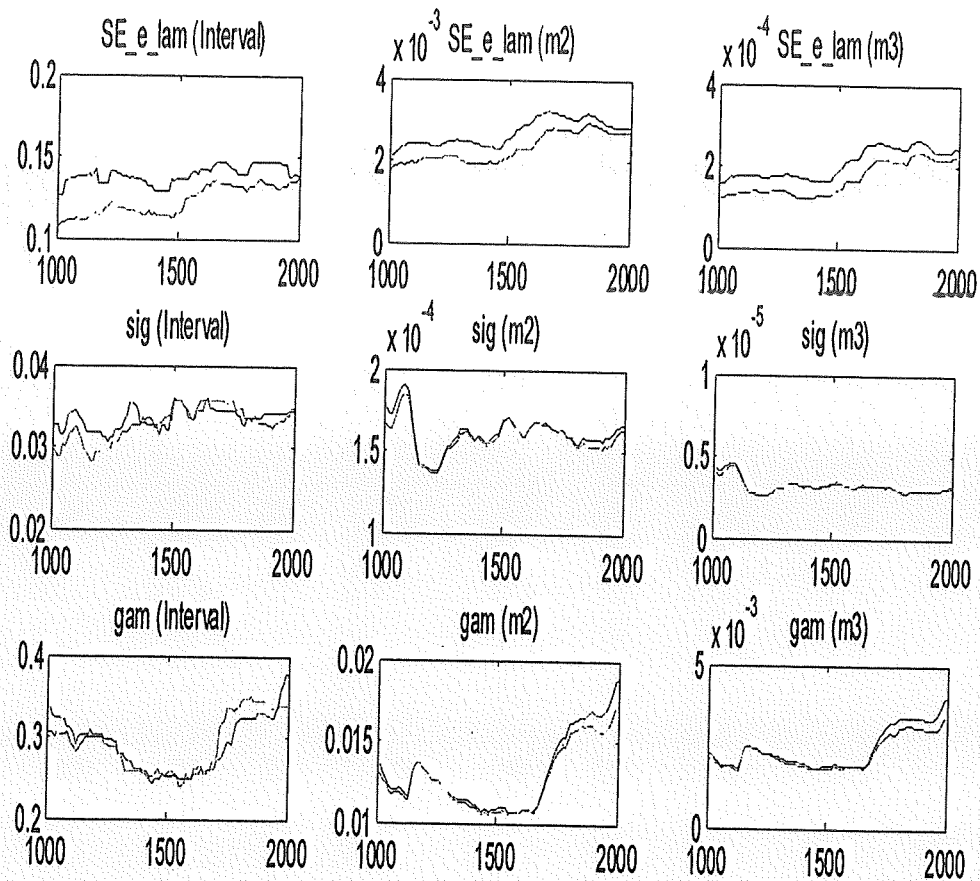


شكل رقم (٣-١)

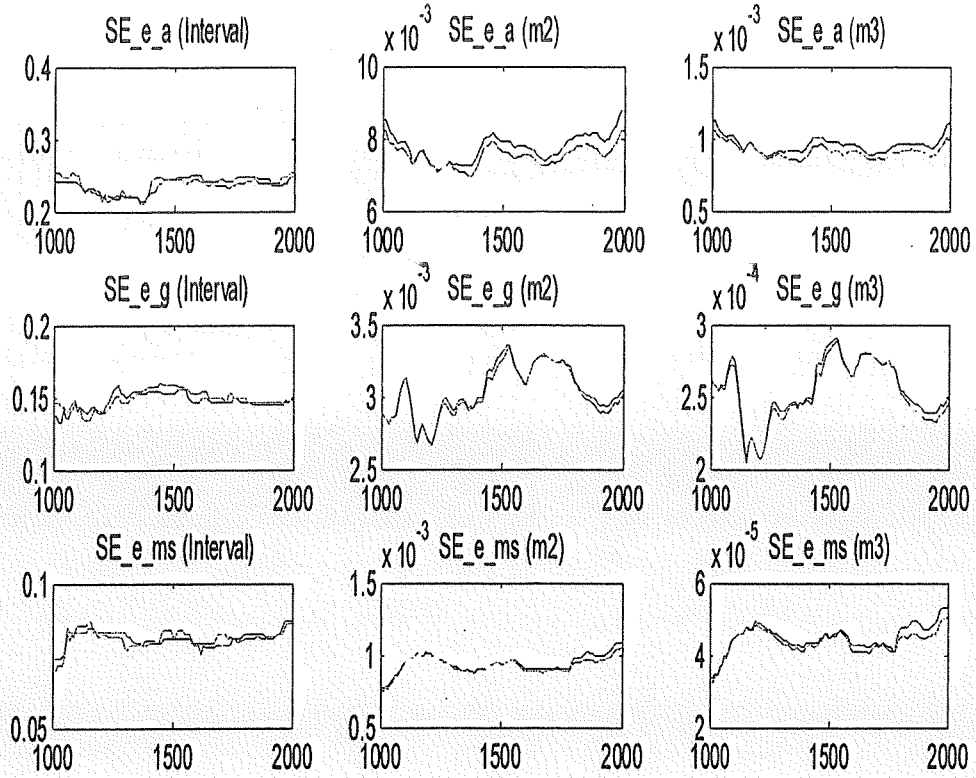
"اختبار الفحص الأحادي" لمعطيات النموذج باستخدام MCMC



شكل رقم (٢-١)
"اختبار الفحص الأحادي" لمعاملات النموذج باستخدام MCMC



شكل رقم (١-١)
"اختبار الفحص الأحادي" لمعطيات النموذج باستخدام MCMC



ملحق إحصائي

and a DSGE Model", **Journal of Applied Econometrics**, Vol. 25, Issue 4 (New York, NY: John Wiley & Sons Ltd., 2010).

4. Griffoli, Tommaso Mancini: **Dynare V4 - User Guide (Public Beta Version)**, January 2013.
(<http://www.dynare.org/documentation-and-support/user-guide/Dynare-UserGuide-Web Beta.pdf>)
5. Guerrón-Quintana, Pablo A. & James M. Nason: **Bayesian Estimation of DSGE Models**, Federal Reserve Bank of Philadelphia Working Paper, No. 12-4 (Pennsylvania, PA: Federal Reserve Bank of Philadelphia, February 2012).
6. Kremer, Jana et al: "Dynamic Stochastic General Equilibrium Models as a Tool for Policy Analysis ", **CESifo Economic Studies**, Vol. 52, No. 4/2006, November 2006 (Oxford: Oxford University Press).
7. Liu, G. & R. Gupta: "A Small-Scale DSGE Model for Forecasting the South African Economy", **South African Journal of Economics**, Vol. 75, Issue 2, 2007, (Oxford: Blackwell Publishing Ltd.).
8. Ngalawa, Harold P.E. & Nicola Viegi: **Interaction of Formal and Informal Financial Markets in Quasi-Emerging Market Economies**, Economic Research Southern Africa (ERSA) working paper, No. 326 (Cape Town: ERSA, January 2013).
9. Peiris, Shanaka J. & Magnus Saxegaard: **An Estimated DSGE Model for Monetary Policy Analysis in Low-Income Countries**, IMF Working Paper, No. 07/282, December 2007.
10. Rabanal, Pau & Juan F. Rubio-Ramírez: **Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach**, Federal Reserve Bank of Atlanta Working Paper, No. 2001-22a (Georgia, GA: Federal Reserve Bank of Atlanta, February 2003).
11. Tovar, C. E.: **DSGE Models and Central Banks**, Bank for International Settlements (BIS) Working Paper, No. 258, September 2008 (Basel: BIS).

