

جامعة القاهرة

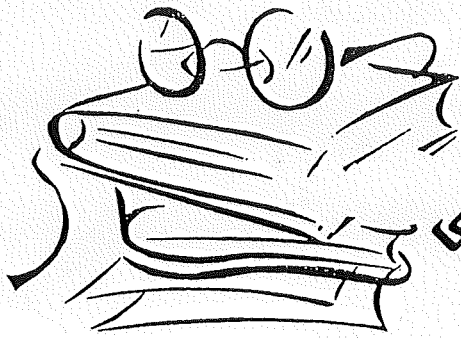


معهد الدراسات والبحوث الإحصائية

المؤتمر السنوى الرابع والأربعون

للإحصاء وعلوم الحاسب

وبحوث العمليات

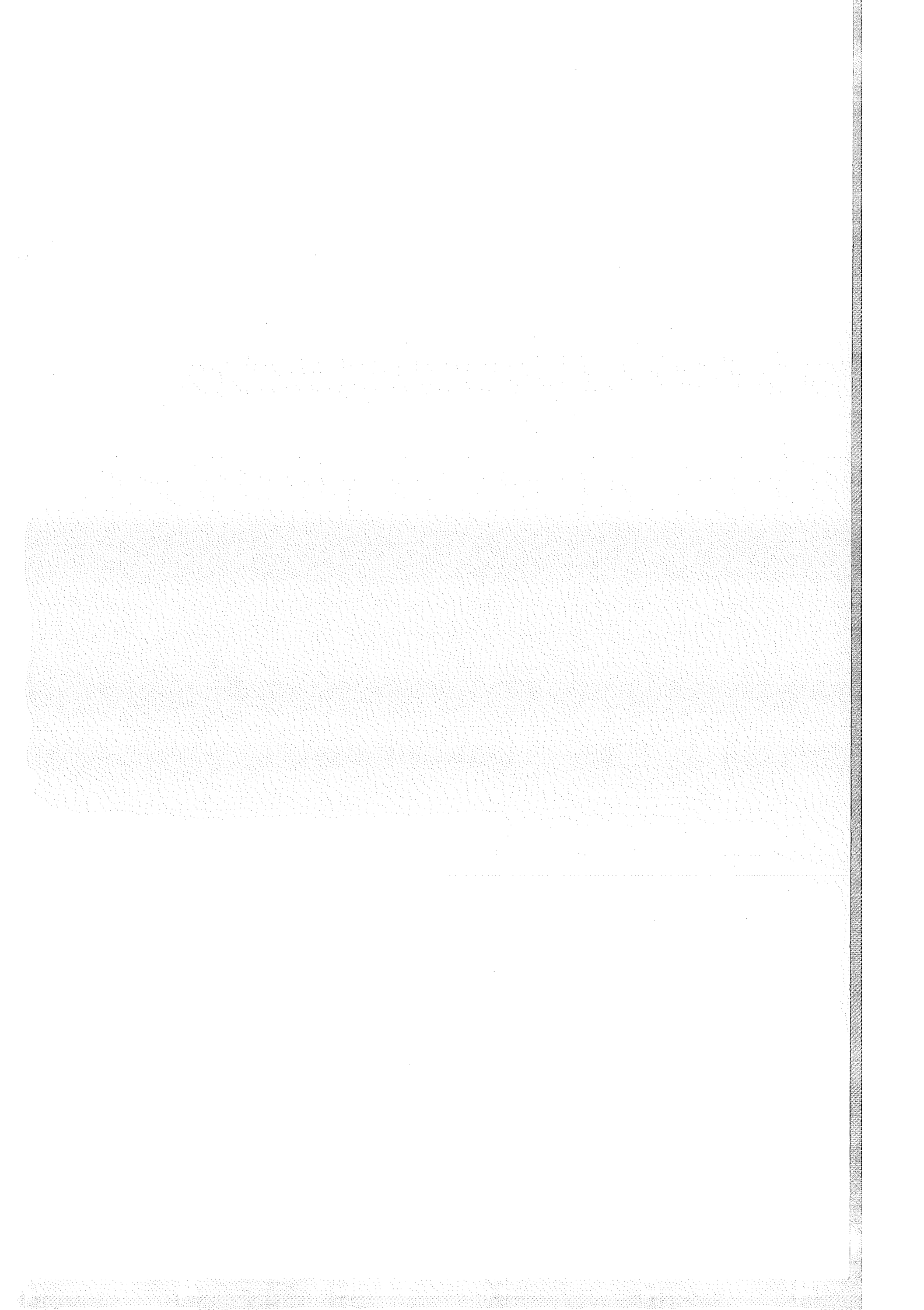


مجلد

الأحصاء السكانى



٧-١٠ ديسمبر ٢٠٠٩



فهرس الإحصاء السكاني

٤-١	كلمة العميد	١
٨-٥	أمانة المؤتمر	٢
٢٣-٩	تقييم وتنقيح بيانات التعداد مع التطبيق على محافظة القليوبية بجمهورية مصر العربية لعام ٢٠٠٦ م عبد الغنى محمد عبد الغنى ، سحر رأفت ، عبد الله محمود سرج	٣
٤٦-٢٤	التعمر الديموجرافى فى مصر خلال الفترة ١٩٧٤ - ٢٠٠٦ والتوقعات المستقبلية عبد الغنى محمد عبد الغنى ، مدحت مصطفى أحمد	٤
٧٢-٤٧	الاحتياجات الحالية والمستقبلية من الوحدات السكنية فى مصر خلال الفترة من (٢٠٠٧-٢٠٢٢) جلاء حلمى سليم	٥
٩٨-٧٣	السكان والغذاء فى مصر فيليب فهمى فانوس	٦

كلمة

المؤتمر السنوي الرابع والأربعون "للإحصاء وعلوم الحاسب وبحوث العمليات" ٧ - ١٠ ديسمبر ٢٠٠٩

أساتذتي وزملائي الاعزاء
السيدات والسادة الحضور الكريم
لقد أصبح مؤتمرنا السنوي "للإحصاء وعلوم الحاسب وبحوث العمليات" أحد
السمات المميزة لأنشطة معهد الدراسات والبحوث الإحصائية ، كما أصبح محفلاً
علمياً رهيناً بما يتضمنه من جلسات علمية متخصصة فضلاً عن الندوات العلمية
من مختلف المجالات ذات الصلة ، كما أنه يمثل فرصة سنوية للتواصل بين
معهدنا ومختلف الجامعات والمراكز البحثية والوزارات والهيئات والجمعيات
والقطاع الخاص ، ويبرز من خلال تلك الفرصة السنوية دور معهد الإحصاء في
المساهمة في المشاركة في حل القضايا المجتمعية.

أساتذتي وزملائي الاعزاء
السيدات والسادة الحضور الكريم
تعقد فعاليات أنشطة مؤتمرنا السنوي هذا العام تحت شعار:

"نحو إحصاء لتحقيق التنمية الإنسانية"

فقد خلصت خبرات السنوات الماضية عن حقيقة لاحتاج إلى برهان مفادها
ان أي خطه أو برنامج أو مشروع لا يمكن انجازه في غياب قواعد البيانات
وبالتالى الإحصاء كأداة لاتخاذ القرارات. وفي هذا السياق فان اعتزازنا نحو تبني
هذا الشعار. فالتنمية الإنسانية طبقاً لمختلف أبعادها تعتبر وسيلة وغاية لأي جهد
تنموي من شتى المجالات الاقتصادية والاجتماعية والثقافية والبيئية. حيث أن
التغيرات السريعة والمتلاحقة التي طرأت على العالم خلال العقد الأول من الألفية
الثالثة تضيف تحديات جديدة في سبيل تحقيق التنمية الإنسانية كما أنها تتيح -

فى نفس الوقت فرصاً لتحقيق هذا الهدف . ومن أهم المتغيرات العالمية التى طرأت هى الإنتقال من المجتمع الصناعي إلى مجتمع المعرفة .
ولقد أصبح من المسلمات أن المصدر الرئيسى لإنتاج المعرفة يتمثل فى التعليم والبحث العلمى والتكنولوجيا ، وإن تطوير التعليم وتشجيع البحوث هو السبيل الرئيسى لتحقيق مجتمع المعرفة .
ولا يغيب عنا جميعاً الدور الهام للإحصاء فى حياة البشرية كأداة للمساعدة فى إتخاذ القرارات ورسم وتبنى السياسات الناجحة على كافة المستويات .

أساتذتى وزملاي الأعزاء
السيدات والسادة الحضور الكريم
إن دور الجامعات ومؤسسات البحث العلمى فى حل مشكلات المجتمع هو محور رئيسى للمشاركة المجتمعية ، وإن لمعهد الإحصاء بجامعة القاهرة دوراً واضحاً فى التفاعل مع البيئة المحيطة وخدمة المجتمع وذلك من خلال عدة أنشطة أذكر منها على سبيل المثال لا الحصر دور المعهد فى المشاركة فى التسجيل الالكترونى لمكتب التنسيق للعام الدراسى الحالى ، وكذلك القوافل التى تم تنظيمها بالتعاون مع مؤسسة النور الخيرية (مغربى) .

السيدات والسادة الحضور الكريم
لقد واکب مؤتمرنّا فى العام الماضى إدراج معهد الإحصاء بجامعة القاهرة فى مشروع الجودة للتعليم ، حيث خطى المعهد خطوات سوف تساعد فى التأهيل للتقدم لمشروع CIQAP والذى نرى فيه أملاً كبيراً حتى أن يحقق معهدنا رؤيته ورسالته فى إقصر وقت ممكن ، وأن ريادة المعهد التى تتمثل فى إنفراده لمنح درجة الدبلوم والماجستير والدكتوراه فى الإحصاء وعلوم الحاسب وبحوث العمليات تملى علينا أهمية مواكبة العصر من تكنولوجيا المعلومات وطرق التعليم والتعلم والتى تضيف الى علوم الإحصاء وجه جديد بحيث تعمل على تحقيق مجتمع المعرفة . وفى هذا الإطار فإن المعهد ينظم سلسلة من حلقات العمل ضمن أنشطة وحدة ضمان الجودة لأعضاء هيئة التدريس وسوف تعقد الحلقة الاولى يوم الثلاثاء ٢٢/١٢/٢٠٠٩ تحت عنوان :

"صياغة مخرجات التعلم وتحديد أدوات التقييم وتحقيق الترابط
بين المقرر والبرنامج الدراسي والخطة الدراسية"
وسيقوم أ.د. أحمد سيف النصر الأستاذ غير المتفرغ بقسم الإحصاء السكاني بالمعهد
ونائب رئيس جامعة دبي للشؤون الأكاديمية (سابقاً) بالإعداد وأدارة الحلقة.

أساتذتي وزملائي الأعزاء
الحضور الكريم

تتعدد أنشطة مؤتمرنا السنوي هذا العام بحيث يشتمل على (٧) جلسات
علمية تحتوي على مايقرب (٥٠) بحثاً في الإحصاء السكاني والتطبيقي والرياضي
وعلوم الحاسب والمعلومات وبحوث العمليات.

كما أن معهدنا كعهدة دائماً يتواصل مع الجامعات والمراكز البحثية
والوزارات والهيئات والجمعيات من خلال عقد ندوة علمية مركز علمي أذكر
منها:

- ندوة : "ماذا بعد مؤتمر القاهرة والسكان (خمس عشرة عاماً)" بالتعاون
مع المكتب الاقليمي لصندوق الامم المتحدة للسكان
- ندوة: "دور المشروع العربي لصحة الأسرة بجامعة الدول العربية فى
التخطيط وإتخاذ القرارات"
بالتعاون مع المشروع العربي لصحة الأسرة - جامعة الدول العربية.
- ندوة: "رؤية اقتصادية للتنمية الزراعية".
بالتعاون مع مركز البحوث الزراعية - معهد بحوث الاقتصاد الزراعي -
وزارة الزراعة
- ندوة: "السكان والتنمية"
بالتعاون مع الاتحاد العربي لحماية البيئة - الامانة العامة
- ندوة "قياس وتحسين الإنتاجية"
بالتعاون مع مصلحة الكفاية الإنتاجية والتدريب - وزارة التجارة
والصناعة

- ندوة:

**"Statistics for Good Governances: the Ibrahim Index
of African Governance" (Dr. Ali Hadi)**

بالتعاون مع مركز معلومات ودعم إتخاذ القرار - مجلس الوزراء

ويتواصل المؤتمر بمحاضرتين عامتين من خيرة علماء مصر بالخارج بل سفرائها
من خلال مشروع نقل المعرفة والخبرة اكااديمية البحث العلمى والتي نتشرف بهما
وهما:

• الاستاذ الدكتور سمير فريد للحديث عن:

"New Studying the Dynamics of International Migration"

• الاستاذ الدكتور عادل المغربي للحديث عن:

"Information Assurance: Challenges Research Directions"

السادة الحضور الكريم:

لايسعنى فى نهاية كلمتى إلا أن أتقدم بخالص الشكر والامتنان لسعادة الأستاذ
الدكتور/ حسام محمد كامل رئيس جامعة القاهرة علي كريم رعايته لمؤتمرنا ،
والذى لا يألوا جهداً في دعم جميع أنشطة المعهد.

كما أتقدم بالشكر والامتنان لجميع الجهات والهيئات الداعمة لمؤتمرنا هذا
العام ، وجميع الجهات والهيئات المشاركة فيه ، كما أتقدم بالشكر للجنة المنظمة
على الجهد المبذول في تنظيم المؤتمر والشكر موصول لحضراتكم جميعاً علي
الحضور والمشاركة . واتمنى أن تخرج جميع أنشطة مؤتمرنا بتوصيات بناءه
تساعد متخذى القرار على جميع المستويات.

وأدعو الله العزيز أن يوفقنا جميعاً لما فيه الخير لمصرنا العزيزة.

والسلام عليكم ورحمه الله وبركاته ،،،

عميد المعهد

أ.د. محمد نجيب عبد الفتاح

أمانة المؤتمر والهيئة العلمية

رعاية المؤتمر

تحت رعاية السيد الاستاذ الدكتور/ رئيس الجامعة أ.د. حسام كامل
رئاسة المؤتمر

رئيس المؤتمر
وعميد معهد الدراسات والبحوث الإحصائية أ.د. محمد نجيب عبد الفتاح

أمانة المؤتمر

أمين عام المؤتمر
ووكيل المعهد لشئون خدمة المجتمع وتنمية البيئة أ.د. يوسف مصطفى محجوب

الهيئة العلمية مرتبة أجداً

الاسم	الجهة
ابراهيم السيد البطل	معهد الإحصاء - جامعة القاهرة
ابراهيم فرج عبد الرحمن	كلية الحاسبات - جامعة القاهرة
أبو العلا عطيفي	كلية الحاسبات - جامعة القاهرة
أحمد أمين الشيخ	معهد الإحصاء - جامعة القاهرة
أحمد حسن يوسف	معهد الإحصاء - جامعة القاهرة
أحمد سيف النصر	معهد الإحصاء - جامعة القاهرة
أحمد فؤاد محمد عطية	رئيس قسم الإحصاء الرياضي بالمعهد
أشرف عبد المنعم الهادي	معهد الإحصاء - جامعة القاهرة
أكرم إبراهيم صلاح	كلية الحاسبات - جامعة القاهرة
الحسيني عبد البر راضى	معهد الإحصاء - جامعة القاهرة
السيد احمد الشربينى	معهد الإحصاء - جامعة القاهرة
الهام شكري محمد	معهد الإحصاء - جامعة القاهرة
أمانى موسى محمد موسى	معهد الإحصاء - جامعة القاهرة
حجازي محمد زاهر	معهد الإحصاء - جامعة القاهرة

رئيس قسم الإحصاء الحيوي والسكاني بالمعهد
عميد كلية الحاسبات والمعلومات - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
قائم بأعمال رئيس قسم الإحصاء التطبيقي بالمعهد
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
كلية الحاسبات - جامعة القاهرة
وكيل معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
كلية الحاسبات والمعلومات - جامعة عين شمس
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
جامعة أكتوبر للعلوم الحديثة والآداب
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
معهد الإحصاء - جامعة القاهرة
كلية الحاسبات - جامعة القاهرة
عميد كلية الحاسبات - جامعة الاهرام الكندية
رئيس قسم بحوث العمليات بالمعهد
كلية الحاسبات - جامعة القاهرة

دعد محمد فؤاد
ريم محمد رضا بهجت
سحر أحمد نيازى
سعد الدين محمد الشبال
سمية محمد كامل العطار
سمية محمود السعدني
سمير كامل عاشور
سناء العلا حنفى احمد
سيد مشعال تاج الدين
شعبان عبد الحميد شعبان
عاطف محمد عبد المنعم
عبد البديع سالم
عبد الغني محمد عبد الغني
عبد الهادي نبيه أحمد
عصام على أمين
علاء عبد العزيز احمد
على محمد على رجب
عماد الدين حسين حسن
غزال عبد العزيز عامر
فاروق تمام على شعيب
فاطمة عبده إبراهيم الليثي
ليلي عبد العزيز على خليل
ليلي ناصف عطيه
مجدى سيد مصطفى
محمد بيومي
محمد جمال الدين درويش
محمد حسن جاد الله
محمد حسن رسمى

معهد الإحصاء - جامعة القاهرة	محمد مصطفى حسن
عميد معهد الإحصاء - جامعة القاهرة	محمد نجيب عبد الفتاح
معهد الإحصاء - جامعة القاهرة	محمد يحيى البسيونى
معهد الإحصاء - جامعة القاهرة	محمود رياض محمود
معهد الإحصاء - جامعة القاهرة	مرفت حسن غيث
نائب الجامعة البريطانية بمصر	معتز محمد خورشيد
معهد الإحصاء - جامعة القاهرة	منى عبد الفتاح خليفة
معهد الإحصاء - جامعة القاهرة	مها مصطفى كمال
معهد الإحصاء - جامعة القاهرة	نجلاء رجاء سعيد حسن
معهد الإحصاء - جامعة القاهرة	نجلاء عبد المنعم مراد
وكيل كلية الحاسبات - جامعة القاهرة	هدى أنسي
قائم بأعمال رئيس قسم علوم الحاسب بالمعهد	هشام احمد حفنى
معهد الإحصاء - جامعة القاهرة	هشام حسن مخلوف
كلية الحاسبات - جامعة القاهرة	هشام محمد عبد السلام
كلية الحاسبات - جامعة القاهرة	هشام نبيه المهدي
وكيل المعهد لشئون خدمة المجتمع وتنمية البيئة	يوسف مصطفى محبوب