ويوضح الجدول رقم (٥) النتائج الخاصة بـ"تجزئ التباين" (Variance Decomposition)، والتي يمكن من خلالها التعرف على نسبة مساهمة كل صدمة من الصدمات الأربع التي يتضمنها النموذج في إلحاق التقلبات/التغيرات بكل من الناتج المحلي الإجمالي، ومعدل التضخم، وسعر الفائدة، والأجور. ومن هذه النتائج، يمكن ملاحظة أن الصدمات التي تلحق بالسياسة النقدية وتفضيل المستهلك هي من أكثر الصدمات الأربع تأثيراً في كل من الناتج المحلي الإجمالي (٦٤،٧% و ٢٧،٤%) ومستوى الأجور (٧١،٤% و ٢٧،٩%) في تنزانيا. في حين تتأثر التقلبات التي يشهدها كل من معدل التضخم (٧٠،٢% و ٢٧،٤%) وسعر الفائدة (٩٢،٤% و ٧،٢%) في تنزانيا بشدة بالصدمات التي تلحق بالسياسة النقدية والتكنولوجية. كما تبدو الصدمات النقدية أكثر تأثيراً على الناتج مقارنة بمعدل التضخم.

٦. خاتمة

لقد توصلت هذه الورقة البحثية إلى عدة نتائج، والتي يمكن إيجازها في النقاط الثلاث التالية:

- نجاح "بنك تنزانيا" في استهداف معدل التضخم نتيجة لالتزامه إلى حد كبير باتباع سياسة نقدية لا يشوبها تغيرات جذرية.
- الصدمات النقدية والتكنولوجية من أكثر الصدمات الهيكلية تأثيراً في الناتج المحلي الإجمالي، ومعدل التضخم، والأجور.
- أهمية الدور الذي تلعبه السياسة النقدية في الاقتصاد التنزاني. فمن خلال رسم سياسة نقدية سليمة، يستطيع "بنك تنزانيا" تحقيق الاستقرار في المستوى العام للأسعار والأجور، كما أنه يستطيع تحفيز معدلات النمو الاقتصادي.

٧. المراجع

أولاً: مراجع باللغة العربية

- ولاء محمد محروس: استخدام نماذج التوازن العام العشوائية الديناميكية في تحليل السياسات الاقتصادية في الدول الأفريقية دراسة تطبيقية على السياسة النقدية في تنزانيا، رسالة دكتوراه غير منشورة (القاهرة: معهد البحوث والدراسات الإفريقية، قسم السياسة والاقتصاد، د. ت.).

ثانياً: مراجع باللغة الإنجليزية

1. An, S. & F. Schorfheide: "Bayesian Analysis of DSGE Models", **Econometric Reviews**, Vol. 26, Issue 2-4, May 2007 (London: Taylor & Francis Group).
2. Berg, Andrew et al: **The Macroeconomics of Medium-Term Aid Scaling-Up Scenarios**, International Monetary Fund (IMF) Working Paper, No. 10/60 (Washington, D. C.: IMF, February 2010).
3. Edge, Rochelle M. et al: "A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced-Form Models,

– وفيما يخص "المعادلة الديناميكية لمنحنى IS"، تشير النتائج إلى أن قيمة الوسط الحسابي للمعلمة σ تبلغ حوالي ٠،١٧؛ أي أن اتباع سياسة نقدية توسعية (من خلال خفض r بنقطة مئوية واحدة) يساهم في زيادة الناتج المحلي الإجمالي وزيادة فجوة الناتج (بحوالي ٠،١٧) والعكس صحيح.

وتوضح الأشكال البيانية رقم (٥)، و(٦)، و(٧)، و(٨) بالملحق الإحصائي دوال استجابة المتغيرات المُشاهدة – والتي تم تقديرها بالمنهج البيزي وتُعرف بـ "دوال الاستجابة للصدمات البيزية" (Bayesian IRF) ^١ - لكل صدمة من الصدمات الأربع التي يتضمنها النموذج. وفيما يتعلق بدوال استجابة المتغيرات المُشاهدة للصدمة النقدية، نجد أن حدوث صدمة نقدية إيجابية (Positive Monetary Shock) يؤدي إلى انخفاض كل من فجوة الناتج ومعدل التضخم ومستوى الأجور وارتفاع سعر الفائدة، وذلك خلال السبعة أرباع الأول اللاحقة لحدوث الصدمة [انظر الشكل رقم (٥) بالملحق الإحصائي].

الجدول رقم (٥)

"تجزيء التباين" (%)

الصدمة المتغير	الصدمة التكنولوجية	صدمة تفضيل المستهلك	الصدمة النقدية	صدمة هامش الربح
الناتج المحلي الإجمالي	7.74	27.44	64.73	0.09
معدل التضخم	70.2	0.44	27.41	1.99
سعر الفائدة	7.21	0.19	92.47	0.31
الأجور	0.47	27.98	71.45	0.1

المصدر: من إعداد الباحثة

^١ للمزيد من التفاصيل عن هذه الدوال وكيفية تقديرها بالمنهج البيزي، انظر:

Ibid., pp. 28-34, p. 52.

تابع الجدول رقم (٤)

المعلمة	الوسط الحسابي	فترة ثقة للوسط الحسابي (بدرجة ثقة ٩٠%)
σ_{ε^a}	0.8394	[0.9869 - 0.7101]
σ_{ε^g}	0.6315	[0.7189 - 0.5515]
σ_{ε^m}	0.6181	[0.6731 - 0.5670]
$\sigma_{\varepsilon^{\lambda}}$	0.7226	[0.8290 - 0.6537]

المصدر: من إعداد الباحثة

ويوضح الجدول رقم (٤) قيم الأوساط الحسابية للتوزيعات الاحتمالية البعدية لمعاملات النموذج، والتي تم تقديرها باستخدام المنهج البيزي، كما يوضح فترة الثقة لهذه الأوساط الحسابية بدرجة ثقة ٩٠%. هذا، وسيتم التركيز فيما يلي على تحليل النتائج التي تتعلق بكل من السياسة النقدية، ومعدل التضخم، ومعدل النمو الاقتصادي:

- ففيما يتعلق بـ "قاعدة تايلور"، تشير النتائج إلى أن قيمة الوسط الحسابي للمعلمة ρ_r تبلغ حوالي ٠،٥؛ أي أن "بنك تنزانيا" التزم إلى حد كبير بإتباع سياسة نقدية لا يشوبها تغيرات جذرية. كما تشير النتائج إلى أن قيمة الوسط الحسابي للمعلمة γ_{π} تبلغ حوالي ١،٠٧، في حين تبلغ قيمة الوسط الحسابي للمعلمة γ_r حوالي ٠،١٦؛ ويدل ذلك على أن السياسة النقدية التي يطبقها "بنك تنزانيا" قد نجحت في استهداف معدل التضخم بصورة أكبر من استهداف/تحفيز معدل النمو الاقتصادي.

- وبالنسبة لـ "المعادلة الديناميكية لمنحنى فيليبس"، تشير النتائج إلى أن قيمة الوسط الحسابي للمعلمة θ تبلغ حوالي ٢،٠٣٦. وبناء عليه، تبلغ قيمة θ_r حوالي ٠،٦٧ وقيمة κ_p حوالي ٠،٠٣٧٩. ويدل ذلك على وجود علاقة طردية ضعيفة بين معدل التضخم والتكلفة الحدية للمنتجين المحليين في تنزانيا.

الإحصائي- أن الصدمات الأربع بالنموذج، والتي تم تقديرها باستخدام MCMC، تتقلب صعوداً وهبوطاً حول القيمة الصفرية؛ مما يؤكد على الوصول لحل وحيد للنموذج^١.

الجدول رقم (٤)
التوزيعات الاحتمالية البعدية لمعاملات النموذج

المعلمة	الوسط الحسابي	فترة ثقة للوسط الحسابي (بدرجة ثقة ٩٠%)
δ	0.36	-
β	0.99	-
ε	6	-
σ	0.1748	[0.1952-0.1539]
γ	0.9289	[1.1618-0.7248]
ρ_r	0.4510	[0.5118-0.3960]
γ_π	1.7066	[2.3049-1.2717]
γ_y	0.1577	[0.3172-(-)0.0606]
ρ_a	0.4768	[0.5376-0.4412]
ρ_g	0.4855	[0.5743-0.4161]
θ	2.0367	[2.1138-0.19341]

^١ للمزيد من التفاصيل عن "اختبار الصدمات الممهدة"، انظر:

Ibid., p.57.

تابع الجدول رقم (٣)

المعلمة	كود المعلمة في برنامج Dynare	التوزيع	الوسط الحسابي	الانحراف المعياري
θ	Thetabig	توزيع جاما (Gamma Distribution)	٢	١,٤٢
σ_{ε^a}	SE_e_a	توزيع منتظم	0.5	0.2887
σ_{ε^g}	SE_e_g	توزيع منتظم	0.5	0.2887
σ_{ε^m}	SE_e_ms	توزيع منتظم	0.5	0.2887
$\sigma_{\varepsilon^{\lambda}}$	SE_e_lam	توزيع منتظم	0.5	0.2887

٥. نتائج التقدير وتحليلها

يشير كلٌّ من "اختبار الفحص الأحادي" (Univariate Diagnostic Test) و"اختبار الفحص المتعدد" (Multivariate Diagnostic Test) لمعلمات النموذج - انظر الشكل البياني رقم (١) والشكل البياني رقم (٢) بالملحق الإحصائي - إلى تمتع عزوم التوزيعات الاحتمالية البعدية لهذه المعلمات بالاستقرار، وأن قيم هذه العزوم تتقارب في نهاية المراجعات المتكررة (وعدها ٢٠٠٠) لتعطي حل وحيد للنموذج (أي أنه نموذج ذو توازن مستقر أو محدد)^١. كذلك يوضح "اختبار الصدمات المُمهدة" (Smoothed Shocks Test) -- انظر الشكل البياني رقم (٣) بالملحق

^١ يتم إجراء هذه الاختبارات باستخدام Markov Chain Monte Carlo (MCMC) على مستوى ثلاثة من العزوم الخاصة بالتوزيعات الاحتمالية البعدية لمعلمات النموذج؛ وهي: فترة ثقة بدرجة ٨٠% للوسط الحسابي (Interval)، والتباين (العزم الثاني) (m2)، والعزم الثالث (m3). لمزيد من التفاصيل حول هذه الاختبارات، انظر: Ibid., pp. 55-57.

الجدول رقم (٣)
التوزيعات الاحتمالية القبلية لمعاملات النموذج

المعلمة	كود المعلمة في برنامج Dynare	التوزيع	الوسط الحسابي	الانحراف المعياري
δ	Delta	مُعَايرة	0.36	-
β	Beta	مُعَايرة	0.99	-
ε	epsilon	مُعَايرة	6	-
σ	Sig	توزيع معكوس جاما (Inverse-Gamma Distribution)	0.67	0.9
γ	Gam	توزيع طبيعي (Normal Distribution)	1	0.5
ρ_r	Rho	توزيع منتظم (Uniform Distribution)	0.5	0.2887
γ_π	gampie	توزيع طبيعي	1.5	0.5
γ_y	Gamy	توزيع طبيعي	0.125	0.125
ρ_a	Rhoa	توزيع منتظم	0.5	0.2887
ρ_g	Rhog	توزيع منتظم	0.5	0.2887

تابع الجدول رقم (٢)

المتغير	القيمة
$w-p$	0
mrs	0

وتتمثل ثاني خطوات أسلوب المحاكاة في حل النموذج باستخدام البرنامج الإحصائي Dynare 4.3.2^١ والذي تم تشغيله من خلال برنامج Matlab - والحصول على البيانات التركيبية للمتغيرات المُشاهدة الأربعة في النموذج.

٤. المنهج البيزي

بعد الحصول على البيانات التركيبية للمتغيرات المُشاهدة في النموذج، يقوم البرنامج الإحصائي Dynare باستخدام هذه البيانات في إعادة تقدير معلمات النموذج وتتبع أثر الصدمة النقدية على عدد من المتغيرات الاقتصادية الكلية من خلال تطبيق المنهج البيزي. ويقوم المنهج البيزي على تحديد التوزيعات الاحتمالية القبلية لمعلمات النموذج، وتقييم دالة الإمكان باستخدام "تنقية كالمان"، ثم تطبيق "طريقة متروبوليس-هاستينجز" للحصول/لتوصيف التوزيعات الاحتمالية البعدية لتلك المعلمات^٢. ويوضح الجدول رقم (٣) التوزيعات الاحتمالية القبلية لمعلمات النموذج، وقد تم الاعتماد في اختيار هذه التوزيعات الاحتمالية القبلية على كل من: (Liu & Gupta 2007) و (Berg et al 2010) و (Rabanal & Rubio-Ramírez 2003).

^١ Dynare هو برنامج إحصائي تم تطويره لأول مرة بواسطة Michel Juillard في عام ١٩٩٦ ليستخدم في حل ومحاكاة وتقدير نماذج DSGE بطرق يسهل على الباحثين الاقتصاديين تطبيقها. وهو برنامج غير قائم بذاته (Toolbox)؛ أي يعتمد في تشغيله على برامج أخرى، وهي: Matlab و GNU Octave. لمزيد من التفاصيل، انظر:

Ibid., pp. xi-4.