اللجنة الفنية للمؤتمر

سكرتارية المؤتمر

أ. نادية حلمي عواد

أمين المعهد:

أ. سراج الدين إبراهيم

مساعدو اللجنة الفنية للمؤتمر

أ. فوزية أحمد على

أ. جليلة كامل

أ. أمال عبد الكريم

أ. سوسن عبد الحليم

أ. مصطفى عباس

أ. سالمة محمود أحمد

أ. شيماء إسماعيل شحاته

أ. صفاء محمود

أ. ريهام محمد عبد الواحد

أ. هالة إسماعيل

أ. ريهام أحمد شافعي

أ. نسرين إمام

أ. سمر صبري

العلاقات العامة (التدريب)

أ. مجدي حسن عبد الرحمن

أ. رضا أحمد محمد

أ. هناء عبد الحليم

وحدة التصوير والطباعة والتجليد

أ. غريب محمود سالم

الإستعلامات

أ. سلامة عبد الواحد

أ. أبو العلا كمال أبو العلا

تقييم وتنقيح بيانات التعداد السكاني
مع التطبيق على محافظة القليوبية بجمهورية مصر العربية لعام ٢٠٠٦م

* عبدالغني محمد عبدالغني ** سحر عادل رافت
*** عبدالله محمد محمود سرج

١- المقدمة

إن البيانات والمؤشرات الديموجرافية لها أهمية قصوى في حياتنا فعلى أساسها يتم التخطيط ، وتقدير الاحتياجات القومية سواء من السلع أو من الخدمات المختلفة وخاصة ما هو مرتبط منها بالتعليم ، والصحة و الإسكان ... الخ . وعلى الرغم من الحرص الشديد على أن تكون البيانات السكانية التي يتم جمعها على درجة عالية من الدقة ، إلا أن الواقع العملي يثبت تعرض هذه البيانات وخصوصا ببياني العمر والنوع بطريقة أو بأخرى إلى العديد من الأخطاء ، وغالبا لا يخلو أي تعداد في أي بلد من تعرض بياناته لأخطاء الشمول و أخطاء المحتوى وهما يمثلان الأخطاء في البيانات السكانية التي يوفرها التعداد وتلك الأخطاء هي كما يلي: (عبد الغني وآخرون ، ١٩٩١)

أ- خطأ الشمول :

الشمول بالنسبة لتعدادات السكان هو الحصر الشامل لكل أفراد المجتمع في لحظة التعداد ، ويتعرض شمول العد الى نوعين من الأخطاء هما العد الزائد بأن يكون عدد السكان أكبر من حقيقته بسبب عد بعض الأفراد أكثر من مرة والعد الناقص والذي ينتج عن اغفال عد بعض الأفراد داخل حدود الدولة اما بسبب استخدام خرنط غير حديثة أو لعدم امكان الوصول لبعض الأسر في الأماكن النائية أو العشوائية أو لغياب بعض الأسر أثناء عملية العد ، كما ينتج أيضا خطأ الشمول من اغفال العداد لبعض الأسر أو اهمال بعض الأسر ذكر بعض أفرادها عمدا أو سهوا . (فراج ، ١٩٧٥)

ب- خطأ المحتوى :

خطأ المحتوى هو الإبلاغ الخاطئ عن بيانات الأفراد التي يتعرض لها بعض بيانات التعداد وخاصة بيانات العمر و النوع وغيرها والخطأ في بيانات العمر أكثر منها في بيانات النوع ، والتبليغ الخاطئ عن العمر والنوع هو المعني بخطأ المحتوى في هذه الدراسة وقد يحدث هذا الخطأ نتيجة التحيز لأعمار معينة مثلا التي تنتهي بالصفير أو الخمسة أو بالأرقام الزوجية وقد يحدث بشكل متعمد وأيضا قد يرجع هذا الخطأ لجهل المبلغ بالبيان الصحيح وخاصة كبار السن والأميين وبالطبع هذا الخطأ يمكن أن يغير في التركيب العمري والنوعي للسكان عن التركيب العمري والنوعي الحقيقي للمجتمع محل الدراسة .

ويختلف حجم ونوع الأخطاء باختلاف تقدم أو تأخر الدول ومدى خبرتها في إجراء التعدادات ، و أيضا تختلف الأخطاء باختلاف الدوافع والأسباب الكامنة وراء هذه الأخطاء، وتتولد الأخطاء في أي مرحلة من مراحل التعداد ، فقد تظهر في المراحل الأولية بسبب عيب أو قصور في تصميم استمارة التعداد وأسلوب صياغتها ، وقد تأتي الأخطاء من عدم استجابة المبحوثين ، أو من عدم تعاونهم ، كما قد يكون العداد نفسه سببا في وقوع الخطأ عندما يسجل معلومة خاطئة

* استاذ الاحصاء الحيوي والسكاني بمعهد الدراسات والبحوث الاحصائية - جامعة القاهرة ومدير المركز الديموجرافي بالقاهرة .

** استاذ ورئيس قسم الاحصاء والرياضة والتأمين - كلية التجارة - جامعة بنها.

*** مدرس مساعد بقسم الاحصاء والرياضة والتأمين - كلية التجارة - جامعة بنها.

كالعمر مثلا ، كما قد تحدث الأخطاء عند تجهيز البيانات أو عند نشرها ، ولذلك كان من الضروري القيام بعملية تقييم للبيانات السكانية للتأكد والاطمئنان على مدى صحة هذه البيانات ودرجة الاعتماد عليها وذلك باستخدام الطرق والأساليب الصحيحة للوصول إلى بيانات ومؤشرات تتسم بالدقة والوضوح يمكننا الاعتماد عليها والقيام بتحليلها للوصول إلى نتائج صحيحة وغير مضللة لسكان الدولة التي تم فيها إجراء التعداد بل وأيضا تحليل النتائج التي أسفر عنها تقييم بيانات التعداد بحيث تصدر على أساس هذا التحليل دراسات فنية عن الأوضاع السكانية لكي يتم وضع الخطط الاقتصادية-التعليمية - الصحية - الإسكانية ... الخ ، وهذه هي المرحلة التي سيتم التركيز عليها باستفاضة في هذه الدراسة .
(فراج ، ١٩٧٥)

٢- أهداف البحث

ان لهذا البحث أهداف عديدة نعرضها فيما يلي:-

- ١- تقييم بيانات التعداد السكاني لاجمالي جمهورية مصر العربية لتعداد عام ٢٠٠٦م.
- ٢- تقييم بيانات التعداد السكاني لاجمالي محافظة القليوبية لتعداد عام ٢٠٠٦م.
- ٣- تنقيح بيانات التعداد السكاني لاجمالي محافظة القليوبية لتعداد عام ٢٠٠٦م.
- ٤- مقارنة نتائج تقييم بيانات اجمالي جمهورية مصر العربية مع اجمالي محافظة القليوبية لتعداد عام ٢٠٠٦م.

٣- البيانات المستخدمة ومصادرها

لما كان هذا البحث منصب أساسا على تعداد ٢٠٠٦ ، فقد كانت النتائج الأولية والنهائية والتي نشرت بعنوان " النتائج الأولية للتعداد العام للسكان والإسكان والمنشآت مصر ٢٠٠٦ " و " النتائج النهائية للتعداد العام للسكان والإسكان والمنشآت مصر ٢٠٠٦ " هي المصدر الرئيسي للبيانات ، وبالطبع لم تغننا هذه البيانات عن الرجوع إلى بيانات التعدادات المصرية السابقة من تعداد ١٩٩٦ إلى تعداد ١٩٦٠ وأيضا بيانات الإحصاءات الحيوية اللازمة لعمل المقارنات وتطبيق الأساليب والطرق المختلفة .

٤- الرموز والعلاقات المستخدمة

(١): تقييم اجمالي بيانات جمهورية مصر العربية لعام ٢٠٠٦ من حيث الشمول:

٤-١-١- نسبة النوع لإجمالي السكان : (مصطفى ، ١٩٩٢)

فيما يلي حساب نسبة النوع لإجمالي السكان من واقع بيانات تعداد عام ٢٠٠٦م (النتائج النهائية) بالصيغة الآتية :

$$\text{نسبة النوع} = \frac{\text{إجمالي عدد الذكور}}{\text{إجمالي عدد الإناث}} \times 100 = 100 \quad \frac{37219056}{35578975} \times 100 = 104,6$$

ومن الفروض الأساسية لنسبة النوع أنه إذا كانت نسبة النوع لإجمالي السكان تتراوح بين القيمتين ٩٥ ، ١٠٥ % فإنها تكون نسبة مقبولة أما إذا انحرفت عن هذه النسبة بالزيادة أو النقصان دل ذلك على وجود أخطاء شمول في بيانات التعداد لعام ٢٠٠٦ ولكننا وجدنا هنا أن نسبة النوع لإجمالي السكان لتعداد عام ٢٠٠٦ مساوية للقيمة ١٠٤,٦ وهذه نسبة مقبولة لشمول التعداد .

٤-١-٢ معادلة الموازنة: (مصطفى ، ١٩٩٢)

إن أسلوب معادلة الموازنة من أهم الأساليب التي يمكن استخدامها في تقييم بيانات التعداد وتقدير أخطاء الشمول وأبسط صورها الصيغة الآتية:

$$P_1 = P_0 + B - D + I - E$$

حيث:

P_1 = عدد السكان الكلي في التعداد اللاحق

P_0 = عدد السكان الكلي في التعداد السابق

B = عدد المواليد أحياء خلال الفترة بين التعدادين

D = عدد الوفيات خلال الفترة بين التعدادين

I = عدد المهاجرين للداخل (الوافدين) خلال الفترة بين التعدادين

E = عدد المهاجرين للخارج (النازحين) خلال الفترة بين التعدادين

وبتطبيق معادلة الموازنة لتعداد السكان لعام ٢٠٠٦ لاجمالي جمهورية مصر العربية كان الناتج كما يلي:

سكان مصر ٢٠٠٦ = ٧٣٧٧٨٤٤٣ نسمة

$$\therefore \text{نسبة النقص المتوقع في العدد} = \frac{\text{عدد السكان المصحح لعام 2006} - \text{العدد المشاهد لعام 2006}}{\text{العدد المشاهد لعام 2006}} \times 100$$

∴ النقص المتوقع في العدد = ١,٤% □ ١% لتعداد ٢٠٠٦ مقابل ٢,٨٤% □ ٣% لتعداد ١٩٩٦ ، وهذه نسبة مقبولة جدا لشمول تعداد ٢٠٠٦م.

٤-١-٣ معدلات النمو التعدادية: (العيسائي ، ١٩٩٣)

وقد تم حساب هذا المعدل كما يلي :

١- حساب معدل النمو السنوي من تعداد عام ١٩٩٦ الى تعداد عام ٢٠٠٦ =

$$\frac{١٩,٥ + ١٩,١ + ١٩,٣ + ١٩,٧ + ٢٠,١ + ٢٠,٥ + ٢١,١ + ٢٠,٦ + ٢١,٥ + ٢١,٠ + ٢١,٨}{١١} = ٢,٠٤\%$$

٢- حساب العدد المصحح ونقص الشمول لتعداد ٢٠٠٦ بالطريقة الآسية وذلك كما يلي :

$$P_{2006} = P_{1996} e^{r(2006-1996)}$$

$$P_{2006} = (59312914) e^{0.0204(10)} = 72735317$$

عدد السكان المقدر في التاريخ ٢٠٠٦

$$100 \times \frac{\text{عدد السكان المقدر في التاريخ 2006}}{\text{عدد السكان المشاهد في التاريخ 2006}} = \text{درجة اكتمال العدد}$$

$$\therefore \text{درجة اكتمال العدد} = 100 \times \frac{72735317}{72798031} = 99,9\%$$

نقص الشمول = ١٠٠ - ٩٩,٩ = ٠,١

وهذا دليل على أن تعداد عام ٢٠٠٦ قد حقق شمولا بدرجة مقبولة لإجمالي عدد السكان في مصر.

٤-١-٤ نسب البقاء: (عبد الغني وآخرون ، ١٩٩١)

وقد تم استخدام هذه الطريقة كما يلي :

$$\text{عدد السكان (١٥+)} ٢٠٠٦ = \frac{٤٩٧١٦٣٩٣}{٥٢٤٥٧٦٧٢} = ٠,٩ \text{ عدد السكان (٥+)} ١٩٩٦$$

وهذه نسبة مقبولة حيث أنها لا تتعدى المدى المسموح به من ٠,٨ إلى ٠,٩ أي أن تعداد ٢٠٠٦ حقق شمولا بدرجة مقبولة.

٤-١-٥ طريقة التوزيع العمري الحقيقي: (عبد الغني وآخرون ، ١٩٩١)

وقد تم حساب هذه الطريقة كما يلي:

(١) حساب عدد المواليد الذكور والإناث للعام الذي منتصفه ٢١ نوفمبر ٢٠٠٦ وهو يوم العد وذلك عن طريق إحصائيات المواليد والوفيات عام ٢٠٠٦ (الجهاز المركزي للتعبئة العامة والإحصاء) ومن المفروض أخذ متوسط ثلاث سنوات وهذه السنوات هي ٢٠٠٥ ، ٢٠٠٦ ، ٢٠٠٧ وذلك لعدم التحيز لمواليد سنة معينة و أيضا للتخلص من التذبذبات ، التغيرات العشوائية التي تحدث في عدد المواليد ذكورا وإناثا نتيجة لقصور التسجيل، ولكن لعدم توافر بيانات عن مواليد عام ٢٠٠٧ فسيتم الاكتفاء بأخذ مواليد ووفيات تعداد عام ٢٠٠٦ فقط كما يلي:

عدد المواليد الذكور = ٩٥١٠٨٥

عدد المواليد الإناث = ٩٠٢٦٦١

(٢) تم حساب قيم L_t من جدول الحياة المحسوب في بحث اسقاطات السكان المستقبلية لمحافظة مصر -

المركز الديموجرافي بالقاهرة $\Leftarrow L_t = n L_x / L_0$ حيث L_0 أساس الجدول = ١٠٠٠٠٠٠

(٣) لحساب معدلات النمو العمرية اللحظية يوم التعداد تم استخدام اقل نموذج لكل فئة عمرية تمثل خمس قراءات في التعدادات التالية ١٩٦٠ ، ١٩٧٦ ، ١٩٨٦ ، ١٩٩٦ ، ٢٠٠٦ على أساس أن عدد السكان في كل فئة دالة في الزمن (t) ، تم حساب معدلات النمو العمرية r_i من العلاقة التالية :

$$r_i = \frac{1}{p_i} \frac{dp_i}{dt}$$

حيث :

$i = ١, ٢, ٣, \dots, ١٦$ والتي تناظر الفئات (٠ - ٤) ، (٥ - ٩) ، ... (٧٥ +)

r_i = معدل النمو السكاني للفئة العمرية (i) في لحظة التعداد

p_i = عدد السكان في الفئة العمرية (i)

t = الزمن

وبتطبيق العلاقة التالية تم الحصول على التوزيع العمري الموضح بالجدول التالي :

$$N_i = B L_i \exp \left[- \int_0^a r_i du \right]$$

جدول رقم (٥ - ٢)

التوزيع العمري للسكان باستخدام طريقة التوزيع العمري الحقيقي

فئات العمر	الذكور	الإناث	المجموع
٥ - ٠	٤٠٢٩٠٥٢	٣٧٩٣٩٢٤	٧٨٢٢٩٧٦
١٠ - ٥	٤٠١٣٤٨٨	٣٧٣٤٠١٦	٧٧٤٧٥٠٤
١٥ - ١٠	٤٠٧٤٢٠٢	٣٧٤٨٧٧٩	٧٨٢٢٩٨١
٢٠ - ١٥	٤٤٥٩٧٤٧	٤١٩٥٢٢٩	٨٦٥٤٩٧٦
٢٥ - ٢٠	٤١٠٤٠٩٩	٣٨٧٥١٧٧	٧٩٧٩٢٧٦
٣٠ - ٢٥	٣٢١٣٥٣١	٣٢٦٣٠٦٥	٦٤٧٦٥٩٦
٣٥ - ٣٠	٢٤٦٢٦٠٤	٢٣٣٤٦٢٢	٤٧٩٧٢٢٦
٤٠ - ٣٥	٢٣٥٣٦٩٤	٢٣٦٥٢٣٤	٤٧١٨٩٢٨
٤٥ - ٤٠	٢١٠٧٣٥٢	٢٠٤٠٠٣٧	٤١٤٧٣٨٩
٥٠ - ٤٥	١٩١٢١١٠	١٨١١٧٤٩	٣٧٢٣٨٥٩
٥٥ - ٥٠	١٥٧١٧١٦	١٥٣٠٥٨٥	٣١٠٢٣٠١
٦٠ - ٥٥	١٢٣٦٣٩٧	١٠٦٠٠٩٣	٢٢٩٦٤٩٠
٦٥ - ٦٠	٩١٥٧٤٠	٨١٣٠٠٢	١٧٢٨٧٤٢
٧٠ - ٦٥	٦٥٢٧٧٩	٥٥٧٢٠٠	١٢٠٩٩٧٩
٧٥ - ٧٠	٤٠٩٨٥٣	٣٩٠٦٦٤	٨٠٠٥١٧
+٧٥	٣٨٣٦٢٢	٣٦٥٠٨١	٧٤٨٧٠٣
المجموع	٣٧٨٩٩٩٨٦	٣٥٨٧٨٤٥٧	٧٣٧٧٨٤٤٣

المصدر: حسابات الباحث بالتطبيق في صيغة التوزيع العمري الحقيقي وباستخدام البرنامج الإحصائي PAS

٠. النقص المتوقع في العدد = ١,٤ % □ ١ % لتعداد ٢٠٠٦ م وهذه نسبة مقبولة جدا لشمول تعداد ٢٠٠٦ م.