^٢ لمزيد من التفاصيل حول خطوات المنهج البيزي في تقدير معلمات نموذج DSGE، انظر المبحث الثالث من الفصل الأول من:

ولاء محمد محروس: استخدام نماذج التوازن العام العشوائية الديناميكية في تحليل السياسات الاقتصادية في الدول الأفريقية دراسة تطبيقية على السياسة النقدية في تنزانيا، رسالة دكتوراه غير منشورة (القاهرة: معهد البحوث والدراسات الإفريقية، قسم السياسة والاقتصاد، د. ت.).

تابع الجدول رقم (١)

المعلمة	القيمة
σ_{ε^g}	0.5
σ_{ε^m}	0.5
$\sigma_{\varepsilon^{\lambda}}$	0.5

الجدول رقم (٢)
القيم الأولية للمتغيرات عند حالة الثبات

المتغير	القيمة
y	0
Δw	0
r	0
Δp	0
a	0
g	0
n	0
mc	0

الجدول رقم (١)
المعاملات المعيارية

المعلمة	القيمة
δ	0.36
β	0.99
ε	6
σ	0.1624
γ	1
ρ_r	0.5
γ_π	2.003
γ_y	0.5227
ρ_a	0.5
ρ_g	0.5
θ	2
σ_{ε^a}	0.5

حيث $\rho_g, \rho_g \in [0,1]$ ، ويعبر كل من ε_i^a و ε_i^g و ε_i^m و ε_i^l عن الخطأ العشوائي الذي يتبع التوزيع الطبيعي (أي أن $\varepsilon_i^l \sim N(0, \sigma_i^2)$).

٣. أسلوب المحاكاة

تتمثل المتغيرات المُشاهدة لهذا النموذج في أربعة متغيرات؛ هي: الناتج المحلي الإجمالي الاسمي، ومعدل التضخم، والأجور الاسمية، وسعر الفائدة الاسمي. ونظراً لعدم توافر سلاسل البيانات ربع السنوية لثلاثة من هذه المتغيرات - الناتج المحلي الإجمالي الاسمي، ومعدل التضخم، والأجور الاسمية - في حالة الاقتصاد التنزاني، تم اللجوء إلى أسلوب المحاكاة للحصول على بيانات تركيبية للمتغيرات الأربعة. وتتمثل أولى خطوات أسلوب المحاكاة في معايرة معلمات النموذج وتحديد القيم الأولية للمتغيرات عند حالة الثبات (Initial Steady State Values) [انظر الجدول رقم (١) والجدول رقم (٢)]. وقد تم الاعتماد في معايرة هذه المعلمات على كلٍ من: Liu & Gupta (2007)، و Berg et al (2010)، و Ngalawa & Viegi (2013). كما قُلت الباحثة باختبار القيم الصفرية كقيم أولية للمتغيرات عند حالة الثبات^١.

^١ لمزيد من التفاصيل، انظر:

- G. Liu & R. Gupta: "A Small-Scale DSGE Model for Forecasting the South African Economy", *South African Journal of Economics*, Vol. 75, Issue 2, 2007(Oxford: Blackwell Publishing Ltd.), pp. 179-193.
 - Andrew Berg et al: *The Macroeconomics of Medium-Term Aid Scaling-Up Scenarios*, IMF Working Paper, No. 10/60, February 2010, pp. 1-46.
 - Harold P.E. Ngalawa & Nicola Viegi: *Interaction of Formal and Informal Financial Markets in Quasi-Emerging Market Economies*, Economic Research Southern Africa (ERSA) working paper, No. 326 (Cape Town: ERSA, January 2013), pp. 1-25.
- ^٢ ويرجع السبب في اختيار قيم صفرية كقيم أولية للمتغيرات عند حالة الثبات إلى أنه قد تم التعبير عن هذه المتغيرات في صيغة خطية لوغاريتمية. لمزيد من التفاصيل، انظر:

Tommaso Mancini Griffoli: *Dynare V4 - User Guide (Public Beta Version)*, January 2013, pp. 21-24.

(<http://www.dynare.org/documentation-and-support/user-guide/Dynare-UserGuide-WebBeta.pdf>)

$$K_p = \frac{(1-\delta)(1-\theta_p\beta)(1-\theta_p)}{\left[\theta_p\left(1-\delta\left(\frac{\bar{\lambda}}{\bar{\lambda}-1}\right)\right)\right]}$$

حيث $\theta_p = \frac{\theta}{\theta+1}$ ، و $\bar{\varepsilon} = \frac{\bar{\lambda}}{(\bar{\lambda}-1)}$ هي قيمة ε عند حالة الثبات. وتعبّر كل من المعادلة رقم

(١٥) و(١٦) و(١٧) عن معدل الإحلال الحدي بين الاستهلاك وعدد ساعات العمل mrs_t ، والتكلفة الحدية للإنتاج mc_t ، والأجر الحقيقي لعنصر العمل $w_t - p_t$ على الترتيب عند حالة الثبات. وتمثل قاعدة إغلاق النموذج في المعادلة رقم (١٨) التي توضح العوامل المحددة للمسار الزمني لسعر الفائدة r_t (أى المعادلة التي تقوم بتوصيف السياسة النقدية المتبعة)، والتي تُعرف بـ "قاعدة تايلور".

يفترض النموذج أن كلا من الصدمة التكنولوجية a_t والصدمة التي تلحق بتفضيل المستهلك g_t تتحدد وفقاً لمعادلة الانحدار الذاتي من الدرجة الأولى (AR(1)) كالتالي^١:

$$a_t = \rho_a a_{t-1} + \varepsilon_t^a \quad (18)$$

$$g_t = \rho_g a_{t-1} + \varepsilon_t^g \quad (19)$$

كما يفترض النموذج أن كلا من الصدمة النقدية ms_t والصدمة التي تلحق بهامش الربح λ_t تتبع التوزيع الطبيعي كالتالي^٢:

$$ms_t = \varepsilon_t^m \quad (20)$$

$$\lambda_t = \varepsilon_t^\lambda \quad (21)$$

^١ Ibid., p. 11.

^٢ Idem.

حيث تشير $M(h)$ إلى خلق النقود (Money Creation).

■ التوازن الآني للأسواق^١

بإعادة كتابة المعادلات السابقة في صيغة خطية لوغاريتمية (Log-Linear Form) عند حالة الثبات وحلها أنياً^٢، نحصل على المعادلات التالية:

$$y_t = E_t y_{t+1} - \sigma (r_t - E_t \Delta p_{t+1} + E_t g_{t+1} - g_t) \quad (12)$$

$$y_t = a_t + (1 - \delta) n_t \quad (13)$$

$$\Delta p_t = \beta E_t \Delta p_{t+1} + \kappa_p (m c_t + \lambda_t) \quad (14)$$

$$m r s_t = g_t + \frac{1}{\sigma} y_t + \gamma n_t \quad (15)$$

$$m c_t = w_t - p_t + n_t - y_t \quad (16)$$

$$w_t - p_t = w_{t-1} - p_{t-1} + \Delta w_t - \Delta p_t \quad (17)$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r) [\gamma_\pi \Delta p_t + \gamma_y y_t] + m s_t \quad (18)$$

وتعبر المعادلة رقم (١٢) عن "المعادلة الديناميكية لمنحنى IS"، والمعادلة رقم (١٣) عن دالة الإنتاج في صيغة خطية لوغاريتمية، والمعادلة رقم (١٤) عن "المعادلة الديناميكية لمنحنى فيليبس". ويتم الحصول على قيمة κ_p من خلال المعادلة التالية:

^١ Ibid., p. 10.

^٢ حالة الثبات هي حالة التوازن (استقرار الأوضاع) التي تبلغ عندها قيمة معدل التغير في كل من المستوى العام للأسعار والأجور صفراً.

$$P(h_t) = \left[\theta_p P(h_{t-1})^{1-\varepsilon(h_t)} + (1-\theta_p) P^*(h_{t-1})^{1-\varepsilon(h_t)} \right]^{\frac{1}{1-\varepsilon(h_t)}} \quad (9)$$

■ سوق العمل^١

من خلال تعظيم دالة المنفعة الخاصة بكل أسرة داخل القطاع العائلي - في ظل قيد الميزانية المفروض عليها ومع الأخذ في الاعتبار دالة الطلب على عنصر العمل التي يحددها المنتجون (المعادلة رقم ٦) - يتم تحديد الحجم الأمثل لكل من: الاستهلاك، السندات المشروطة، السندات غير المشروطة، عدد ساعات العمل، الرصيد الحقيقي من النقود. ويتمثل شرطي الدرجة الأولى (First-order Conditions) لتعظيم هذه الدالة في المعادلتين التاليتين:

$$G(h_t)C(h_{t,j})^{\frac{1}{\sigma}} = \beta \sum_{h_{t+1}} \pi(h_{t+1} | h_t) \left\{ G(h_{t+1})C(h_{t+1,j})^{\frac{1}{\sigma}} \left[\frac{R(h_t) P(h_t)}{P(h_{t+1})} \right] \right\} \quad (10)$$

$$G(h_t)C(h_{t,j})^{\frac{1}{\sigma}} \frac{W(h_t, j)}{P(h_t)} = \vartheta N(h_t, j)^{\gamma} \quad (11)$$

حيث β هي عامل الخصم، و $\pi(h_{t+1} | h_t) = \frac{\pi(h_{t+1})}{\pi(h_t)}$ هو الاحتمال الشرطي لوقوع الحدث h_{t+1} مع العلم بالحدث h_t ، و $\vartheta = \frac{\phi}{\phi - 1}$ إلى الفرق بين الأجر الحقيقي ومعدل الإحلال الحدي بين عدد ساعات العمل والاستهلاك (Marginal Rate of Substitution between Consumption and Labor).

■ الحكومة^٢

لا تسعى الحكومة إلى إحداث فائض أو عجز في الموازنة العامة، ومن يتمثل القيد على الموازنة العامة في المعادلة التالية:

$$\int_0^1 T(h_t, j) dj = M(h_t) - M(h_{t-1})$$

^١ Ibid., p. 7.

^٢ Idem.

$$N(h_i, i, j) = \left[\frac{W(h_i, j)}{W(h_i)} \right]^{-\phi} \left[\frac{Y(h_i, i)}{A(h_i)} \right]^{\frac{1}{1-\delta}} \quad \forall j \quad (6)$$

حيث تمثل $W(h_i)$ المؤشر عام للأجور/مؤشر عام لإجمالي الأجور/إجمالي الأجور (Aggregate Wage Index)، والذي يمكن الوصول إليه من خلال المعادلة التالية:

$$W(h_i) = \left[\int_0^1 W(h_i, j)^{1-\phi} dj \right]^{\frac{1}{1-\phi}} \quad (7)$$

المرحلة الثانية: في هذه المرحلة، يسمى منتج السلع الوسيطة إلى تحديد أسعار سلعهم بالشكل الذي يعظم أرباحهم. ولكن يمكن لبعض المنتجين فقط، والتي تصل نسبتهم إلى حوالى $(1-\theta_p)$ ، إعادة تحديد أسعار السلع التي ينتجونها في أى فترة زمنية. بينما يظل البعض الآخر، والتي تصل نسبتهم إلى حوالى θ_p ، غير قادر على تغيير أسعار السلع التي ينتجها (أى تظل أسعار السلع التي يتم إنتاجها ثابتة دون تغيير)؛ فيما يُعرف بـ "قيد كالفو" (Calvo-Type Restriction). ومن ثم، يسعى حوالى $(1-\theta_p)$ من المنتجين إلى معرفة/تحديد السعر الأمثل $P^*(h_i, i)$ الذي يعظم القيمة السوقية الحالية لأرباحهم، ويمكن التعبير عن ذلك كالتالي:

$$\max \sum_{\tau=1}^{\infty} \sum_{h_{i+\tau}|h_i} \theta_p^{\tau} S(h_{i+\tau}|h_i) \left\{ \left[\frac{P^*(h_i, i)}{P(h_{i+\tau})} - \Lambda(h_i) \bar{MC}(h_{i+\tau}, i) \right] \bar{Y}(h_{i+\tau}, i) \right\} = 0 \quad (8)$$

حيث تشير $S(h_{i+\tau}|h_i)$ إلى عامل الخصم العشوائي للأرباح/العائد الاسمي (Stochastic Discount Factor for Nominal Payoffs)

$$\bar{MC}(h_{i+\tau}, i) = \frac{W(h_{i+\tau}) \left[\frac{\bar{Y}(h_{i+\tau}, i)}{A(h_i)} \right]^{\frac{\delta}{1-\delta}}}{(1-\delta)P(h_{i+\tau})} \quad \text{وتشير}$$

الأمثل الذي يعظم أرباح منتجي السلع النهائية، يمكن تطبيق المعادلة التالية:

عنصر رأس المال فى الأجل القصير، ويعبر عن هذا الفرض بـ $\bar{K}$ ؛ حيث تمثل δ نصيب رأس المال من الناتج (Capital Share of Output).