(٢): تقييم إجمالي بيانات جمهورية مصر العربية لعام ٢٠٠٦ من حيث المحتوى:

١-٢-٤ تحليل نسبة النوع : (العيسائي ، ١٩٩٣)

إن نسبة النوع لأي مجتمع تأخذ نمطا ثابتا فهي تتراوح ما بين القيمتين ١٠٤ ، ١٠٧ وذلك عند الميلاد وعند غياب التغيرات الناجمة عن الحروب والأوبئة والكوارث ، ثم تأخذ في النقصان تدريجيا حتى القيمة ١٠٠ في الأعمار المتوسطة ثم تأخذ أيضا في النقصان تدريجيا مع تقدم العمر حيث من المتوقع أن تكون معدلات الوفيات للذكور أعلى منها للإناث في هذه الأعمار المتقدمة ، وفيما يلي يوضح الجدول التالي نسبة النوع للفئات الخمسية وانحرافاتهن عن الرقم ١٠٠ :

جدول (١)

تحليل نسب النوع حسب السن لتعداد عام ٢٠٠٦ م

الانحراف عن ١٠٠	نسبة النوع %	الإناث	الذكور	فئات العمر
٥,٢	١٠٥,٢	٣٧٦٢٢٥٦	٣٩٥٦٦٦٤	٥-٠
٦,٤	١٠٦,٤	٣٧٠٢٨٤٧	٣٩٤١٣٨٠	١٠-٥
٧,٦	١٠٧,٦	٣٧١٧٤٨٨	٤٠٠١٠٠٣	١٥-١٠
٥,٣	١٠٥,٣	٤١٦٠٢١١	٤٣٧٩٦٢١	٢٠-١٥
٤,٨	١٠٤,٨	٣٨٤٢٨٣٠	٤٠٣٠٣٦٢	٢٥-٢٠
٢,٥-	٩٧,٥	٣٢٣٥٨٢٨	٣١٥٥٧٩٥	٣٠-٢٥
٤,٤	١٠٤,٤	٢٣١٥١٣٥	٢٤١٨٣٦٠	٣٥-٣٠
١,٥-	٩٨,٥	٢٣٤٥٤٩١	٢٣١١٤٠٦	٤٠-٣٥
٢,٣	١٠٢,٣	٢٠٢٣٠٠٨	٢٠٦٩٤٩١	٤٥-٤٠
٤,٥	١٠٤,٥	١٧٩٦٦٢٦	١٨٧٧٧٥٦	٥٠-٤٥
١,٧	١٠١,٧	١٥١٧٨٠٩	١٥٤٣٤٧٧	٥٥-٥٠
١٥,٥	١١٥,٥	١٠٥١٢٤٥	١٢١٤١٨٤	٦٠-٥٥
١١,٥	١١١,٥	٨٠٦٢١٥	٨٩٩٢٨٧	٦٥-٦٠
١٦	١١٦	٥٥٢٥٤٩	٦٤١٠٥١	٧٠-٦٥
٣,٩	١٠٣,٩	٣٨٧٤٠٣	٤٠٢٤٨٩	٧٥-٧٠
٤,١	١٠٤,١	٣٦٢٠٣٤	٣٧٦٧٣٠	+٧٥
		٣٥٥٧٨٩٧٥	٣٧٢١٩٠٥٦	الإجمالي

المصدر : حسابات الباحث اعتمادا على النتائج النهائية لتعداد السكان والاسكان والمنشآت مصر ٢٠٠٦ " القاهرة ، ٢٠٠٨ .

ونلاحظ من الجدول السابق وجود أخطاء في بعض الأعمار لبيانات تعداد عام ٢٠٠٦ لإجمالي الجمهورية.

٢-٢-٤ تحليل نسبة العمر : (العيساني ، ١٩٩٣)

يتم حساب نسبة العمر في فئة معينة عن طريق خارج قسمة عدد الأفراد في الفئة المختارة المحددة على الوسط الحسابي لعدد الأفراد في الفئة العمرية السابقة واللاحقة ويتم حساب نسبة العمر لكل من الذكور والإناث كل على حدة ، وأي انحراف كبير عن الرقم ١٠٠ يوضح وجود أخطاء تبليغ عن العمر مع استثناء الأعمار المتقدمة الكبيرة والجدول التالي يوضح نسب العمر للفئات العمرية الخمسية المختلفة لتعداد عام ٢٠٠٦ وانحرافاتهما عن الرقم ١٠٠ وذلك كما يلي:

جدول (٢)

نسبة العمر وانحرافاتهما في الفئات العمرية لتعداد عام ٢٠٠٦

فئات العمر	نسبة العمر		الانحرافات عن القيمة ١٠٠	
	ذكور	إناث	ذكور	إناث
٥-١٠				
١٠-١٥	٩٩,١	٩٩	١٠٩-	١-
١٥-٢٠	٩٦,٢	٩٤	٣,٨-	٦-
٢٠-٢٥	١٠٩,١	١١٠,١	٩,١	١٠,١
٢٥-٣٠	١٠٧	١٠٣,٩	٧	٣,٩
٣٠-٣٥	٩٧,٩	١٠٥,١	٢,١-	٥,١
٣٥-٤٠	٨٨,٥	٨٣	١١,٥-	١٧-
٤٠-٤٥	١٠٣	١٠٨,١	٣	٨,١
٤٥-٥٠	٩٨,٨	٩٧,٧	١,٢-	٢,٣-
٥٠-٥٥	١٠٣,٩	١٠١,٥	٣,٩	١,٥
٥٥-٦٠	٩٩,٨	١٠٦,٦	٠,٢-	٦,٦
٦٠-٦٥	٩٩,٤	٩٠,٥	٠,٦-	٩,٥-
٦٥-٧٠	٩٦,٩	١٠١,٥	٣,١-	٠,٥
٧٠-٧٥	٩٨,٥	٩٢,٦	١,٥-	٧,٤-

المصدر : حسابات الباحث اعتمادا على النتائج النهائية لتعداد السكان والاسكان والمنشآت مصر ٢٠٠٦ " القاهرة ، ٢٠٠٨ .

ونلاحظ من الجدول السابق وجود أخطاء في بعض الأعمار لبيانات تعداد عام ٢٠٠٦ لاجمالي الجمهورية.

٣-٢-٤ مقياس سكرتارية الأمم المتحدة: (United Nations, 1955)

ان مقياس سكرتارية الأمم المتحدة يعطي مؤشرا على وجود أخطاء في بيانات العمر والنوع وذلك في ظل الظروف المستقرة للبلاد .

ويوضح الجدول التالي كيفية حساب مقياس سكرتارية الأمم المتحدة لبيانات تعداد عام ٢٠٠٦:

جدول (٣)

حساب مقياس العمر بأسلوب سكرتارية الأمم المتحدة لتعداد ٢٠٠٦

نسبة العمر للإناث		نسبة العمر للذكور		الفروق المتتالية	نسبة النوع		الفئات العمرية
الفرق عن ١٠٠	القيمة	الفرق عن ١٠٠	القيمة		القيمة	الفرق	
					١٠٥,٢		٥-١٠
١-	٩٩	٠,٩-	٩٩,١	١,٣	١٠٦,٤		١٠-١٥
٥,٤-	٩٤,٦	٣,٨-	٩٦,٢	١,٢	١٠٧,٦		١٥-٢٠
١٠,١	١١٠,١	٩,١	١٠٩,١	٢,٤-	١٠٥,٣		٢٠-٢٥
٣,٩	١٠٣,٩	٧	١٠٧	٠,٥-	١٠٤,٨		٢٥-٣٠
٥,١	١٠٥,١	٢,١-	٩٧,٩	٧,٣-	٩٧,٥		٣٠-٣٥
١٧-	٨٣	١١,٥-	٨٨,٥	٦,٩	١٠٤,٤		٣٥-٤٠
٨,١	١٠٨,١	٣	١٠٣	٥,٩-	٩٨,٥		٤٠-٤٥
٢,٣-	٩٧,٧	١,٢-	٩٨,٨	٣,٨	١٠٢,٣		٤٥-٥٠
١,٥	١٠١,٥	٣,٩	١٠٣,٩	٢,٢	١٠٤,٥		٥٠-٥٥
٦,٦	١٠٦,٦	٠,٢-	٩٩,٨	٢,٨-	١٠١,٧		٥٥-٦٠
٩,٥-	٩٠,٥	٠,٦-	٩٩,٤	١٣,٨	١١٥,٥		٦٠-٦٥
٠,٥	١٠٠,٥	٣,١-	٩٦,٩	٤-	١١١,٥		٦٥-٧٠
٧,٤-	٩٢,٦	١,٥-	٩٨,٥	٤,٥	١١٦		٧٠-٧٥
				١٢,١-	١٠٣,٩		٧٥-٨٠
					١٠٤,١		٨٠-٨٥
							٨٥-٩٠
							٩٠-٩٥
							٩٥-١٠٠
							١٠٠+
							متوسط الانحرافات المطلقة
	٦		٣,٧		٤,٩		

المصدر : حسابات الباحث اعتمادا على النتائج النهائية لتعداد السكان والاسكان والمنشآت مصر ٢٠٠٦ " القاهرة ، ٢٠٠٨

• مقياس سكرتارية الأمم المتحدة = ٣ (متوسط القيم المطلقة لفروق القيم المتتالية الخاصة بالنوع) + متوسط القيم المطلقة لفروق قيم نسب العمر للذكور + متوسط القيم المطلقة لفروق قيم نسب العمر للإناث

٠. مقياس سكرتارية الأمم المتحدة = $(٤,٩)٣ + ٣,٧ + ٦ = ٢٤,٤$

سنة التعداد	١٩٦٠	١٩٧٦	١٩٨٦	١٩٩٦	٢٠٠٦
قيمة المقياس	٥٠,١	٥٢,١٦	٤٥,٧	٣٨,٤	٢٤,٤

المصدر : عبد الغني وآخرون ، ١٩٩١ وحسابات الباحث لتعداد ٢٠٠٦ .

من الجدول السابق نلاحظ أن قيمة المقياس كانت في الثلاث تعدادات قبل تعداد ١٩٩٦ كبيرة وأكبر من الرقم ٤٠ مما يدل على وجود أخطاء كبيرة في بيانات العمر والنوع بينما انخفضت قيمة المقياس في تعداد ١٩٩٦ لتصبح ٣٨,٤ ثم انخفضت أيضا في تعداد ٢٠٠٦ لتصبح ٢٤,٤ ويدل ذلك على وجود تحسن ملحوظ في البيانات إلا أن ذلك لا يفي بوجود الأخطاء الكبيرة في بيانات العمر والنوع لتعداد عام ٢٠٠٦ .

(٣) : تقييم إجمالي بيانات محافظة القليوبية لعام ٢٠٠٦ من حيث الشمول:

٤-٣-١ نسبة النوع لإجمالي السكان :

وبتطبيق هذه الطريقة كان الناتج كما يلي:

$$\text{نسبة النوع} = \frac{2187288}{2064384} \times 100 = 106\%$$

وهذه النسبة غير مقبولة لشمول تعداد ٢٠٠٦ لمحافظة القليوبية ومعنى ذلك أن هناك إما عد زائد لإجمالي عدد الذكور أو عد ناقص لإجمالي عدد الإناث.

٤-٣-٢ معادلة الموازنة:

وبتطبيق هذه الطريقة كان الناتج كما يلي:

$$\text{عدد سكان محافظة القليوبية لتعداد ٢٠٠٦} = ٣٣٠١٢٤٤ + ١٠٠٥١٢٥ - ٢٤٧٥٤٨ = ٤٠٥٨٨٢١$$

عدد السكان المصحح لتعداد ٢٠٠٦ - العدد المشاهد لتعداد ٢٠٠٦

$$\text{النقص المتوقع في العد} = \frac{\text{العدد المشاهد لتعداد ٢٠٠٦}}{١٠٠ \times}$$

العدد المشاهد لتعداد ٢٠٠٦

$$٤٢٥١٦٧٢ - ٤٠٥٨٨٢١$$

$$\text{النقص المتوقع في العد} = \frac{٤٢٥١٦٧٢}{٤٠٥٨٨٢١} \times ١٠٠ = ١٠٤,٥$$

ومعنى أن النقص المتوقع في العد بالسالب = ١٠٤,٥ أي أن هذا يدل على أن هناك زيادة في العد قدرت ب ١٠٤,٥ أي أن تعداد محافظة القليوبية لعام ٢٠٠٦ لم يحقق شمو لا مقبولا وأن هناك زيادة في إجمالي عدد الذكور وقصور تسجيل أو (عد ناقص) لإجمالي عدد الإناث ، وذلك بفرض صحة الإحصاءات الحيوية لمحافظة القليوبية.

٤-٣-٣ نسب البقاء:

وبتطبيق هذه الطريقة كان الناتج كما يلي:

$$\text{عدد السكان (١٥+) ٢٠٠٦} = \frac{٢٨٩٥٨٨٣}{٢٩٢٩٨٩٨} \times \text{عدد السكان (٥+) ١٩٩٦}$$

وهذه النسبة غير مقبولة لشمول التعداد لمحافظة القليوبية لأنها تتعدى المدى المسموح به وهو ٠,٨ ، ٠,٩ ، وهذا يدل على أن تعداد ٢٠٠٦ لمحافظة القليوبية لم يحقق شمو لا مقبولا .

٤-٣-٤ معدلات النمو التعدادية:

وقد تم حساب هذا المعدل كما يلي :

١- حساب معدل النمو السنوي من تعداد عام ١٩٩٦ إلى تعداد عام ٢٠٠٦ لمحافظة القليوبية =

$$19,5 + 19,1 + 19,3 + 19,7 + 20,1 + 20,5 + 21,1 + 20,6 + 21,5 + 21,0 + 21,8 = 2,04\%$$

١١

٢- حساب العدد المصحح ونقص الشمول لتعداد ٢٠٠٦ لمحافظة القليوبية بالطريقة الآسية وذلك كما يلي :

$$P_{2006} = P_{1996} e^{r(2006-1996)}$$

$$P_{2006} = (3301244) e^{0.0196(10)} = 4016052$$

عدد السكان المقدر في التاريخ ٢٠٠٦ التعداد الحالي

$$100 \times \frac{\text{عدد السكان المقدر في التاريخ ٢٠٠٦ التعداد الحالي}}{\text{عدد السكان المشاهد في التاريخ ٢٠٠٦ التعداد الحالي}} = \text{درجة اكتمال العدد}$$

$$4016052$$

$$94,5\% = 100 \times \frac{4016052}{4251672} = \text{نسبة النقص المتوقع في العدد}$$

$$5,5\% = 100\% - 94,5\% = \text{نقص الشمول}$$

وهذا دليل على أن تعداد عام ٢٠٠٦ لمحافظة القليوبية لم يحقق شمو لا بدرجة مقبولة .

(٤): تقييم اجمالي بيانات محافظة القليوبية لعام ٢٠٠٦ من حيث المحتوى:

٤-٤-١ تحليل نسبة النوع:

وبتطبيق هذه الطريقة كان الناتج كما يلي:

جدول (٤)

تحليل نسب النوع حسب السن لمحافظة القليوبية لتعداد عام ٢٠٠٦

فئات العمر	الذكور	الإناث	نسبة النوع %	الانحراف عن ١٠٠
٥-١٠	٢٤٦١٠٧	٢٣٣٧٨٣	١٠٥,٣	٥,٣
١٠-١٥	٢٣٣١٠١	٢١٩٦٥٣	١٠٦,١	٦,١
١٥-٢٠	٢١٨٢١٧	٢٠٤٩٢٨	١٠٦,٥	٦,٥
٢٠-٢٥	٢٣٧٨٢٨	٢٢٩٠٠٦	١٠٣,٩	٣,٩
٢٥-٣٠	٢٣٥٥١٩	٢٣٧٢٥٩	٩٩,٣	-٠,٧
٣٠-٣٥	٢٠١٩١٩	٢٠٧٦٦٥	٩٧,٢	-٢,٨
٣٥-٤٠	١٦٠٥٧٠	١٤٥٠٢٥	١١٠,٧	١٠,٧
٤٠-٤٥	١٤٨٧٢٤	١٣٥٠٧٣	١١٠,١	١٠,١

٩,٥	١٠٩,٥	١١٦٠٠٣	١٢٧٠٨٠	٤٥-٤٠
١٠,٤	١١٠,٤	٩٨٠٦٤	١٠٨٢٥٢	٥٠-٤٥
٦,٤	١٠٦,٤	٨٣٤٤٠	٨٨٨١١	٥٥-٥٠
٢١,١	١٢١,١	٥٥٨٢٦	٦٧٦٠١	٦٠-٥٥
٢٠,٨	١٢٠,٨	٣٩٥٩٤	٤٧٨١٥	٦٥-٦٠
٢٠,١	١٢٠,١	٢٥٠٢٩	٣٠٠٥٧	٧٠-٦٥
٦,٩	١٠٦,٩	١٧٦٤٤	١٨٨٥٧	٧٥-٧٠
٢,٧	١٠٢,٧	١٦٣٩٢	١٦٨٣٠	+٧٥
				الإجمالي

المصدر : حسابات الباحث اعتمادا على بيانات محافظة القليوبية النهائية للسكان و الاسكان والمنشآت مصر ٢٠٠٦.

ويتضح من الجدول السابق وجود تأرجح بين القيم من الزيادة الى النقصان والعكس صحيح وهذا دليل على وجود خطأ في بيان الغمر لمحافظة القليوبية ٢٠٠٦.

٢-٤-٤ تحليل نسبة العمر:

وبتطبيق هذه الطريقة كان الناتج كما يلي:

جدول (٥)

نسبة العمر وانحرافاتهما في الفئات العمرية لمحافظة القليوبية لتعداد عام ٢٠٠٦

نسبة العمر		الانحرافات		فئات العمر
ذكور	إناث	ذكور	إناث	
				٥-٠
١٠٠,٤	١٠٠,١	٠,٤	٠,١	١٠-٥
٩٢,٧	٩١,٤	٧,٣-	٨,٦-	١٥-١٠
١٠٤,٨	١٠٣,٦	٤,٨	٣,٦	٢٠-١٥
١٠٧,١	١٠٨,٧	٧,١	٨,٧	٢٥-٢٠
١٠٢	١٠٨,٦	٢	٨,٦	٣٠-٢٥
٩١,٦	٨٤,٦	٨,٤-	-١٥,٤	٣٥-٣٠
١٠٣,٤	١٠٣,٥	٣,٤	٣,٥	٤٠-٣٥
٩٨,٩	٩٩,٥	١,١-	-٠,٥	٤٥-٤٠
١٠٠,٣	٩٨,٣	٠,٣	-١,٧	٥٠-٤٥
١٠١	١٠٨,٤	١	٨,٤	٥٥-٥٠

٦٠-٥٥	٩٩	٩٠,٧	١-	-٩,٣
٦٥-٦٠	٩٧,٩	٩٧,٩	٢,١-	-٢,١
٧٠-٦٥	٩٠,٢	٨٧,٥	٩,٨-	-١٢,٥
٧٥-٧٠				
+٧٥				

المصدر : حسابات الباحث اعتمادا على بيانات محافظة القليوبية النتائج النهائية للسكان و الاسكان والمنشآت مصر

٢٠٠٦ .

ومن الجدول السابق يتضح وجود انحرافات واضحة عن الرقم ١٠٠ وهذا يدل على وجود اخطاء ايضا في بيان العمر لتعداد ٢٠٠٦ لمحافظة القليوبية.

٤-٣ مقياس سكرتارية الأمم المتحدة:

وينطبق هذه الطريقة كان الناتج كما يلي:

جدول (٦)

حساب مقياس العمر بأسلوب سكرتارية الأمم المتحدة لمحافظة القليوبية لتعداد ٢٠٠٦

نسبة العمر للإناث		نسبة العمر للذكور		نسبة النوع		الفئات العمرية
القيم	الفرق عن ١٠٠	القيم	الفرق عن ١٠٠	القيم	الفرق المتتالية	
				١٠٥,٣		٥-١
١٠٠,١	٠,١	١٠٠,٤	٠,٤	١٠٦,١	٠,٨	١٠-٥
٩١,٤	-٨,٦	٩٢,٧	-٧,٣	١٠٦,٥	٠,٤	١٥-١٠
١٠٣,٦	٣,٦	١٠٤,٨	٤,٨	١٠٣,٩	-٢,٦	٢٠-١٥
١٠٨,٧	٨,٧	١٠٧,١	٧,١	٩٩,٣	-٤,٦	٢٥-٢٠
١٠٨,٦	٨,٦	١٠٢	-٢	٩٧,٢	-٢,١	٣٠-٢٥
٨٤,٦	-١٥,٤	٩١,٦	-٨,٤	١١٠,٧	١٣,٥	٣٥-٣٠
١٠٣,٥	٣,٥	١٠٣,٤	٣,٤	١١٠,١	٠,٦	٤٠-٣٥
٩٩,٥	٠,٥	٩٨,٩	-١,١	١٠٩,٥	٠,٦	٤٥-٤٠
٩٨,٣	-١,٧	١٠٠,٣	٠,٣	١١٠,٤	٠,٩	٥٠-٤٥
١٠٨,٤	٨,٤	١٠١	١	١٠٦,٤	-٤	٥٥-٥٠
٩٠,٧	-٩,٣	٩٩	-١	١٢١,١	١٤,٧	٦٠-٥٥
٩٧,٩	-٢,١	٩٧,٩	-٢,١	١٢٠,٨	٠,٣	٦٥-٦٠

١٢,٥-	٨٧,٥	٩,٨-	٩٠,٢	٠,٧-	١٢٠,١	٧٠-٦٥
				١٣,٢-	١٠٦,٩	٧٥-٧٠
					١٠٢,٧	+٧٥
	٦,٤		٣,٧		٤,٢	متوسط الانحرافات المطلقة

المصدر : حسابات الباحث اعتمادا على بيانات محافظة القليوبية النتائج النهائية للسكان و الاسكان والمنشآت مصر ٢٠٠٦.

والجدول التالي يوضح قيمة المقياس في آخر خمس تعدادات

سنة التعداد	١٩٦٠	١٩٧٦	١٩٨٦	١٩٩٦	٢٠٠٦
قيمة المقياس	٥٤,٥	٥٨,٤	٤٧,٨	٣٨	٢٢,٧

ونلاحظ من الجدول السابق أن قيمة المقياس كانت في الثلاث تعدادات التي كانت قبل تعداد ١٩٩٦ كبيرة وأكبر من الرقم ٤٠ مما يدل على وجود أخطاء كبيرة في بيانات العمر والنوع بينما انخفضت قيمة المقياس في تعداد ١٩٩٦ لتصبح ٣٨ ثم انخفضت أيضا في تعداد ٢٠٠٦ لتصبح ٢٢,٧ مما يدل ذلك على وجود تحسن ملحوظ في البيانات إلا أن ذلك لا يلغي وجود أخطاء في بيانات العمر والنوع لتعداد محافظة القليوبية لعام ٢٠٠٦ .

(٥) تنقيح بيانات تعداد محافظة القليوبية ٢٠٠٦ :

وقد تم التنقيح باستخدام طريقة سترونج والتي كان ناتجها كما يلي : (عبد الغني وآخرون ، ١٩٩١)

جدول (٧)

التوزيع العمري و النوعي و التوزيع النسبي المعدل باستخدام طريقة سترونج لمحافظة القليوبية لتعداد ٢٠٠٦

فئات العمر	أعداد السكان المعدلة بالنسبة للمحتوى		أعداد السكان المعدلة بالنسبة لشمول	التوزيع النسبي المعدل باستخدام طريقة سترونج		
	ذكور	إناث		جملة	ذكور	إناث
٥-٠	٢٣٠٢٤٩	٢١٦٦٦٥	٤٤٦٩١٤	٤٥١٨٢٥	١١,٠٢	١٠,٩٩
١٠-٥	٢٢٧٣٩٤	٢١٦٣٦٧	٤٤٣٧٦١	٤٣٨٨٥٠	١١,٨٩	١٠,٩٣
١٥-١٠	٢٢٣٤٨٦	٢١٤٨٩٢	٤٣٨٣٧٨	٤٢٧٢٦٩	١٠,٧٠	١٠,٨٠
٢٠-١٥	٢١٤٣٠٩	٢٠٧٥٣٧	٤٢١٨٤٦	٤٢٢٦٦١	١٠,٢٦	١٠,٥٣
٢٥-٢٠	٢٠٥١٧١	٢٠١٠٨٥	٤٠٦٢٥٦	٤٤٤٩٤٧	٩,٨٢	١٠,٢٠
٣٠-٢٥	١٨٧٤٩٢	١٨٢٥١٥	٣٧٠٠٠٧	٣٩٧٧٠٩	٨,٩٨	٩,٢٦
٣٥-٣٠	١٦٥١٣٦	١٥٧٠٣٨	٣٢٢١٧٤	٣٠٤٢٣٣	٧,٩١	٧,٩٤

٦,٨٦	٦,٨٣	٦,٨٩	٢٥٨٦٣٦	٢٧٨٥٦٧	١٣٤٥٥٥	١٤٤٠١٢	٤٠-٣٥
٥,٧٣	٥,٦٢	٥,٨٣	٢٣٢٣٩٢	٢٣٢٦٧٠	١١٠٨٣٧	١٢١٨٣٣	٤٥-٤٠
٤,٧٨	٤,٦٦	٤,٩٠	١٩٦٧٨٤	١٩٤١٩٠	٩١٨٥٩	١٠٢٣٣١	٥٠-٤٥
٣,٨٩	٣,٧٨	٤,٠٠	١٥٩٤٩٤	١٥٨٠٩٩	٧٤٦٠٤	٨٣٤٩٥	٥٥-٥٠
٣,٠٧	٢,٩٧	٣,١٧	١٢٢٨٧٩	١٢٤٨١٦	٥٨٦٦٧	٦٦١٤٩	٦٠-٥٥
٢,٢٣	٢,١٣	٢,٣٢	٧٩٩٢٩	٩٠٢٩٦	٤٢٠٣٩	٤٨٢٥٧	٦٥-٦٠
١,٦٠	١,٥٤	١,٦٦	٥٦١٥٤	٧٦٦٠١	٣٠٣٢١	٣٤٥٦٥	٧٠-٦٥
١,٠٥	١,٠٣	١,٠٧	٣٨٧٨١	٤٢٧٣٧	٢٠٣٥٧	٢٢٣٨٠	٧٥-٧٠
٠,٥٩	٠,٦٢	٠,٥٦	٢٧٨٠٥	٢٣٨٤٨	١٢١٤٧	١١٧٠١	+٧٥
١٠٠	١٠٠	١٠٠	٤٠٦٠٣٤٧	٤٠٥٩٤٤٧	١٩٧١٤٨٧	٢٠٨٧٩٦٠	الإجمالي

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء " النتائج النهائية لتعداد السكان والاسكان والمنشآت مصر ٢٠٠٦ " القاهرة ، ٢٠٠٨

وحسابات الباحث اعتمادا على الصيغ المذكورة سابقا وباستخدام البرنامج الإحصائي PAS

وهذا هو التوزيع العمري المعدل لاجمالي محافظة القليوبية لتعداد عام ٢٠٠٦ والذي سيتم الأخذ به في الاعتبار.

٥- نتائج المقارنات

١-٥ مقارنة تقييم جمهورية مصر العربية ومحافظة القليوبية لتعداد عام ٢٠٠٦ من حيث الشمول :

بعد النظر في نتائج تقييم اجمالي جمهورية مصر العربية واجمالي محافظة القليوبية لتعداد عام ٢٠٠٦ من جهة الشمول

نجد أن اداء الجمهورية كاملة أفضل من أداء محافظة القليوبية لنفس العام. وهذا يدل على التحسن الواضح لشمول

الجمهورية لعام ٢٠٠٦ عن نظيره في التعدادات السابقة.

٢-٥ مقارنة تقييم جمهورية مصر العربية ومحافظة القليوبية لتعداد عام ٢٠٠٦ من حيث المحتوى :

بعد القيام بتقييم فئات العمر لكل من اجمالي جمهورية مصر العربية واجمالي محافظة القليوبية لتعداد عام ٢٠٠٦ وجدنا أن

هناك فئات أفضل لاجمالي الجمهورية و فئات أفضل لمحافظة القليوبية ولكن الاداء العام الأفضل لاجمالي الجمهورية

ولكن مع وجود أخطاء في بيانات العمر لكل منهما .

٦- المراجع

٦-١ المراجع العربية:

١- العيسائي ، محمد مفرح صالح ١٩٩٣ " تقييم وتنقيح البيانات الأساسية في تعداد ١٩٨٦ مع تقدير المعلومات الخاصة بالإنجاب والوفيات في المحافظات الشمالية للجمهورية اليمنية " رسالة ماجستير مقدمة في معهد الدراسات والبحوث الإحصائية ، جامعة القاهرة

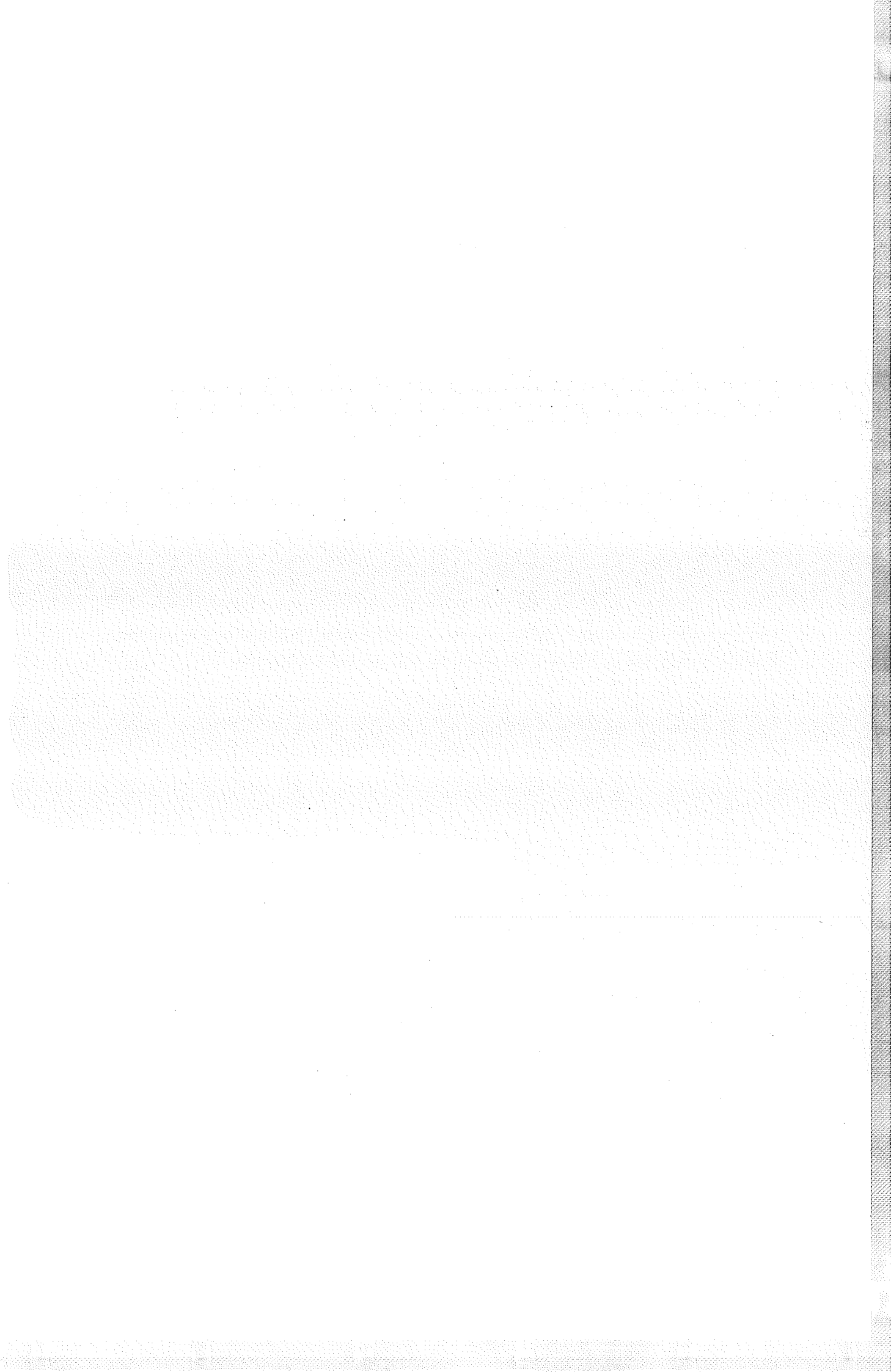
٢- عبد الغنى ، عبد الغنى محمد و آخرون ديسمبر ١٩٩١ " تقييم البيانات السكانية مع التركيز على تعداد ١٩٨٦ " المؤتمر السنوي السادس والعشرون للإحصاء وعلوم الحاسب و بحوث العمليات ، ندوة تقييم البيانات السكانية مع التركيز على تعداد ١٩٨٦ ، معهد الدراسات والبحوث الإحصائية ، جامعة القاهرة

٣- فراج ، عبد المجيد مصطفى ١٩٧٥ " الأسس الإحصائية للدراسات السكانية " دار النهضة العربية ، القاهرة.

٤- مصطفى ، أحمد فتحي ١٩٩٢ " مقدمة في التحليل الديمجرافى " مذكرات غير منشورة.

٦-٢ المراجع الأجنبية :

United Nations, "Methods of Appraisal of Quality of Basic Data for Population Estimate" Manual II, Population Studies, NO. 23, New York, 1955.



التعمر الديموجرافي في مصر
خلال الفترة (١٩٤٧ - ٢٠٠٦)
والتوقعات المستقبلية

مدحت مصطفى أحمد

عبد الغني محمد عبد الغني

١-١: مقدمة:

أصبحت الشيخوخة من أهم الظواهر الديموجرافية في العالم، حيث يتزايد حجم السكان من كبار السن بشكل سريع للغاية، وذلك نتيجة لانخفاض الوفيات والخصوبة الذي حدث في الماضي ويتسبب هذا التراجع في معدلات الوفيات والخصوبة الذي يصاحب البلدان عبر مراحلها الديموجرافية الانتقالية في زيادة نسب كبار السن فيها بمرور الوقت وبصورة لا يمكن تجنبها، ويتوقع الديموجرافيون أن تصبح شيخوخة السكان أو تقدمهم في السن ظاهرة عامة تتمثل في كافة مناطق العالم خلال العقود القليلة القادمة ويعتبر الحفاظ على الصحة والرفاهية النفسية والاقتصادية لهذه المجموعة المتزايدة من سكان العالم النامي تحدياً كبيراً لأكثر من سبب. فمن المعروف أن الصحة تميل إلى الضعف مع كبر السن وكثيراً ما يتوقف كبار السن عن العمل مما يستدعي تقديم مزيج فريد من الدعم الجسدي والمادي لهم.

ولقد مر المجتمع المصري بمرحلة من التغير الاجتماعي والاقتصادي نتيجة لظروف التنمية. فقد شهد زيادة وتضخم في حجم السكان، وانخفاض ملحوظ في معدل الوفيات الخام صاحبه انخفاض في الخصوبة. كما حدث تغير في التركيب العمري للسكان حيث انخفضت نسبة الأطفال أقل من ١٥ سنة واتجهت نسبة المسنين ٦٠ سنة فأكثر إلى التزايد. وهذا يعني أن المجتمع المصري نحو حركة سكانية (تحول ديموجرافي) سيصل بنسبة المسنين فيها تقريباً إلى النسب المرتفعة الموجودة بالدول المتقدمة. ومع ذلك فالمجتمع المصري ما زال في المراحل المبكرة للتعمر الديموجرافي والذي يعنى ارتفاع في نسبة كبار السن بمعدل أكبر من معدل الزيادة في السكان، وتبدو مظاهره في ارتفاع العمر الوسيط للسكان وزيادة توقع الحياة عند الميلاد وخصوصاً عند الأعمار المتقدمة وانخفاض نسبة الذكور للإناث بين المسنين وزيادة عبء إعالة المسنين ٦٠ سنة فأكثر. وهذه الظاهرة تتمثل في صورة زيادة في أعداد المسنين، والتي تتطلب دراسة واستعداد كامل من المسؤولين في الدولة لمواجهة الزيادة المضطردة في أعدادهم في الحاضر والمستقبل، وتخطيط وتنفيذ برامج الرعاية المتكاملة، والتي تتفق مع خصائص وطبيعة المسنين، وإلا فإنه سترتب على ذلك مشاكل اجتماعية ونفسية وصحية وديموجرافية كثيرة.