بينما تخضع السلع النهائية - والتي تعتمد فى إنتاجها على السلع الوسيطة- لدالة الإنتاج التالية:

$$Y(h_t) = \left[\int_0^1 Y(h_t, i)^{\frac{\varepsilon(h_t)-1}{\varepsilon(h_t)}} di \right]^{\frac{\varepsilon(h_t)}{\varepsilon(h_t)-1}} \quad (4)$$

حيث أن $\varepsilon(h_t)$ هى مرونة الإحلال بين السلع الوسيطة المختلفة، و $\Lambda(h_t) = \frac{\varepsilon(h_t)}{\varepsilon(h_t)-1}$ هو هامش الربح الذى يحققه منتج السلع النهائية (وهو متغير عبر الزمن)^١. ويسعى منتج السلع النهائية إلى تعظيم أرباحهم فى ظل قيد دالة الإنتاج (المعادلة رقم (٤))، ويتمثل السعر الذى يقوم بتعظيم هذه الأرباح فيما يلي:

$$P(h_t) = \left[\int_0^1 P(h_t, i)^{1-\varepsilon(h_t)} di \right]^{\frac{1}{1-\varepsilon(h_t)}} \quad (5)$$

وكما سبق الذكر يعمل منتج السلع الوسيطة فى ظل شروط المنافسة الاحتكارية. ويسعى هؤلاء المنتجون إلى تعظيم أرباحهم من خلال المرور بمرحلتين:

المرحلة الأولى: فى هذه المرحلة، يسعى المنتجون- فى ظل تلقيهم لأجور عنصر العمل/العمال (Given All Wages)- إلى اختيار التوليفة المثلى من عنصر العمل (Optimal Labor Mix) التى تساعد فى تعظيم الأرباح. وتتمثل دالة الطلب لكل منتج i على عنصر العمل z فى المعادلة التالية:

^١ وكما سبق الذكر، تسيطر المنافسة الكاملة على سوق السلع النهائية؛ أى أن منتجي السلع النهائية متلقون للأسعار سواء لأسعار السلع الوسيطة $P(h_t, i)$ أو لأسعار السلع النهائية $P(h_t)$. وتتمثل دالة الطلب على السلع الوسيطة (مدخلات الإنتاج) من قبل منتجي السلع النهائية فيما يلي:

$$Y(h_t, i) = \left[\frac{P(h_t, i)}{P(h_t)} \right] Y(h_t) \quad \forall i$$

$$P(h_t)C(h_{t,j}) + M(h_{t,j}) - M(h_{t-1,j}) + \sum_{T=1}^{\infty} \sum_{h_{t+T}|h_t} Q(h_{t+T}|h_t) D(h_{t+T}, j) + \frac{B(t+1, j)}{R(h_t)} \\ = W(h_t, j)N(h_t, j) + \Pi(h_t, j) + T(h_t, j) + D(h_t, j) + B(t, j) \quad (2)$$

حيث أن $D(h_{t+T}, j)$ هي سندات مشروطة (Contingent Bonds) يمتلكها القطاع العائلي و $Q(h_{t+T}|h_t)$ هي سعر هذه السندات^١، و $B(t+1, j)$ هي سندات غير مشروطة يحتفظ بها القطاع العائلي (Uncontingent Bond) و $\frac{1}{R(h_t)}$ هي سعر هذه السندات (وهو معكوس/مقلوب

(Inverse) سعر الفائدة الاسمي الإجمالي (Gross Nominal Interest Rate)، و $W(h_t, j)$ الأجر الاسمي في الساعة، و $\Pi(h_t, j)$ هي أرباح المنشآت التي يملكها القطاع العائلي (تملكها الأسر)، و $T(h_t, j)$ هي التحويلات/الإعانات الحكومية الاسمية (Nominal Transfers) للقطاع العائلي أو إجمالي الضرائب المدفوعة للحكومة (Lump-sum Taxes).

■ المنتجون (سوقي السلع الوسيطة والنهائية)^٢

تخضع السلع الوسيطة لدالة الإنتاج التالية:

$$Y(h_t, i) = A(h_t) K^{\delta} \left\{ \left[\int_0^1 N(h_t, i, j)^{\frac{\phi-1}{\phi}} dj \right]^{\frac{\phi}{\phi-1}} \right\}^{1-\delta} \quad (3)$$

حيث أن $A(h_t)$ هي الصدمة التي تلحق بالمستوى التكنولوجي المستخدم، و $N(h_t, i, j)$ هي عدد ساعات العمل للعامل j الذي يقوم بتأجيره المنتج i ، و ϕ هي مرونة الإحلال بين الأنواع غير المتجانسة لعنصر العمل المختلفة (Different Types of Labor). كما يفترض النموذج ثبات

^١ ويُقصد بالسندات المشروطة تلك السندات التي يرتبط حصول حاملها على سعر الفائدة عليها بوقوع شرط/حدث معين؛ وفي حالة النموذج محل الدراسة يحصل حاملها هذه السندات على سعر الفائدة عليها (والذي يُقدر بدولار واحد) في اللحظة $t + \tau$ إذا ما وقع الحدث $h_{t+\tau}$ ، بينما لا يتم الحصول على ذات سعر الفائدة في حالة عدم وقوع هذا الحدث.

^٢ Ibid., pp. 5-7.

٤. عدم حياد دور السياسة النقدية في الاقتصاد الكلي.

وفيما يلي شرح مفصل للمعادلات الرياضية ولقرارات الأمثلية التي يتخذها المنتجين والأفراد (المستهلكين/القطاعات العائلية) في هذا النموذج.

■ القطاع العائلي^١

في كل فترة زمنية s_t ، يتخذ الفاعلون الاقتصاديون في النموذج مجموعة من القرارات.

ومن ثم، يشير $h_t = \{s_0, s_1, s_2, \dots, s_t\}$ إلى سلسلة من الفترات الزمنية التي تبدأ بـ s_0 وتنتهي بـ s_t . وتسعى كل أسرة - والتي يتم الرمز إليها بـ j - إلى تعظيم منفعتها المتوقعة من كل من الاستهلاك، وعدد ساعات العمل، والرصيد من النقود الحقيقية (Real Money Balance Holdings) منذ الوقت الحاضر إلى ما لا نهاية. وتأخذ دالة المنفعة (دالة الهدف Objective Function) لكل أسرة الشكل التالي:

$$\max \sum_{t=0}^{\infty} \sum_{h_t} \beta^t \pi(h_t) \left[\frac{G(h_t) C(h_{t,j})^{\frac{1}{1-\sigma}}}{1 - \frac{1}{\sigma}} + \frac{v}{1-\xi} \frac{M(h_{t,j})^{1-\xi}}{P(h_t)} - \frac{N(h_{t,j})^{1-\gamma}}{1+\gamma} \right] \quad (1)$$

حيث أن $C(h_{t,j})$ هي الكمية المستهلكة من السلع النهائية المنتجة، و $\frac{M(h_{t,j})}{P(h_t)}$ هي الرصيد من

النقود الحقيقية، و $N(h_{t,j})$ هي عدد ساعات العمل للأسرة j ، و $P(h_t)$ هي سعر السلعة النهائية،

و β هي عامل الخصم، و σ هي مرونة الإحلال فيما بين الفترات الزمنية (Elasticity of Intertemporal Substitution)، و ξ هي مرونة الاحتفاظ بالنقود (Elasticity of Money

Holdings)، و $\frac{1}{\gamma}$ هي مرونة عرض العمل للأجر الحقيقي، و v هي مقياس للأهمية التي يمثلها

الرصيد من النقود الحقيقية في دالة المنفعة، و $G(h_t)$ هي الصدمة التي تلحق بتفضيل المستهلك

(Preference-Shifter Shock) وتؤدي إلى تغيير قراراته فيما يتعلق باستهلاك السلع النهائية،

ومن ثم تؤثر على المنفعة الحدية للاستهلاك. ويتمثل قيد الميزانية المفروض على دالة المنفعة فيما يلي:

^١ Ibid., p. 4.