وعموماً فإن الحكومات والمنظمات والهيئات الأهلية والدولية توجه المزيد من الاهتمام والرعاية للمسنين والذي أبرزته المؤتمرات والتوصيات المختلفة وفقاً لمبادئ الأمم المتحدة، والحوار الفكري المتواصل بين مختلف الدول والأمم جعلنا نستلهم واقع الخصائص والاتجاهات الحديثة للمسنين في المجتمع المصري.

ومما لاشك فيه أن لهذا التغير آثاراً شديدة على الأوضاع الديموجرافية والاجتماعية تتمثل في عبء إعالة الكبار. فينبغي مواجهة هذه الظاهرة بتفهم احتياجات هذه الفئة من المجتمع - فئة المسنين - ودراسة الإمكانيات المتاحة لتحقيق أهداف وخطط التنمية الاقتصادية والاجتماعية بالدولة. لذا يقدم هذا البحث وصفاً لظاهرة من ظواهر المجتمع المصري الذي انتابه الكثير من التغيرات الجذرية.

٢-١: هدف البحث:

يهدف البحث إلى التعرف على حجم واتجاهات التغير، والتعرف على خصائص المسنين الديموجرافية والاجتماعية والاقتصادية، وذلك للمساهمة في التعرف على مشاكلهم والخدمات المطلوب إشباعها وفق هذه الخصائص وطبقاً لاحتياجات المسنين لها، مما يفيد في تخطيط وتطوير تلك الخدمات مستقبلاً. لذلك تتلخص أهداف البحث في النقاط التالية:

١. التعرف على بعض مقاييس التغير.
٢. تطور حجم ونمط الفئات العمرية للمسنين في مصر.
٣. أهم الخصائص الديموجرافية والاجتماعية والاقتصادية التي يتسم بها المسنون.
٤. التوقعات المستقبلية لفئة المسنين في مصر.

٣-١: المنهجية المستخدمة:

المنهجية المتبعة هي استخدام منهج التحليل الوصفي للتعرف على اتجاهات ونمط التغير في مصر (كبار السن ٦٠ سنة فأكثر) وذلك من خلال دراسة التركيب العمري للسكان، وكذلك التعرف على مقاييس التغير من خلال أهم التعدادات التي تمت، وكذلك الخصائص الديموجرافية والاجتماعية للمسنين خلال الفترة الزمنية التي يشملها البحث. كما تم استخدام برنامج Spectrum لتقدير أعداد السكان، وأعداد المسنين في مصر حتى عام ٢٠٣١، اعتماداً على بيانات نتائج تعداد ٢٠٠٦.

أما عن الفترة الزمنية التي يشملها البحث هي الفترة الزمنية (١٩٤٧-٢٠٠٦). وذلك لتوافر البيانات التعدادية عن هذه الفترة التي تتيح دراسة واقع هذه الفئة العمرية في مصر. كما يشمل البحث توقعات الفترة الزمنية المستقبلية حتى عام ٢٠٥٠.

٤-١: مصادر البيانات:

لتحقيق هدف البحث تم الاعتماد على تعداد السكان في الفترة (١٩٤٧-٢٠٠٦). كذلك إسقاطات الأمم المتحدة (UN, The 2008 Revision) وذلك للاستعانة بتوقعاتها المستقبلية للسكان.

٥-١: مفاهيم البحث:

١. التمر:

ظاهرة ديناميكية متحركة في ازدياد مستمر، وتعرف بأنها عملية التحول من المجتمع الفتي إلى المجتمع الهرم، ويقصد بها التغيرات الجوهرية التي تحدث في التركيب العمري للسكان مما يؤدي إلى ارتفاع نسبة السكان في الفئات العمرية الكبيرة (٦٥ فأكثر) وفي بعض المجتمعات (٦٠ فأكثر)، ويحدث هذا التغير نتيجة التحول الديموجرافي في المجتمع - حسب نظرية الانتقال الديموجرافي - من خصوبة عالية إلى خصوبة منخفضة، كما أن تحسن الرعاية الصحية أدت إلى ارتفاع احتمال البقاء على قيد الحياة وتعكسه الزيادة في أعداد كبار السن. وقد اصطلح الديموجرافيون على اعتبار أن المجتمعات التي تقل فيها نسبة كبار السن عن ٥% هو مجتمع فتي، والمجتمعات التي تتراوح فيها نسبة كبار السن بين ٥-١٠% تعيش مرحلة انتقالية، بينما المجتمعات التي تزيد فيها نسبة كبار السن على ١٠% فهي في المرحلة الأخيرة (النضج) أو ما يسمى بالتممر.

٢. تعريف المسن:

اختلفت الآراء في تعريف المسن، وهي في اللغة تعني الرجل الكبير، وفي الاصطلاح هو كل فرد أصبح عاجز عن رعاية وخدمة نفسه إثر تقدمه في العمر، وليس بسبب إعاقة. وعموماً المسن يعرف طبقاً لمبادئ وتوصيات الأمم المتحدة لتعدادات السكان والإسكان بأن المسن هو كل شخص عمره ٦٠ عاماً فأكثر. وتأخذ الدراسة بهذا التعريف على الرغم من وجود عدة تعريفات للمسن تختلف من مجتمع لآخر ففي الدول المتقدمة والمقارنات الدولية يكون المسن هو الشخص الذي عمره ٦٥ عاماً فأكثر. ويعتبر عمر التقاعد في مصر هو سن الستين عاماً

لأغلبية الفئات وبداية مرحلة التعمر (القانون رقم ٧٩ لسنة ١٩٧٥، قانون التأمين الاجتماعي) فيما عدا رجال السلطة القضائية والجامعات وخريجي الثانوية الأزهرية والتي يزيد فيها عمر التقاعد إلى ٦٥ عاماً.

٦-١: الدراسات السابقة:

توجد دراسات كثيرة على المستوى الدولي، والعربي، والقومي تعرضت لظاهرة التعمر السكاني. نستخلص منها أن التغير في التركيب العمري للسكان يعد ظاهرة حديثة نسبياً لم تؤثر بعد في بلاد كثيرة فجميع البلاد النامية ذات تركيبات عمرية متشابهة ومن ثم فإن تعمير المجموعات السكانية في قمة الهرم السكاني ظاهرة تتميز بها الدول المتقدمة. وترجع أسباب هذا التغير لانخفاض الخصوبة والوفيات. فيتوقع تقرير المؤتمر الدولي الثاني للشيخوخة للأمم المتحدة (٢٠٠٢) أن يفوق عدد كبار السن على مستوى العالم عدد الأطفال لأول مرة في عام ٢٠٤٧.

كما أن أغلبية المسنين في البلدان النامية يميلون إلى العيش في أسر كبيرة، مع تزايد احتمالات أن تكون إقامتهم مع ابن بالغ، وغالباً ما يكون ذكراً. وإن كبار السن الذين يعيشون بمفردهم أكثر عرضة للعزلة الاجتماعية والحرمان الاقتصادي، ولذلك قد يحتاجون إلى دعم خاص. وأن الغالبية من المسنين في كل بلدان إقليم شرق المتوسط تعيش في مناطق ريفية. فالهجرة من الأرياف إلى المدن هي العامل الرئيسي الذي يحدد ظاهرة شيخوخة السكان في المناطق الريفية، ويتوقع أن يؤدي استمرار هذا التحول من الأرياف إلى المدن إلى ضغط هائل على البنية التحتية في المدن وغيرها من المناطق الحضرية. كما أن الهجرة السريعة من المناطق الريفية إلى الحضرية ستغير في السنوات المقبلة الوضع الحالي، نظراً لأن الذين هاجروا حديثاً للمدن سيتقدمون في العمر فإن تركّز المسنين سينزاح تدريجياً من المناطق الريفية إلى المدنية كما حدث في البلدان المتقدمة. كما يوجد تباين بين المدن فيما يتعلق بنسبة المسنين، وخصائصهم والخدمات المقدمة إليهم.

كما يختلف نمط الحالة الزوجية لدى الرجال المسنين عن النساء المسنات، حيث يشكل النساء الشريحة الأكبر من كبار السن، وأن أغلبهن أرمال، كما أظهرت الدراسات تدنى المستويات التعليمية للمسنين، فنسبة الأمية مرتفعة بين المسنين أغلبهم من النساء. كما أن المسنين ذوو المستويات التعليمية الأعلى لديهم قدرة أكبر على العناية بأنفسهم لأنهم يتمتعون بصحة أفضل، ولديهم قدرة أكبر في الحصول على دخل.

وتبين الدراسات أيضاً القصور في الخدمات الطبية ونظام الرعاية الاجتماعية المقدمة للمسنين بسبب النقص الواضح في عدد الأخصائيين سواء الأطباء أو النفسيين أو الاجتماعيين العاملين في مجال تقديم الخدمات الطبية والإرشادية والنفسية والاجتماعية للمسنين. كما أن العاملون في الرعاية الصحية الأولية إما لم يتدربوا على الرعاية الصحية للمسنين أو أنهم لا يستطيعون تمييز الاحتياجات الخاصة لهذا القطاع من السكان. كما أن هناك فئة خاصة من المسنين تحتاج إلى مزيد من الرعاية والاهتمام، وهي تشمل المسنات اللاتي لم يسبق لهن الزواج، والمطلقات والأرامل، هذا بالإضافة إلى أن هذه الفئة تحتاج إلى إعالة كاملة بصفة خاصة إذا كن لا يتقاضين أي معاش أو ليس لديهن مورد ثابت للرزق.

وعلى الرغم من أن التعمر الديموجرافي في مصر لا يزال في مراحله المبكرة، إلا أنه سيلقي بأعباء على المجتمع، وستلحق ظاهرة التعمر ببعض الظواهر الاقتصادية منها: فقدان القوة الإنتاجية في المجتمع لخبرات وطاقات المسنين، وارتفاع معدل الإعالة مما يؤدي إلى انخفاض المدخرات والاستثمارات، وإضعاف عملية التنمية نتيجة إبعاد هذه الشريحة الهامة من أفراد المجتمع. كما يترتب على التقاعد تدنى مستوى المعيشة نتيجة نقص الدخل، كما يترتب عليه أيضاً الاكتئاب نتيجة المرض والعجز والإحساس بتدني القيمة، وضمور العلاقات الاجتماعية.

٢. ظاهرة التعمر الديموجرافي في مصر

٢-١: تطور حجم المسنين:

يعرض الجدول رقم (٢-١) التوزيع العددي والنسبي لسكان مصر في التعدادات التي أجريت في مصر خلال الفترة من ١٩٤٧ إلى ٢٠٠٦. وفي هذا الجدول نجد أن العدد الكلي للسكان قد زاد خلال الفترة من ١٨,٩٦ مليون نسمة في تعداد ١٩٤٧ إلى ٧٢,٦١ مليون في تعداد ٢٠٠٦. أي بزيادة قدرها ٢٨٣% تقريباً. ومن ناحية أخرى، صاحب ذلك زيادة في عدد المسنين (٦٠ سنة فأكثر) فقد زاد العدد من ١,١٤ مليون إلى ٤,٤١ مليون مسن خلال تلك الفترة أي أن عدد المسنين زاد بمقدار ٢٨٧% تقريباً عن عام ١٩٤٧. كما نجد في الجدول أن هناك انخفاض في المجموعة الأولى من السكان — فئة الأطفال (٠-١٤) — وهي فئة التكوين والنمو. فقد كانت نسبة الفئة الأولى ٤٠% عام ١٩٨٦ انخفضت هذه النسبة انخفاض ملحوظ حوالي ٨% في عام ٢٠٠٦ فكانت نسبة هذه الفئة ٣١,٧% من إجمالي السكان، وقد يعزى هذا الانخفاض إلى الانخفاض في معدلات المواليد. وعموماً كان هذا الانخفاض لصالح المجموعتين

وتبين الدراسات أيضاً القصور في الخدمات الطبية ونظام الرعاية الاجتماعية المقدمة للمسنين بسبب النقص الواضح في عدد الأخصائيين سواء الأطباء أو النفسيين أو الاجتماعيين العاملين في مجال تقديم الخدمات الطبية والإرشادية والنفسية والاجتماعية للمسنين. كما أن العاملون في الرعاية الصحية الأولية إما لم يتدربوا على الرعاية الصحية للمسنين أو أنهم لا يستطيعون تمييز الاحتياجات الخاصة لهذا القطاع من السكان. كما أن هناك فئة خاصة من المسنين تحتاج إلى مزيد من الرعاية والاهتمام، وهي تشمل المسنات اللاتي لم يسبق لهن الزواج، والمطلقات والأرامل، هذا بالإضافة إلى أن هذه الفئة تحتاج إلى إعالة كاملة بصفة خاصة إذا كن لا يتقاضين أي معاش أو ليس لديهن مورد ثابت للرزق.

وعلى الرغم من أن التمر الديموجرافي في مصر لا يزال في مراحله المبكرة، إلا أنه سيلقي بأعباء على المجتمع، وستلحق ظاهرة التمر ببعض الظواهر الاقتصادية منها: فقدان القوة الإنتاجية في المجتمع لخبرات وطاقات المسنين، وارتفاع معدل الإعالة مما يؤدي إلى انخفاض المدخرات والاستثمارات، وإضعاف عملية التنمية نتيجة إبعاد هذه الشريحة الهامة من أفراد المجتمع. كما يترتب على التقاعد تدنى مستوى المعيشة نتيجة نقص الدخل، كما يترتب عليه أيضاً الاكتئاب نتيجة المرض والعجز والإحساس بتدني القيمة، وضمور العلاقات الاجتماعية.

٢. ظاهرة التمر الديموجرافي في مصر

٢-١: تطور حجم المسنين:

يعرض الجدول رقم (٢-١) التوزيع العددي والنسبي لسكان مصر في التعدادات التي أجريت في مصر خلال الفترة من ١٩٤٧ إلى ٢٠٠٦. وفي هذا الجدول نجد أن العدد الكلي للسكان قد زاد خلال الفترة من ١٨,٩٦ مليون نسمة في تعداد ١٩٤٧ إلى ٧٢,٦١ مليون في تعداد ٢٠٠٦. أي بزيادة قدرها ٢٨٣% تقريباً. ومن ناحية أخرى، صاحب ذلك زيادة في عدد المسنين (٦٠ سنة فأكثر) فقد زاد العدد من ١,١٤ مليون إلى ٤,٤١ مليون مسن خلال تلك الفترة أي أن عدد المسنين زاد بمقدار ٢٨٧% تقريباً عن عام ١٩٤٧. كما نجد في الجدول أن هناك انخفاض في المجموعة الأولى من السكان — فئة الأطفال (٠-١٤) — وهي فئة التكوين والنمو. فقد كانت نسبة الفئة الأولى ٤٠% عام ١٩٨٦ انخفضت هذه النسبة انخفاض ملحوظ حوالي ٨% في عام ٢٠٠٦ فكانت نسبة هذه الفئة ٣١,٧% من إجمالي السكان، وقد يعزي هذا الانخفاض إلى الانخفاض في معدلات المواليد. وعموماً كان هذا الانخفاض لصالح المجموعتين

الأخيرتين — فئة السكان في عمر النشاط (١٥-٥٩ سنة)، وفئة كبار السن (٦٠ سنة فأكثر) — فنجد زيادة ملحوظة في نسبة الفئة الثانية من السكان حوالي ٧,٨% بين عامي ١٩٨٦، ٢٠٠٦. فقد ارتفعت النسبة من ٥٤,٤% عام ١٩٨٦، إلى ٦٢,٢%. وقد يرجع ذلك إلى دخول سكان فترة طفرة المواليد التي شوهدت في فترة ما بعد الحرب.

جدول (١-٢)

التوزيع العددي والنسبي للسكان حسب فئات عمرية عريضة في تعدادات مصر خلال الفترة ١٩٤٧ - ٢٠٠٦

سنة التعداد	عدد السكان			عدد السكان الكلية	النسبة المئوية		
	14-0	59-15	+60		14-0	59-15	+60
1947	7220517	10605684	1140569	18966770	38.07	55.92	6.01
1960	11110019	13296648	1577434	25984101	42.76	51.17	6.07
1966	12645524	15374266	1827347	29847137	42.37	51.51	6.12
1976	14629395	19709493	2287316	36626204	39.94	53.81	6.25
1986	19281187	26246031	2727020	48254238	39.96	54.39	5.65
1996	22345496	33554721	3412697	59312914	37.67	56.57	5.75
2006	23053263	45145873	4414216	72613352	31.75	62.17	6.08

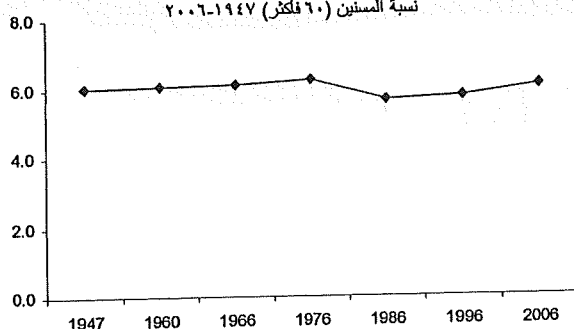
المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٤٧ - ٢٠٠٦)

أما الفئة الأخيرة — فئة كبار السن (٦٠ سنة فأكثر) — فنجد نسب الزيادة طفيفة جداً حيث نجد أن النسبة ارتفعت من ٥,٧% عام ١٩٨٦، إلى ٦,١% تقريباً عام ٢٠٠٦. وبالرغم من ذلك نجد أن هذه الزيادة في نسبة كبار السن قد تذبذبت خلال الفترات التعدادية (١٩٤٧ - ٢٠٠٦)

حول ٦%، حيث يظهر ذلك من الشكل رقم (١-٢)، وهذه النسبة تقع ضمن نطاق المرحلة الانتقالية. حيث أن الديموجرافيين قد اصطاحوا على اعتبار أن المجتمعات التي نقل فيها نسبة كبار السن عن ٥% هو مجتمع فتى، والمجتمعات التي تتراوح فيها نسبة كبار السن بين ٥-١٠% تعيش مرحلة انتقالية، بينما المجتمعات التي تزيد فيها نسبة كبار السن على ١٠% هي مرحلة التمر.