أداء السياسة النقدية في تنزانيا؛ باعتبار أن هذا المجال أصبح من أكثر المجالات استخداماً لهذه النماذج في ضوء اعتماد البنوك المركزية في الدول المتقدمة وعدد من الدول النامية على هذه النماذج في هذا الغرض. وبناء عليه، يتم وضع توصيفاً كاملاً لهذا النموذج، ثم عرض خطوات كل من أسلوب المحاكاة الذي يتم استخدامه للحصول على بيانات تركيبية لمتغيرات النموذج، والمنهج البيزي الذي يتم استخدامه في عملية التقدير. وأخيراً، يتم تحليل النتائج وتتبع مسار آثار السياسة النقدية على عدد من المتغيرات الاقتصادية الكلية في تنزانيا.

٢. توصيف النموذج

يتكون النموذج من أربعة فاعلين اقتصاديين؛ القطاع العائلي الذي يحتوي على عدد لا نهائي من الأسر (ويقوم الأفراد داخل كل أسرة بالمشاركة في سوق العمل)، والمنتجين/المنشآت الصناعية (عدد لا نهائي من منتجي السلع الوسيطة والنهائية)، والحكومة، وأخيراً البنك المركزي المسئول عن رسم السياسة النقدية. ويخضع سوقي السلع الوسيطة والعمل في هذا النموذج إلى شروط المنافسة الاحتكارية، بينما يخضع سوق السلع النهائية إلى شروط المنافسة الكاملة. ويواجه منتجو السلع الوسيطة قيوداً في تحديد الأسعار الخاصة بسلعهم (Restrictions in Price-setting Process). كما يحتوي النموذج على أربع صدمات خارجية؛ وهي: صدمة تكنولوجية، وصدمة تلحق بتفضيل/اختيار المستهلك (Preference Shock)، وصدمة تلحق بهامش الربح، وأخيراً صدمة نقدية. ومن ثم، يمكن إيجاز الفروض التي يعتمد عليها النموذج في النقاط الأربع التالية:

١. اقتصاد مغلق بدون استثمار خاص وهناك دور محايد للإنفاق الحكومي (السياسة المالية)؛ أي الناتج الكلي يتم استهلاكه.

٢. سيطرة المنافسة الاحتكارية على سوقي السلع الوسيطة والعمل؛ أي أن هناك تنوع في السلع الوسيطة (Differentiated Goods)/السلع المنتجة غير متماثلة أو غير متجانسة، ومن ثم يتم تحديد أسعار السلع من قبل المنتجين/يستطيع المنتجون التحكم في أسعار السلع المنتجة (أي أنهم غير متلقين للأسعار)، كما أن هناك عدم تجانس/عدم تماثل في عنصر العمل، ومن ثم يتم تحديد الأجور من قبل الأسر/يستطيع الأفراد العاملون داخل الأسر التحكم في تحديد الأجور التي يحصلون عليها (أي أنهم غير متلقين للأجور). بينما تسيطر شروط المنافسة الكاملة على سوق السلع النهائية.

٣. سيطرة الجمود على هيكل أسعار السلع الوسيطة/أو أن هناك قيود مفروضة على "آلية مواءمة الأسعار" (Price Adjustment Mechanism)؛ ويتم ذلك من خلال افتراض أن بعض المنتجين فقط هم القادرون على إعادة تعيين/تحديد أسعار هذه السلع في كل فترة (In Any Given Period).

^٤ ويرجع السبب وراء اختيار هذا النموذج إلى سهولة تطبيقه (لإمكانية تطبيقه)، وملامته لحالة تنزانيا، وتوافر الأكواد الخاصة بتطبيقه على البرنامج الإحصائي Dynare. لمزيد من التفاصيل، انظر:

Pau Rabanal & Juan F. Rubio-Ramírez: **Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach**, Federal Reserve Bank of Atlanta Working Paper, No. 2001-22a (Georgia, GA: Federal Reserve Bank of Atlanta, February 2003), pp. 1-39.

² Ibid., pp. 3-4.

يتخذونها عبر الزمن، وتقييم تأثيرها على المتغيرات الاقتصادية الكلية. كما تساعد هذه النماذج على تحليل أثر الصدمات العشوائية على الاقتصاد الكلي؛ حيث تقدم هذه النماذج توصيفاً دقيقاً للقنوات التي تنتقل من خلالها تلك الصدمات مما يمكن من تتبع وتقييم آثارها المباشرة وغير المباشرة على المتغيرات الاقتصادية الكلية^١. كذلك تؤكد العديد من الدراسات التطبيقية الحديثة على أن التوقعات التي يمكن الحصول عليها باستخدام نماذج DSGE تتمتع بانخفاض الجذر التربيعي لمتوسط الخطأ المربع (Root Mean Squared Error) مقارنة بتلك التي يمكن الحصول عليها من خلال نماذج "متجه الانحدار الذاتي" (Vector Auto Regressions (VARs)) ونماذج "متجه الانحدار الذاتي البيزي" (Bayesian Vector Auto Regressions (BVARs))، شريطة أن يحتوي النموذج على عدد كاف من الصدمات وأن تكون معلمات النموذج محددة بشكل جيد (Well Identified)^٢.

وتوجد العديد من الطرق القياسية المتعارف عليها والتي يتم استخدامها من قبل الباحثين لتعيين قيم لمعاملات (Parameterize) نماذج DSGE، وتقييم تلك النماذج (Evaluate DSGE Models). وتتضمن هذه الطرق: "المعايرة"، و"طريقة العزوم المعممة" (Generalized Method of Moments (GMM))، و"دالة الاستجابة للصدمات" (Impulse Response Function (IRF))، و"طريقة الإمكان الأكبر" (Maximum Likelihood Method)، وأخيراً "المنهج البيزي". ويتزايد الاعتماد في الفترة الأخيرة على المنهج البيزي في تقدير نماذج DSGE نظراً لنجاحه في التغلب على مشاكل التمييز (Identification Problems) المتعلقة بنماذج DSGE والتي لا تستطيع طرق الاقتصاد القياسي الكلاسيكي معالجتها، كما يمكنه التغلب على المشاكل المتعلقة بتوصيف نماذج DSGE لأنه يفترض عدم وجود نموذج صحيح/موصف توصيفاً صحيحاً (All Models are False). كما يساعد هذا المنهج من خلال استخدام التوزيعات الاحتمالية القبلية للمعاملات - والتي يعتمد اختيارها على النظرية الاقتصادية والدراسات الاقتصادية التطبيقية - على إدراج معلومات إضافية في عملية تقدير المعلمات. كذلك يقدم هذا المنهج مجموعة واسعة من الأدوات التي تمكن الباحثين من تقدير وتقييم نماذج DSGE - خاصة النماذج متوسطة وكبيرة الحجم (Evaluate Medium- to Large- Scale Models)^٣.