شكل رقم (١-٢)

نسبة المسنين (٦٠ فأكثر) ١٩٤٧-٢٠٠٦



المصدر: محسوب بمعرفة الباحث من التعدادات خلال الفترة

(١٩٤٧-٢٠٠٦)

٢-٢: التركيب العمري والنوعي للمسنين:

عند دراسة التركيب العمري للمسنين في مصر للتعرف على التفاوت في اتجاه التعمر، فمن

خلال شكل (٢-٢) يمكننا التعرف على

هذا التفاوت في نسبة المسنين، حيث

نلاحظ انخفاض نسبة المسنين في الفئة

الأولى لهم (٦٠-٦٤ سنة)، وكذلك الفئة

الثالثة (٧٠-٧٤ سنة). بينما نجد أن

الفئة الثانية للمسنين (٦٥-٦٩ سنة) في

زيادة واضحة، كما نجد أن الفئة

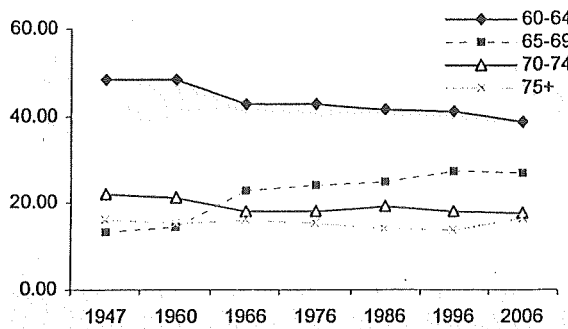
الأخيرة للمسنين (٧٥ فأكثر) تتذبذب

بين الارتفاع تارة وبين الانخفاض تارة

أخرى.

شكل رقم (٢-٢)

اتجاه التركيب العمري لفئات المسنين في مصر (١٩٤٧-٢٠٠٦)



المصدر: محسوب بمعرفة الباحث من التعدادات خلال الفترة

(١٩٤٧-٢٠٠٦)

كما أننا نجد من جدول (٢-٢) أن نسبة الفئة الأولى كانت ٤٨,٨% عام ١٩٦٠ من إجمالي المسنين، انخفضت إلى ٤٢,٩% عام ١٩٦٦، ثم إلى ٤١,٥% عام ١٩٨٦، وانخفضت انخفاض طفيف جداً عام ١٩٩٦ فكانت نسبة هذه الفئة هي ٤١%. وكانت عام ٢٠٠٦ هذه النسبة ٣٨,٥% من إجمالي المسنين في مصر. وفي المقابل نلاحظ ارتفاع في نسب الفئة الثانية من فئات المسنين (٦٥-٦٩ سنة) من ١٣,٣% في عام ١٩٤٧ إلى ٢٣% عام ١٩٦٦. ثم ارتفعت إلى ٢٥% عام ١٩٨٦، إلى أن بلغت ٢٧% عام ٢٠٠٦. كما نجد من الجدول أن غالبية المسنين تقع في الفئة العمرية الأولى (٦٠-٦٤ سنة) حيث تخطت نصيب هذه الفئة ٤٠% في التعدادات التي تمت في مصر بين عام ١٩٤٧، وعام ١٩٩٦، بينما انخفضت قليلاً إلى ٣٨,٥% عام ٢٠٠٦.

جدول (٢-٢)

التوزيع النسبي للمسنين حسب فئات العمر والنوع خلال الفترة (١٩٤٧-٢٠٠٦)

الفئات العمرية	1947			الفئات العمرية	1960		
	ذكور	إناث	الجملة		ذكور	إناث	الجملة
-٦٠	48.55	48.36	48.44	-٦٠	43.40	42.21	42.77
-٦٥	16.14	13.29	14.59	-٦٥	22.17	20.22	21.13
-٧٠	20.76	22.13	21.51	-٧٠	18.07	20.01	19.10
+٧٥	14.55	16.22	15.46	+٧٥	16.35	17.56	16.99
المجموع	45.64	54.36	100.00	المجموع	46.85	53.15	100.00

تابع جدول (٢-٢)

التوزيع النسبي للمسنين حسب فئات العمر والنوع خلال الفترة (١٩٤٧-٢٠٠٦)

1976			الفئات
الجملة	إناث	ذكور	العمرية
42.42	42.07	42.79	-٦٠
22.35	20.80	23.96	-٦٥
19.12	20.15	18.05	-٧٠
16.11	16.98	15.20	+٧٥
100.00	50.96	49.04	المجموع

1966			الفئات
الجملة	إناث	ذكور	العمرية
42.86	43.42	42.23	-٦٠
22.95	20.73	25.42	-٦٥
18.04	18.75	17.26	-٧٠
16.15	17.10	15.09	+٧٥
100.00	52.59	47.41	المجموع

1996			الفئات
الجملة	إناث	ذكور	العمرية
40.99	42.17	39.90	-٦٠
27.29	25.78	28.68	-٦٥
18.10	18.38	17.84	-٧٠
13.62	13.67	13.58	+٧٥
100.00	48.14	51.86	المجموع

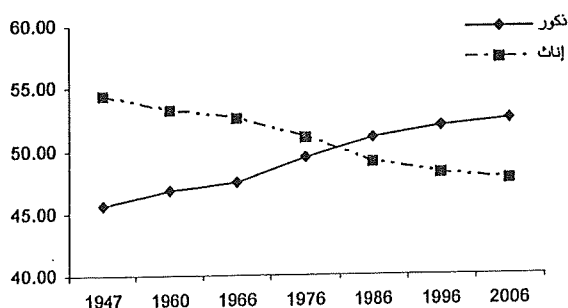
1986			الفئات
الجملة	إناث	ذكور	العمرية
41.48	43.23	39.80	-٦٠
24.98	24.73	25.22	-٦٥
19.41	17.74	21.02	-٧٠
14.13	14.30	13.96	+٧٥
100.00	48.96	51.04	المجموع

2006			الفئات
الجملة	إناث	ذكور	العمرية
38.52	38.25	38.76	-٦٠
26.96	26.22	27.64	-٦٥
17.84	18.38	17.36	-٧٠
16.68	17.15	16.25	+٧٥
100.00	47.63	52.37	المجموع

المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٤٧-٢٠٠٦)

شكل رقم (٢-٢)

اتجاه التركيب النوعي للمسنين خلال الفترة (١٩٤٧-٢٠٠٦)



المصدر: محسوب بمعرفة الباحث من التعدادات خلال الفترة

(١٩٤٧-٢٠٠٦)

أما عن التركيب النوعي لفئات المسنين فيلاحظ من شكل (٢-٢ب)، أن نسبة الإناث المسنات كانت أكبر من نسبة المسنين الذكور وذلك في التعدادات من (١٩٤٧-١٩٧٦)، ثم على النقيض في التعدادات (١٩٨٦-٢٠٠٦) فكانت نسبة المسنين الذكور أكبر من نسبة الإناث المسنات.

كما يلاحظ من الجدول السابق أيضاً أن نسبة الإناث المسنات ٥٤,٤% عام ١٩٤٧، قد انخفضت إلى ٥١% عام ١٩٧٦ ثم إلى ٤٧,٦% عام ٢٠٠٦. في المقابل نجد زيادة في نسب الذكور المسنين فمن ٤٥,٦% عام ١٩٧٦، إلى ٤٩% عام ١٩٧٦، ثم إلى ٥٢,٤% عام ٢٠٠٦. كما يلاحظ أيضاً أن نسبة الإناث المسنات ٧٠ سنة فأكثر أكبر من نسبة الذكور المسنين في نفس الفئة.

٢-٣: توزيع المسنين حسب محل الإقامة:

وعند دراسة المسنين وفقاً للإقامة (حضر - ريف) في الفترة ١٩٦٠ وحتى ٢٠٠٦، يتضح من جدول (٢-١٣) أن نسبة إجمالي المسنين في الريف أعلى من الحضر حيث يوجد تباين واضح بين الريف والحضر. فيظهر ذلك التباين بشكل كبير في تعداد ١٩٦٠ حيث بلغت نسبة المسنين الذين يقيمون في الحضر ٣٢% في مقابل الضعف تقريباً في الريف حيث كانت نسبة المسنين ٦٨%. أي أن الفجوة بين الحضر والريف حوالي ٣٦% لصالح الريف.

كما يلاحظ من الجدول أيضاً أن هذه الفجوة في انخفاض مستمر حيث يظل التباين كبير حتى عام ١٩٨٦ فكانت هذه الفجوة حوالي ١٤% حيث كانت نسبة المسنين في الحضر ٤٢,٨% مقابل ٥٧,٢% في الريف، وتقلصت هذه النسبة إلى ١٠% عام ١٩٩٦ فكان نسبة سكان الحضر ٤٥,٠% مقابل ٥٥,٠% للريف. إلى أن أصبح الفرق بين الحضر والريف حوالي ٥% تقريباً لصالح الريف، فكانت نسبة المسنين عام ٢٠٠٦ في الحضر مقابل الريف هي ٤٧,٧%، ٥٢,٣% على التوالي. قد يرجع السبب في هذا الانخفاض إلى عدم عودة أعداد كبيرة من المهاجرين المحالين على المعاش والذين استقروا بأسرهم في المناطق الحضرية إلى موطنهم الأصلي كما كان يحدث في الماضي (مخلوف، ٢٠٠٣).

جدول (٢-١٣)

التوزيع النسبي للمسنين (٦٠ سنة فأكثر) في مصر حسب محل الإقامة (حضر - ريف) والنوع خلال الفترة (١٩٦٠-٢٠٠٦)

سنة التعداد	حضر			ريف		
	ذكور	إناث	جملة	ذكور	إناث	جملة
1960	50.1	49.9	32.0	45.3	54.7	68.0
1976	51.8	48.2	38.9	47.3	52.7	61.1
1986	53.5	46.5	42.8	49.3	50.7	57.2
1996	54.2	45.8	45.0	49.9	50.1	55.0
2006	54.7	45.3	47.7	50.3	49.7	52.3

المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٦٠-٢٠٠٦)

وعند دراسة النوع في كل من الحضر والريف يلاحظ من الجدول أن نسبة الذكور المسنين في الحضر أكبر من نسبة الإناث المسنات في جميع سنوات التعداد من ١٩٦٠ وحتى ٢٠٠٦، وعلى العكس من ذلك نجد أن نسبة الإناث المسنات في الريف أعلى من الذكور المسنين ولكن هذه النسبة تتناقص للإناث وتزايد للذكور فنجد أن نسبة الإناث ٥٤,٧% مقابل ٤٥,٣% في عام ١٩٦٠، وتظل نسبة الإناث أعلى من الذكور حتى عام ١٩٩٦ فنجد أن نسبة الإناث والذكور في الريف كانت ٥٠,١%، ٤٩,٩% على التوالي. وفي عام ٢٠٠٦ كانت نسبة الذكور أعلى من نسبة الإناث حيث كانت ٥٠,٣% مقابل ٤٩,٧% على التوالي.

وعند تناول توزيع المسنين حسب محل الإقامة (حسب المنطقة) يلاحظ من جدول (٢-٣) أن

جدول (٢-٣)

التوزيع النسبي للمسنين حسب محل الإقامة (حسب الأقاليم)
خلال الفترة (١٩٨٦-٢٠٠٦)

محل الإقامة	سنوات التعداد		
	١٩٨٦	١٩٩٦	٢٠٠٦
المحافظات الحضرية	20.8	21.8	21.4
محافظات الوجه البحري	40.5	40.6	45.0
محافظات الوجه القبلي	37.9	36.7	32.5
محافظات الحدود	0.9	0.9	1.1
الإجمالي	100.0	100.0	100.0

المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٨٦-٢٠٠٦)

نسبة المسنين في الوجه البحري تتعدى الـ ٤٠% من جملة المسنين في مصر، وذلك في تعداد ١٩٨٦، ١٩٩٦، بينما تصل نسبة المسنين في تعداد ٢٠٠٦ إلى ٤٥%. يلي ذلك محافظات الوجه القبلي ويظهر من الجدول أن هذه النسب تميل إلى الانخفاض فمن ٣٨% عام ١٩٨٦ تقل إلى ٣٢,٥% عام ٢٠٠٦. كما يلاحظ أن نسبة

المسنين في المحافظات الحضرية وحدها خمس المسنين تقريباً، وقد يرجع ذلك لوجود أعداد كبيرة من المهاجرين من محافظات الوجه القبلي إلى القاهرة والإسكندرية ومدن القناة بحثاً عن العمل، وعدم عودة هذه الأعداد من المحالين على المعاش الذين استقروا بالمناطق الجديدة بأسرهم إلى موطنهم الأصلي. كما يتضح من الجدول أيضاً أن نسبة المسنين في محافظات الحدود أقل ما يمكن حيث نجد نسبة المسنين فيها ١% من جملة المسنين في مصر خلال الفترة (١٩٨٦-٢٠٠٦).

وعلى مستوى المحافظات وجد في تعداد ٢٠٠٦ أن نسبة المسنين في القاهرة وحدها حوالي ١٣% تقريباً، تليها محافظة الدقهلية، والإسكندرية كلاهما حوالي ٧% تقريباً.

٢-٤: الحالة الزوجية للمسنين:

يوضح جدول (٢-٤) التوزيع النسبي للمسنين حسب الحالة الزوجية في تعداد عامي ١٩٩٦، ٢٠٠٦، ويتبين منه أن الغالبية العظمى من كبار السن الذكور كانوا متزوجين، بينما العكس في الإناث المسنات فأغلبهن من المترملات. ففي الفئة الأولى (٦٠-٦٤) للذكور يلاحظ أن نسبة المتزوجين حوالي ٩٢% في كل من تعداد ١٩٩٦، ٢٠٠٦، كذلك نجد أن نسبة الإناث المتزوجات في هذه الفئة هي ٤٦%، ٥١% في عامي ١٩٩٦، ٢٠٠٦ على التوالي. بينما يلاحظ أن نسبة المترملين الذكور في نفس الفئة العمرية هي ٧%، وكانت نسبة الإناث المترملات ٥٢% عام ١٩٩٦، بينما في عام ٢٠٠٦ كانت نسبة المترملين الذكور ٧%، والإناث المترملات ٤٦% لنفس الفئة العمرية.

جدول (٢-٤)

التوزيع النسبي للمسنين ٦٠ فأكثر حسب الحالة الزوجية والنوع في تعداد ١٩٩٦، ٢٠٠٦

الحالة الزوجية والنوع	تعداد ١٩٩٦				تعداد ٢٠٠٦			
	٦٠-	٦٥-	٧٠-	٧٥+	٦٠-	٦٥-	٧٠-	٧٥+
لم يتزوج ذكور	0.7	0.7	0.7	0.7	0.8	0.8	0.8	1.0
إناث	1.0	1.0	1.0	1.1	1.1	1.1	1.1	1.2
جملة	0.9	0.8	0.9	0.9	1.0	0.9	0.9	1.1
متزوج ذكور	92.2	88.7	82.9	74.1	91.6	88.2	83.0	74.2
إناث	45.6	32.9	18.6	10.4	51.4	35.7	22.9	12.6
جملة	69.1	63.3	51.5	43.3	72.6	63.9	53.5	44.0
مطلق ذكور	0.4	0.4	0.4	0.4	0.5	0.4	0.4	0.4
إناث	1.1	0.9	0.8	0.6	1.1	1.0	0.7	0.6
جملة	0.7	0.6	0.6	0.5	0.8	0.7	0.6	0.5
أرمل ذكور	6.7	10.2	16.0	24.9	7.1	10.6	15.9	24.4
إناث	52.3	65.1	79.6	87.9	46.3	62.1	75.3	85.6
جملة	29.3	35.2	47.1	55.3	25.6	34.5	45.0	54.4
الإجمالي جملة	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

المصدر: محسوب بمعرفة الباحث من تعداد الجهاز المركزي للتعبئة العامة والإحصاء عامي ١٩٩٦، ٢٠٠٦.

كذلك الحال بالنسبة للفئات العمرية الأخرى، ولكن مع انخفاض مستمر في نسبة المتزوجين من الذكور والإناث وارتفاع في نسبة المترملين، ولكن الانخفاض بالنسبة للإناث المتزوجات كان أكبر من الذكور بشكل ملحوظ يقابله زيادة كبيرة في نسبة المترملات الإناث عن نسبة المترملين الذكور، ذلك قد يرجع إلى أن الذكور المسنين قد يكون لديهم القدرة والإمكانية

للزواج مرة أخرى عن الإناث وما يؤكد ذلك أن نسبة الإناث المسنات المطلقات في الجدول السابق أعلى من المسنين الذكور المطلقين في جميع فئات المسنين العمرية الذكور منهم والإناث. وكذلك من إحصاءات الزواج والطلاق لعام ٢٠٠٦ كان عدد عقود الزواج التي تمت عام ٢٠٠٦ لكبار السن كانت بنسبة (٧ : ١) تقريباً بين الذكور المسنين والإناث المسنات، حيث كان عدد عقود الزواج للمسنين الذكور ٧٣٦٢ مقابل ١١٨٨ عقد زواج للإناث المسنات. كما يلاحظ من الجدول أيضاً أن نسبة المسنين الذين لم يتزوجوا بسيطة جداً لم تتعد شخص واحد لكل مائة شخص.

٥-٢: الحالة التعليمية للمسنين:

يوضح جدول (٥-٢) التوزيع النسبي للمسنين ٦٠ سنة فأكثر وفقاً للحالة التعليمية والنوع، والذي يتبين منه أن الغالبية العظمى للمسنين أميين، ولكن نسبة هذه الشريحة الأمية تميل إلى الانخفاض. فلاحظ أن نسبة الأمية للمسنين قد انخفضت من ٧٨,٩% في عام ١٩٧٦ إلى ٦٨,٨% في عام ٢٠٠٦. كما يلاحظ أن الانخفاض في نسبة الأمية لدى المسنين كان لصالح من حصل على التعليم فارتفع الحاصلون على شهادة أقل من متوسط، حيث كانت ٢,١% عام ١٩٧٦ ارتفعت إلى ٥,٧% عام ٢٠٠٦، كما ارتفع نسبة المسنون الحاصلين على شهادة متوسطة من ٢,٣% عام ١٩٧٦، إلى ٧,٨% عام ٢٠٠٦، كما تضاعفت فئة المسنين الأكثر تعليماً (الجامعي فأعلى) لتصل إلى ٦,٥% في عام ٢٠٠٦، مقارنة بـ ١,١% عام ١٩٧٦.

جدول (٥-٢)

التوزيع النسبي للمسنين ٦٠ سنة فأكثر حسب الحالة التعليمية والنوع خلال الفترة (١٩٧٦-٢٠٠٦)

الحالة التعليمية	1976			1986			1996			2006		
	ذكور	إناث	جملة	ذكور	إناث	جملة	ذكور	إناث	جملة	ذكور	إناث	جملة
أمي	65.5	91.7	78.9	68.9	92.8	80.5	44.5	76.9	59.4	56.1	82.7	68.8
يقرأ ويكتب	26.4	4.0	15.0	21.2	5.2	13.4	30.2	14.1	22.8	15.7	6.1	11.1
ابتدائي/إعدادي	3.4	0.9	2.1	2.4	0.9	1.6	6.3	2.9	4.8	7.6	3.7	5.7
متوسط/فوق متوسط	4.0	0.6	2.3	4.4	0.9	2.7	8.8	4.1	6.6	10.7	4.6	7.8
جامعي فأعلى	2.2	0.1	1.1	3.1	0.2	1.7	10.2	1.9	6.4	9.9	2.9	6.5

المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٧٦-٢٠٠٦)

كما تشير نفس البيانات إلى ارتفاع نسبة الأمية بين الإناث المسنات عن نظيرتها بين الذكور المسنين، حيث ٩١,٧% من الإناث المسنات عام ١٩٧٦ أميات، وانخفضت هذه النسبة عام ٢٠٠٦ إلى ٨٢,٧% وذلك مقابل ٦٥,٥%، ٥٦,١% للذكور المسنين عام ١٩٧٦، ٢٠٠٦ على

التوالي. كما يلاحظ من الجدول أيضاً أن الإناث أقل حظاً من الذكور في جميع مستويات التعليم في جميع التعدادات، فعلى سبيل المثال عام ٢٠٠٦، يلاحظ أن نسبة الحاصلات على تعليم أقل من متوسط كانت ٤% مقابل ٨% للذكور، وكانت نسبة الحاصلات على تعليم متوسط أو فوق متوسط ٥% مقابل ١١% للذكور، بينما يلاحظ أن المسنات الحاصلات على تعليم جامعي كن ٣% مقابل ١٠% من الذكور المسنين.

٢-٦: مساهمة المسنين في قوة العمل:

ويوضح جدول (٢-٦) نسب السكان المسنين المساهمين في قوة العمل وفقاً للنوع والفئات العمرية، يلاحظ من الجدول وجود ميل إلى التناقص في نسبة المسنين المساهمين في قوة العمل، حيث انخفضت من ٦,٧% عام ١٩٧٦ إلى ٤,٧% عام ١٩٨٦، ووصلت إلى ٤,٠% عام ١٩٩٦، وأخيراً عام ٢٠٠٦ انخفضت إلى ٣,٢%. كذلك نجد أن حوالي ٦٠% من القوى العاملة للمسنين تقع في الفئة الأولى (٦٠-٦٤ سنة) والذين يمثلون حديثاً إلى المعاش، وهم عادة ما تكون حالتهم أفضل كثيراً من الفئات التي تليها، مما يتيح لهم العمل (مخلف، ٢٠٠٣). كما يلاحظ من الجدول أيضاً ارتفاع نسبة العاملين الذكور المسنين عن الإناث المسنات في جملة القوى العاملة من المسنين.

جدول رقم (٢-٦)

التوزيع النسبي للمسنين المساهمين في قوة العمل حسب فئات السن والنوع خلال الفترة (١٩٧٦-٢٠٠٦)

السن	١٩٧٦			١٩٨٦			١٩٩٦			٢٠٠٦		
	ذكور	إناث	جملة	ذكور	إناث	جملة	ذكور	إناث	جملة	ذكور	إناث	جملة
٦٤-٦٠	٥٨,٤	٦٠,٢	٥٨,٥	٦٣,١	٦٢,٩	٦٣,١	٥٧,٢	٥٩,٧	٥٧,٣	٥٥,٢	٧٨,٣	٥٧,١
٦٩-٦٥	٢١,٣	١٧,٣	٢١,١	٢٤,٣	٢٣,٢	٢٤,٢	٢٦,٦	٢٣,٩	٢٦,٥	٢٦,٧	١٢,٦	٢٥,٥
+٧٠	٢٠,٣	٢٢,٤	٢٠,٤	١٢,٦	١٣,٩	١٢,٧	١٦,٢	١٦,٤	١٦,٢	١٨,١	٩,١	١٧,٤
نسبة المشتغلين	٧,٠	٢,٥	٦,٧	٥,١	٠,٩	٤,٧	٤,٦	٠,٢	٤,٠	٣,٦	١,٥	٣,٢

المصدر: محسوب بمعرفة الباحث من تعدادات الجهاز المركزي للتعبئة العامة والإحصاء في الفترة (١٩٧٦-٢٠٠٦)

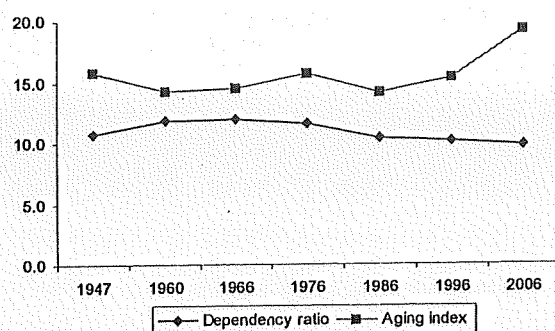
٢-٧: معدل إعالة المسنين، ودليل التعمر:

يعرف معدل إعالة المسنين (٦٠ سنة فأكثر) بأنه نسبة المسنين لإجمالي السكان في الفئة العمرية (١٥-٥٩)، كما تعرف أيضاً بنسبة الإعالة الكبرى، حيث توضح هذه النسبة الأعباء التي يجب أن يتحملها الجزء المنتج من المجتمع نتيجة ارتفاع أعداد كبار السن، كما يساعد في

تقدير الاحتياجات الصحية والاجتماعية والنفسية لهذه الفئة من كبار السن، والتعرف على فرص الاستفادة من خبرات كبار السن في مختلف مجالات التنمية بالدولة. أما عن دليل التعمر فهو النسبة المئوية لكبار السن إلى فئة الأطفال (٠-١٤)، حيث أن الديموجرافيين قد اصطاحوا أن هذه النسبة إذا كانت أقل من ١٥% اعتبر المجتمع فتياً.

ويعرض شكل (٧-٢) معدل إعالة المسنين، ودليل التعمر في مصر خلال الفترة (١٩٤٧-٢٠٠٦)، ومن الملاحظ أن معدل الإعالة يميل إلى الانخفاض، حيث بدأ ينخفض منذ عام ١٩٦٦ حيث انخفض انخفاض طفيف جداً

شكل (٧-٢)
معدل إعالة المسنين، ودليل التعمر خلال التعدادات في الفترة
(١٩٤٧-٢٠٠٦)



المصدر: محسوب بمعرفة الباحث من التعدادات خلال الفترة
(١٩٤٧-٢٠٠٦)

من ١١,٩% إلى ١١,٦% عام ١٩٧٦، ثم كان الانخفاض ملحوظ عام ١٩٨٦ إلى ١٠,٤%، وكانت نسبة هذا المعدل ١٠,٢% عام ١٩٩٦، وأخيراً عام ٢٠٠٦ كان هذا المعدل هو ٩,٨%.

أما عن دليل التعمر الذي يقيس عبء إعالة الكبار إلى عبء إعالة الصغار فإن ما توضحه الأرقام، وشكل (٧-٢) أن تعداد ١٩٨٦ اعتبر المجتمع فتياً حيث كانت نسبة هذا الدليل هي ١٤,١%، ثم

بدأت تميل هذه النسبة إلى الارتفاع حيث ارتفعت هذه النسبة إلى ١٥,٣% عام ١٩٩٦، ثم كانت ١٩,١% في عام ٢٠٠٦، مما يدل على أن مصر في مرحلة التحول نحو الشيخوخة.

٣. التوقعات المستقبلية للمسنين في مصر:

تشير البيانات إلى أن حجم الفئة العمرية (٦٠ فأكثر) بدأت في الزيادة منذ عام ١٩٨٦ حيث كانت نسبة المسنين ٥,٧% من إجمالي الجمهورية، وأصبحت عام ٢٠٠٦ هذه النسبة حوالي ٦,١%، كما أن التوقعات بما سيطرأ من زيادات على عدد ونسب السكان المسنين في مصر وضرورة العمل على تلبية احتياجاتهم سوف تشكل تحدياً حقيقياً يجب المبادرة من الآن بوضع استراتيجيات مناسبة بهدف تحقيق العدالة بين كافة فئات المجتمع، بما في ذلك المسنين. وسوف يتم استعراض هذه التوقعات من خلال ما يلي:

٣-١: إسقاطات المركز الديموجرافي بالقاهرة خلال الفترة (١٩٩٦-٢٠٢١):

بالرغم من الاستقرار التقريبي للتركيب العمري في مصر حتى الآن إلا أن الاسقاط الذي قام به المركز الديموجرافي بالقاهرة يتوقع أن يتغير هذا التركيب في اتجاه زيادة نسبة المسنين في المجتمع. فجدول (٣-١) يستعرض التقدير العددي والنسبي للمسنين حسب النوع خلال الفترة (١٩٩٦-٢٠٢١) تبعاً لفرض الخصوبة المتوسط الذي يشترط انخفاضها إلى مستوى الإحلال (٢,١ طفل لكل سيدة) بحلول عام ٢٠٣٠.

جدول (٣-١)

التقدير العددي والنسبي لكبار السن (٦٠ سنة فأكثر) حسب النوع خلال الفترة (١٩٩٦-٢٠٢١)
تبعاً لفرض الخصوبة المتوسط

العدد بالآلاف

النوع	١٩٩٦	٢٠٠١	٢٠٠٦	٢٠١١	٢٠١٦	٢٠٢١
ذكور	١٧٠١٩	١٩٣٣	٢٢٤١	٢٧٣٤	٣٣٦٠	٤٠٥٥
إناث	١٥٧٤	١٨١٤	٢١٩٥	٢٧٤٥	٣٤٦٨	٤٣٢٤
إجمالي	٣٢٩٣	٣٧٤٧	٤٤٣٦	٥٤٧٩	٦٨٢٨	٨٣٧٩
النسبة	٥,٦	٥,٨	٦,٢	٧,٠	٨,١	٩,٢

المصدر: المركز الديموجرافي بالقاهرة، إسقاطات السكان المستقبلية لمحافظة مصر لأغراض التخطيط والتنمية، ٢٠٠٠.

وتشير بيانات الجدول إلى توقع ارتفاع أعداد المسنين إلى ٨,٤ مليون مسن عام ٢٠٢١ بزيادة حوالي ١٥٤% عن عام ١٩٩٦، كما يتوقع ارتفاع نسبة المسنين من ٥,٦% عام ١٩٩٦، إلى حوالي ٩,٢% تقريباً عام ٢٠٢١، كما تشير البيانات إلى توقع زيادة أعداد الإناث عن الذكور.

٣-٢: التقديرات السكانية للأمم المتحدة لكبار السن في مصر في الفترة (٢٠٠٥-٢٠٥٠):

تشير تقديرات الأمم المتحدة (UN, 2008)، تبعاً لفرض الخصوبة المتوسط والذي يتوقع انخفاضها إلى مستوى الإحلال (٢,١ طفل لكل سيدة) بحلول عام ٢٠٣٥، في جدول (٣-٢) يتبين أن عدد المسنين في مصر (٦٠ سنة فأكثر) سيزداد من ٥,١ مليون مسن بنسبة ٦,٧% عام ٢٠٠٥، إلى ٩ مليون مسن تقريباً بنسبة ٩,١% عام ٢٠٢٠، وهذه النسبة قريبة من تقدير المركز الديموجرافي، كما يتوقع أن يكون عدد المسنين المقدر عام ٢٠٣٠ هو ١٢,٦ مليون مسن بنسبة ١١,٤%، كما سيزداد عدد السكان المسنين إلى ٢٤,٨ مليون مسن، وذلك بنسبة ١٩,٢% من إجمالي السكان المتوقع بحلول عام ٢٠٥٠.

وتشير تقديرات الأمم المتحدة أن مصر سوف تصل إلى مستوى الإحلال عام ٢٠٣٥، وبناء على ذلك يلاحظ من الجدول أن مصر لن تدخل في مرحلة الشيخوخة إلى بعد حلول عام ٢٠٤٠، أي بعد الوصول إلى مستوى الإحلال. كما تشير بيانات الجدول أن أعداد الإناث المسنات ستزيد عن أعداد الذكور المسنين.

جدول (٣-٢)

التقدير العددي والنسبي للأمم المتحدة بالنسبة لكبار السن (٦٠ سنة فأكثر) لمصر حسب النوع خلال الفترة (٢٠٠٥-٢٠٥٠)، تبعاً لفرض الخصوبة المتوسط

العدد بالآلاف

السنة	٦٠ سنة فأكثر				٧٥ سنة فأكثر			
	الذكور	الإناث	الجملة	النسبة	الذكور	الإناث	الجملة	النسبة
2005	2413	2743	5157	6.7	455	599	1054	1.4
2010	2983	3311	6295	7.5	536	711	1247	1.5
2015	3556	3919	7477	8.1	624	824	1449	1.6
2020	4246	4725	8972	9.1	720	934	1655	1.7
2025	5021	5712	10731	10.2	966	1232	2196	2.1
2030	5804	6805	12609	11.4	1179	1553	2732	2.5
2035	6656	7943	14596	12.5	1438	1989	3425	2.9
2040	7880	9464	17343	14.3	1721	2514	4235	3.5
2045	9460	11345	20806	16.5	2004	3066	5071	4.0
2050	11352	13494	24846	19.2	2352	3625	5978	4.6

المصدر: الأمم المتحدة، شعبة السكان التابعة لإدارة الشؤون الاقتصادية والاجتماعية، التوقعات السكانية العالمية، تنقيح عام ٢٠٠٨

ونظراً للتقدم في المجال الصحي وزيادة العمر المتوقع للبقاء على قيد الحياة عند الميلاد من ٧٠ سنة عام ٢٠٠٥، إلى ٧٥,٣ سنة عام ٢٠٣٠، ثم ٧٧,٧ سنة عام ٢٠٥٠ (UN, 2008)، فيلاحظ من الجدول أن شريحة المسنين (٧٥ سنة فأكثر) سيزداد عدد المسنين في هذه الشريحة من مليون مسن عام ٢٠٠٥ والتي تمثل ١,٤% من إجمالي السكان، ٢٠% من إجمالي المسنين إلى ٢,٧ مليون مسن بنسبة ٢,٥% من إجمالي السكان، ٢٢% من إجمالي المسنين (٧٥ سنة فأكثر) في مصر عام ٢٠٣٠، كما تشير البيانات أن عدد المسنين في هذه الشريحة سيرتفع إلى ٦ مليون مسن بحلول عام ٢٠٥٠، تمثل حوالي ٤,٦% من إجمالي السكان، ٢٤,١% من إجمالي المسنين. كما تشير بيانات الجدول أيضاً أن أعداد الإناث المسنات في هذه الشريحة ستزيد عن أعداد الذكور المسنين وستكون هذه الزيادة ملحوظة عام ٢٠٣٠، فسيكون مقابل كل أربعة من المسنين الذكور حوالي ستة من الإناث المسنات عام ٢٠٣٠.

أما عن العمر الوسيط للسكان الذي قد يستخدم في توضيح ما إذا كان الهيكل السكاني يتجه إلى الكبر أو إلى الصغر، حيث نجد أن المجتمعات المعمرة يقع وسيطها في الثلاثينات، كما تشير

تقديرات الأمم المتحدة أن العمر الوسيط للسكان سيرتفع في مصر من ٢٢ سنة عام ٢٠٠٥ إلى ٣٠ سنة تقريباً عام ٢٠٣٠، ثم إلى ٣٦ سنة عام ٢٠٥٠. مما يدل أن مصر ستدخل ضمن المجتمعات المعمرة بعد حلول عام ٢٠٣٠.

٣-٣: التقديرات باستخدام برنامج Spectrum :

قد تم استخدام برنامج Spectrum، تبعاً لفرض الخصوبة المتوسط الذي يشترط انخفاضها إلى مستوى الإحلال بحلول عام ٢٠٣٠، واعتماداً على بيانات تعداد ٢٠٠٦، وذلك لتقدير أعداد السكان والمسنين وكذلك العمر الوسيط للسكان حتى عام ٢٠٣١.

جدول (٣-٣)

التقدير العددي والنسبي لكبار السن (٦٠ سنة فأكثر) في مصر، في الفترة (٢٠٣١-٢٠٠٦)

باستخدام برنامج Spectrum

السنة	إجمالي السكان	العمر الوسيط	٦٠ سنة فأكثر		
			ذكور	إناث	النسبة
2006	72,613,352	23	2,311,940	2,102,276	6.1
2011	79,988,274	25	2,881,123	2,700,429	7
2016	87,360,969	26	3,564,657	3,600,866	8.2
2021	93,997,000	28	4,360,517	4,610,056	9.5
2026	99,675,761	29	5,132,081	5,650,856	10.8
2031	104,696,819	31	5,921,336	6,787,371	12.1

المصدر: محسوب بمعرفة الباحث.

ويشير جدول (٣-٣) إلى التقدير العددي لإجمالي السكان، والتقدير العددي والنسبي للمسنين (٦٠ سنة فأكثر) وكذلك العمر الوسيط للسكان، ويتضح منه أن عدد سكان مصر سوف يزد من ٧٢,٦ مليون نسمة عام ٢٠٠٦، إلى ١٠٤,٧ مليون نسمة عام ٢٠٣١، أي أن عدد السكان سوف يزد بنسبة ٤٤% عام ٢٠٣١ عن عام ٢٠٠٦، كما تشير التقديرات في الجدول أن عدد المسنين سوف يتزايد من ٤,٤ مليون مسن عام ٢٠٠٦، إلى ١٢,٧ مليون مسن عام ٢٠٣١، أي أن أعداد المسنين سوف تزد بنسبة ١٨٨% بين عام ٢٠٠٦ وعام ٢٠٣١، كما يوضح الجدول أن نسبة المسنين سوف ترتفع من ٦,١% عام ٢٠٠٦ إلى ١٢,١% عام ٢٠٣١.