هذا، وتسعى هذه الورقة البحثية إلى استخدام نموذج توازن عام عشوائي ديناميكي لاقتصاد صغير مغلق تم تصميمه من قبل Rabanal و Rubio-Ramírez في عام ٢٠٠٣ صغير مغلق لتحليل

^١ Shanaka J. Peiris & Magnus Saxegaard: **An Estimated DSGE Model for Monetary Policy Analysis in Low-Income Countries**, International Monetary Fund (IMF) Working Paper, No. 07/282 (Washington, D. C.: IMF, December 2007), p. 5.

^٢ للمزيد من التفاصيل، انظر:

Rochelle M. Edge et al: "A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model", **Journal of Applied Econometrics**, Vol. 25, Issue 4 (New York, NY: John Wiley & Sons Ltd., 2010), pp. 720-754.

^٣ S. An & F. Schorfheide: "Bayesian Analysis of DSGE Models", **Econometric Reviews**, Vol. 26, Issue 2-4, May 2007 (London: Taylor & Francis Group), p. 123.

^٤ Pablo A. Guerrón-Quintana & James M. Nason: **Bayesian Estimation of DSGE Models**, Federal Reserve Bank of Philadelphia Working Paper, No. 12-4 (Pennsylvania, PA: Federal Reserve Bank of Philadelphia, February 2012), p. 1.

استخدام نماذج التوازن العام العشوائية الديناميكية في تحليل السياسة النقدية

مع التطبيق على تنزانيا

عراقي عبد العزيز الشربيني^١، ولاء محمد محروس^٢

ملخص البحث

مع التوسع في تطبيق نماذج التوازن العام العشوائية الديناميكية لتحليل السياسات الاقتصادية الكلية، والدورات الاقتصادية، والتنبؤ بالأداء الاقتصادي في الدول المتقدمة والنامية على حد سواء، تسعى هذه الورقة البحثية إلى استخدام نموذج توازن عام عشوائي ديناميكي لاقتصاد صغير مغلق لتحليل أداء السياسة النقدية في تنزانيا. ويتم استخدام المنهج البيزي في عملية تقدير وتتبع مسار آثار السياسة النقدية على عدد من المتغيرات الاقتصادية الكلية لتنزانيا. هذا، وتؤكد النتائج على التزام "بنك تنزانيا" باتباع سياسة نقدية لا يشوبها تغيرات جذرية، هذا بالإضافة إلى نجاحه في استهداف معدل التضخم. كما تشير النتائج إلى الدور الفعال الذي تلعبه السياسة النقدية في الاقتصاد التنزاني؛ حيث يتأثر كل من الناتج ومعدل التضخم بالصدمات النقدية بدرجة كبيرة.

١. مقدمة

نجحت نماذج التوازن العام العشوائية الديناميكية (Dynamic Stochastic General Equilibrium (DSGE) Models أن تحل محل نماذج الاقتصاد القياسي الكلية في عملية تحليل الدورات الاقتصادية، والتنبؤ بالأداء الاقتصادي، ورسم السياسات الاقتصادية خلال العقدين المنصرمين. فقد بدأت العديد من البنوك المركزية في الدول المتقدمة (بنك كندا، وبنك إنجلترا، والبنك المركزي الأوروبي، وبنك الاحتياطي الفيدرالي الأمريكي)، والدول النامية (البنك المركزي في شيلي، والبنك المركزي في بيرو، والبنك المركزي في دولة جنوب أفريقيا، والبنك المركزي في نيجيريا) في استخدام تلك النماذج لتحليل ورسم السياسة النقدية^٣. كما قام صندوق النقد الدولي بتصميم "نموذج الاقتصاد الشامل"، و"النموذج الشامل لتكامل السياسة المالية والنقدية"، و"نموذج السياسة المالية الشامل" لتحليل ورسم السياستين النقدية والمالية على مستوى دول العالم^٤.

ويمكن إرجاع هذا النجاح والانتشار لنماذج DSGE إلى قدرتها على تقديم توصيفاً مفصلاً لهيكل الاقتصاد الكلي ودوافع وسلوك الفاعلين الاقتصاديين، وتتبع مسار العديد من القرارات التي

^١ أستاذ الاقتصاد المساعد بقسم السياسة والاقتصاد، معهد البحوث والدراسات الإفريقية - جامعة القاهرة.

^٢ مدرس مساعد بقسم السياسة والاقتصاد، معهد البحوث والدراسات الإفريقية - جامعة القاهرة. ويود الباحثان التوجه بالشكر والتقدير إلى كل من الدكتور/ Jonathan Benchimol المحاضر بجامعة Sorbonne، والدكتور/ Johannes Pfeifer المحاضر بجامعة Mannheim على ما قدماه من إرشادات سديدة وتوجيهات عديدة أثناء عملية تشغيل النموذج وتحليل النتائج الخاصة به، فلهما جزيل الشكر والتقدير.

^٣ C. E. Tovar: **DSGE Models and Central Banks**, Bank for International Settlements (BIS) Working Paper, No. 258, September 2008 (Basel: BIS), p. 1.

^٤ Jana Kremer et al: "Dynamic Stochastic General Equilibrium Models as a Tool for Policy Analysis ", **CESifo Economic Studies**, Vol. 52, No. 4/2006, November 2006 (Oxford: Oxford University Press), p. 641.

فهرس الإحصاء التطبيقى

٣٣-١	إستخدام نماذج التوازن العام العشوائية الديناميكية فى تحليل السياسة النقدية مع التطبيق على تنزانيا عراقى عبد العزيز الشربيني ، ولاء محمد محروس	١
٥٥-٣٤	نموذج إحصائى للتنبؤ بعدد الشكاوى الواردة بالهيئة القومية للتأمينات الإجتماعية أحمد أمين الشيخ ، محمد سيد أحمد	٢
٧٤- ٥٦	خدمات الرعاية الصحية وتقدير القيم الإتجاهية لمؤشرات تكاليفها بإستخدام الأساليب الكمية كمدخل للتأمين الصحى: دراسة تطبيقية على المنطقة الشرقية بالمملكة العربية السعودية أسامة حنفى محمود	٣
٨٨ - ٧٥	قياس الذكاء الرقمى للأجهزة الإحصائية الرسمية بالتطبيق على الدول العربية أمل خيرى أمين	٤
١٠٥ - ٨٩	ورقة بحثية عن تدفق العمالة فى سوق العمل المصرى بين الفترتين ٢٠١٠ - ٢٠١١ (دراسة مقارنة) أحمد ماهر أمين ، نجلاء عبد المنعم	٥

1. The first part of the paper is a review of the literature on the topic of the paper.

2. The second part of the paper is a description of the methodology used in the study.

3. The third part of the paper is a discussion of the results of the study.

4. The fourth part of the paper is a conclusion.

5.

6.

7.

8.



جامعة القاهرة
معهد الدراسات والبحوث الإحصائية



المؤتمر السنوى الثامن والأربعون للإحصاء
وعلوم الحاسب وبحوث العمليات

مجلد
الإحصاء التطبيقى

٢٣-٢٦ ديسمبر ٢٠١٣