كما تشير بيانات الجدول أن أعداد الإناث المسنات ستزيد عن أعداد الذكور المسنين بدءاً من عام ٢٠١٦، كما يبين الجدول أن العمر الوسيط للسكان سيرتفع من ٢٣ سنة عام ٢٠٠٦ إلى ٣١ سنة عام ٢٠٣١.

مما سبق، يلاحظ أن تقديرات المركز الديموجرافي بالقاهرة، والتقديرات السكانية للأمم المتحدة، والتقديرات المحسوبة، اتفقت على أن مصر لن تدخل في مرحلة الشيخوخة قبل حلول عام ٢٠٤٠.

٤. الخلاصة والتوصيات:

٤-١: الخلاصة:

العرض السابق تناول تطور حجم المسنين (٦٠ سنة فأكثر) في مصر منذ تعداد ١٩٤٧ إلى تعداد ٢٠٠٦، كما تناول بعض الخصائص الديموجرافية والاجتماعية والاقتصادية لهذه الفئة.

ومن خلال ذلك يتبين ما يلي:

- أن الزيادة في نسبة كبار السن قد تذبذبت خلال الفترات التعدادية (١٩٤٧ - ٢٠٠٦) حول ٦%.
- ظاهرة التعمر ناتجة عن عملية التحول الديموجرافي، ومن نتاج المتغيرات المتمثلة في معدل المواليد، ومعدل الوفيات، فكلما انخفضا فإن نسبة صغار السن (٠-١٤) تنخفض، ونسبة كبار السن (٦٠ سنة فأكثر) ترتفع.
- عند دراسة التركيب العمري للمسنين في مصر للتعرف على التفاوت في اتجاه التعمر، تبين أن الغالبية العظمى للمسنين تقع في الشريحة الأولى لهم (٦٠-٦٤)، حيث تخطت نسبة هذه الفئة ٤٠% من إجمالي المسنين في مصر خلال الفترة (١٩٤٧-١٩٩٦)، وبالرغم من ذلك تبين أن هذه الفئة تأخذ اتجاه الانخفاض وذلك لصالح الفئات العمرية الأعلى وبالأخص الفئة العمرية الثانية (٦٥-٦٩)، حيث انخفضت الفئة الأولى إلى ٣٨,٥% من إجمالي المسنين بينما ارتفعت الفئة الثانية إلى ٢٧% عام ٢٠٠٦.
- وبدراسة التركيب النوعي في مصر تبين أن في الفترة (١٩٤٧-١٩٧٦) زيادة أعداد الإناث المسنات عن المسنين الذكور، ولكن منذ تعداد ١٩٨٦ كان الاتجاه معاكساً لذلك، حيث فاقت أعداد الذكور المسنين أعداد الإناث المسنات.
- نسبة إجمالي المسنين في الريف أعلى من المسنين في الحضر، حيث يوجد تباين واضح بينهما، كما يلاحظ أن هذا التباين في انخفاض مستمر حيث تقلص الفرق بين الريف والحضر إلى ٥% عام ٢٠٠٦.

- وعند دراسة النوع في كل من الحضر والريف يلاحظ أن نسبة الذكور المسنين في الحضر أكبر من نسبة الإناث المسنات، وعلى العكس من ذلك نجد أن نسبة الإناث المسنات في الريف أعلى من الذكور المسنين حتى تعداد ١٩٩٦، أما عام ٢٠٠٦ نسبة الذكور المسنين في الريف أعلى من الإناث المسنات.
- يشكل كبار السن في الوجه البحري الغالبية العظمى من المسنين، حيث يوجد أكثر من ٤٠% من إجمالي المسنين، يلي ذلك الوجه القبلي بالرغم أن نسب المسنين به تميل إلى الانخفاض، كما يلاحظ أن نسبة المسنين في المحافظات الحضرية وحدها خمس المسنين تقريباً.
- كما تبين أن الغالبية العظمى من نسبة كبار السن الذكور متزوجين، أما الإناث المسنات فأغلبهن من المترملات. قد يرجع ذلك لأن الذكور المسنين قد يكون لديهم القدرة والإمكانية للزواج مرة أخرى
- عند دراسة الحالة التعليمية للمسنين تبين أن نسبة الأمية مرتفعة للغاية بين المسنين، ولكن نسبة هذه الشريحة الأمية تميل إلى الانخفاض حيث انخفضت من ٧٩% عام ١٩٧٦ إلى ٦٩% عام ٢٠٠٦، كما تضاعفت نسبة المسنين ذوو المستوى الجامعي فأعلى من ١,١% عام ١٩٤٧ إلى ٦,٥% عام ٢٠٠٦، كما ترتفع نسب الأمية بين الإناث المسنات عن نسب الأمية بين الذكور المسنين بدرجة كبيرة، كما أن الإناث أقل حظاً من الذكور في جميع مستويات التعليم.
- عند دراسة مساهمة المسنين في قوة العمل تبين أن التناقص في مساهمتهم في قوة العمل خلال الفترة (١٩٧٦-٢٠٠٦) هو الاتجاه السائد، حيث انخفضت نسبة مساهمتهم من ٦,٧% عام ١٩٧٦ إلى ٣,٢% عام ٢٠٠٦، كذلك تبين أن أكثر من نصف المسنين بقليل (حوالي ٦٠%) من المساهمين في قوة العمل تقع في الفئة الأولى (٦٠-٦٤ سنة). كما ترتفع نسبة العاملين الذكور المسنين عن الإناث المسنات في جملة المساهمة في قوة العمل من المسنين.
- معدل إعالة المسنين يميل إلى الانخفاض، أما عن دليل التعمر بدأ يميل إلى الارتفاع منذ عام ١٩٨٦ حيث أصبحت نسبة هذا الدليل ١٩,١% في عام ٢٠٠٦.
- من المتوقع أن يتغير التركيب العمري لسكان مصر في اتجاه زيادة نسبة المسنين، فتبعاً لفرض الخصوبة المتوسط حيث اتفقت جميع التقديرات المتوقعة أن أعداد الإناث المسنات ستزيد عن أعداد الذكور المسنين، كما أن مصر لن تدخل في مرحلة الشيخوخة قبل حلول

عام ٢٠٤٠. فيقدر المركز الديموجرافي بالقاهرة نسبة المسنين عام ٢٠٢١ بـ ٩,٢% من إجمالي السكان، كما تشير تقديرات الأمم المتحدة (UN, 2008)، أن مصر ستصل إلى مستوى الإحلال (٢,١ طفل لكل سيدة) بحلول عام ٢٠٣٥، وأن نسبة المسنين في مصر (٦٠ سنة فأكثر) ستصل إلى ٩,١% عام ٢٠٢٠، وبحلول عام ٢٠٥٠، سيكون نسبة المسنين ١٩,٢% من إجمالي السكان المتوقع. وعند استخدام بيانات تعداد ٢٠٠٦ في برنامج Spectrum نتج أن نسبة المسنين سترتفع من ٦,١% عام ٢٠٠٦ إلى ١٢,١% عام ٢٠٣١.

٢-٤: التوصيات:

- توفير البيانات الدقيقة والموثقة والقابلة للمقارنة من أجل بناء قاعدة إحصائية شاملة تتيح التعرف على أوضاع المسنين بغية صياغة السياسات والاستراتيجيات المناسبة.
- ضرورة النظر على فئة المسنين كطاقة إنتاجية لا يصح إهدارها نظراً لخبراتها، وذلك بوضع برامج شاملة لإعادة تأهيلهم ليكونوا أكثر قدرة على ممارسة أعمال تتسق مع مرحلتهم العمرية واحتياجاتهم ومتطلبات المجتمع.
- تشجيع وسائل الإعلام على إبراز الصورة الايجابية للمسنين والتركيز على دورهم في المجتمع.
- توفير وتحسين الخدمات الاجتماعية والصحية المرتبطة بالمسنين، وتدريب جميع المتعاملين مع المسنين، وذلك لزيادة كفاءة الخدمات المقدمة.
- توفير المعلومات اللازمة للمسنين حول أنواع الخدمات الصحية المتوافرة، وأماكن توافرها، وكيفية الاستفادة منها.

المراجع

أولاً: المراجع العربية:

١. ارناؤط، سعيد (٢٠٠١). "المنظمات والهيئات الدولية ودورها في رعاية المسنين"، ندوة (المدينة والمسنون .. دور المدن والبلديات في رعاية المسنين)، ٢٠ - ٢٢ فبراير ٢٠٠١، المعهد العربي لإنماء المدن، القاهرة.
٢. الأمم المتحدة (٢٠٠٢)، "تقرير الجمعية العالمية الثانية للشيخوخة"، مدريد ٨-١٢ أبريل ٢٠٠٢، A/CONF.197/9 منشورات الأمم المتحدة، رقم المبيع: A.02/IV.4، نيويورك.
٣. الأمم المتحدة (١٩٥٦)، "تعمر السكان ونتائج الاقتصادية والاجتماعية"، قسم الشؤون الاقتصادية والاجتماعية، فرع السكان، دراسات سكانية رقم ٢٦، رقم المبيع: 1956XIII.6 ترجمة: صلاح الدين الشريف، المركز الديموجرافي لشمال إفريقيا، القاهرة.
٤. الإسكوا (اللجنة الاقتصادية والاجتماعية لغربي آسيا)، (٢٠٠٨). "الملاحم الديموجرافية للبلدان العربية .. شيخوخة السكان في المناطق الريفية"، الأمم المتحدة.
٥. الإسكوا (٢٠٠٧). "شيخوخة السكان في البلدان العربية"، الأمم المتحدة.
٦. الإسكوا (٢٠٠٢). "خطة العمل العربية للمسنين حتى عام ٢٠١٢"، الأمم المتحدة.
٧. البنا، احمد سرور (٢٠٠١). "الخصائص والاتجاهات الحديثة للمسنين في التجمعات الحضرية والمدن المصرية"، ندوة (المدينة والمسنون .. دور المدن والبلديات في رعاية المسنين)، ٢٠ - ٢٢ فبراير ٢٠٠١، المعهد العربي لإنماء المدن، القاهرة.
٨. الجهاز المركزي للتعبئة العامة والإحصاء، إحصاءات التعدادات السكانية والإسكان خلال الفترة (١٩٤٧-٢٠٠٦).
٩. أمين، سعد زغلول (١٩٩١). "تعمر السكان في مصر"، ندوة (نحو رعاية متكاملة للمسنين)، ٣-٥ مارس ١٩٩١، المركز القومي للبحوث الاجتماعية والجنائية، القاهرة.
١٠. بونجارت، جون؛ زيمر، زخاري (٢٠٠١). "التدابير المعيشية الخاصة بالمسنين في العالم النامي: تحليل المسوح الصحية والديموجرافية للأسر المعيشية"، ترجمة: نولة درويش (٢٠٠٦)، مجلس السكان الدولي.

١١. جواهر، محمود محمد (٢٠٠١). "دور التخطيط العمراني في رعاية المسنين"، ندوة (المدينة والمسنون .. دور المدن والبلديات في رعاية المسنين)، ٢٠ - ٢٢ فبراير ٢٠٠١، المعهد العربي لإنماء المدن، القاهرة.
١٢. حجازي، محمد عزت سليم (٢٠٠٤). "كبار السن في مصر: مسح بالعينة لنزلاء دور المسنين"، المركز القومي للبحوث الاجتماعية والجنائية، القاهرة.
١٣. قنديل، شحاتة محمد شحاتة (١٩٩٧). "التعمر الديموجرافي في مصر "احتمالاته وآثاره"، رسالة ماجستير في الإحصاء السكاني، معهد الدراسات والبحوث الإحصائية، جامعة القاهرة.
١٤. مخلوف، هشام؛ نجيب، محمد (٢٠٠٣). "المسنون في مصر ... ديموجرافياً واجتماعياً واقتصادياً"، أوراق في ديموجرافية مصر، رقم ٥، المركز الديموجرافي، القاهرة.
١٥. مخلوف، هشام؛ وآخرون (٢٠٠٠). "إسقاطات السكان المستقبلية لمحافظة مصر لأغراض التخطيط والتنمية"، المركز الديموجرافي، القاهرة.
١٦. نوفل، حلا؛ فاعور، تانيا (٢٠٠٦). "تعمر السكان في البلدان العربية: الواقع والتحديات والاستراتيجيات المطلوبة". المؤتمر العربي الأول لصحة الأسرة والسكان، ١٣-١٦ مايو ٢٠٠٦، جامعة الدول العربية، القاهرة.

ثانياً: المراجع الأجنبية:

17. Chemali, Z.; Chahine, L.M. and Sibai, A.M. (2008) Older adult care in Lebanon: towards stronger and sustainable reforms, Eastern Mediterranean Health Journal, Vol. 14, No. 6, 2008.
18. Soliman, Nagat (1994), "Aging in Egypt", Cairo Demographic Center.
19. United Nations (2008): Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population Prospects: The 2008 Revision
20. United Nations (2007), World Population Prospects "The 2006 Revision", Department of Economic and Social Affairs Population Division, ESA/P/WP.202, New York
21. United Nations (2002), "World Population Ageing 1950-2050", Population Division, DESA.

1871. The first of these was the
the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

the first of these was the

الإحتياجات الحالية والمستقبلية من الوحدات السكنية

في مصر خلال الفترة من (٢٠٠٧ إلى ٢٠٢٢)

جلال حلمى سليم

مقدمة:

تحتل قضية الإسكان فى مصر موقعا هاما ومتقدما فى قائمة الأولويات الاجتماعية والاقتصادية ، باعتبار أن المسكن الملائم يأتى فى مقدمة احتياجات الأسر الجديدة ، كما أنه يعتبر من مقومات تحسين معيشة كثير من الأسر القائمة.

إن الوضع الحالى للإسكان فى مصر يظهر إختلالا فى العلاقة بين الطلب والعرض على مستوى نوعيات الإسكان ، فهناك فجوة بين الاحتياجات من إسكان محدودى الدخل والمتاح منه وذلك لأن تكلفة معظم الوحدات السكنية التى يتم تشييدها حالياً تفوق قدرة معظم المواطنين ذوى الدخل المحدود .

الهدف من الدراسة :

تهدف هذه الدراسة إلى توضيح وتحديد خصائص الشعب المصرى للمساعدة فى وضع الخطط اللازمة للتنمية العمرانية حتى عام ٢٠٢٢ و تتلخص هذه الخصائص فيما يلى :

- التعداد السكانى و معدلات النمو السكانى خلال العشرين عاما سابقا و حتى عام ٢٠٠٦ .
- الهيكل السكانى مقسما لفئات العمر و الجنس و محافظات (ريف / حضر) و اجمالى الجمهورية و عدد الأسر خلال هذه الفترة .

منهجية الدراسة :

تتبع الدراسة المنهج الوصفى التحليلى حيث تستخدم البيانات الأساسية عن الوضع الراهن للوحدات السكنية فى مصر واستخلاص بعض المؤشرات والتقديرات التى تخدم أهداف الدراسة بحيث تكون هناك صورة علمية واضحة يمكن بها وضع السياسات والبرامج اللازم تطبيقها وخاصة لتلبية احتياجات محدودى الدخل من الوحدات السكنية .

الوضع الحالي للإسكان في مصر

أولاً : حصر الوحدات السكنية الحالية ونسبة الأشغال :

يتوقف تقدير الاحتياجات الحالية من الوحدات السكنية على تحديد البنود التالية :

١- الوحدات السكنية المطلوبة مقابل التزاوج وتتمثل في تكديس أكثر من أسرة في منزل واحد ، أو بعدد الأفراد المكديسين بغرفة واحدة وهو من المعدلات التي تشير إلى الأزمة السكانية و أيضاً يشير إلى المستوى المعيشي للأسرة و يتناسب عكسياً مع مستوى دخل الأسرة .

و نظرا لأنه من الصعب وضع حل شامل لمشكلة الزيادة السكانية خلال المرحلة القادمة فسوف نتجاهل حالياً معدل التزاوج حتى يتم تسكين جميع الأسر و حتى يصبح في أماكن كل أسرة الحصول على مسكن ملائم يبدأ بعدها النظر في تخفيض معدل التزاوج و خصوصا أنه بدراسة متوسط عدد أفراد الأسرة في السنوات العشر السابقة اتضح أنه في تناقص مستمر بما يعنى أن معدل التزاوج ينخفض تلقائيا نتيجة الى الانخفاض التدريجي في متوسط حجم الأسرة .

٢ - الوحدات السكنية الملائمة بدلا من الوحدات ذات المستوى المتدنى (المساكن الجوازية) و تتمثل في المساكن غير المعدة أصلاً للسكن و التي لا يصل اليها اى نوع من المرافق العامة (حجرات المقابر- عشة- خيمة- مركب الخ) و التي تقدر بحوالى ٢١٧ ألف وحدة وفقا لنتائج التعداد العام للسكان والمنشآت ٢٠٠٦ .

٣ - الوحدات السكنية المطلوبة للأسر الجديدة و تتمثل في الشباب في سن الزواج أى البالغين ١٨ سنة للذكور ، ١٦ سنة للإناث (لم يتزوج أبداً + عقد قران).

$$\begin{aligned} & \text{الوحدات السكنية المطلوبة للأسر الجديدة} \\ & \text{لم يتزوج أبداً + عقد قران} \\ & = \\ & \frac{2}{2} \\ & = \frac{(13609214 + 2250250)}{2} \\ & = 7929732 \text{ وحدة} \end{aligned}$$

٤ - الوحدات السكنية المطلوبة مقابل الإحلال و التجديد و ذلك لتعرض نسبة من المساكن الموجودة للاهلاك و التقادم و عدم صلاحيتها للاستخدام بشكل جيد (المساكن الآيلة للسقوط) و تمثل هذه النسبة ١٪ من المساكن القائمة (عدد الأسر القائمة) و ذلك لإحلال مساكن جديدة محل المساكن القديمة (و هى النسبة المتعارف عليها و المقدرة في الأبحاث الدولية و القومية المشابهة)، و تقدر بحوالى ١٧٢ ألف وحدة.

٥ - الوحدات السكنية المطلوبة مقابل الاحتياطي و هى عبارة عن المساكن الاضافية الى عدد المساكن المأهولة بالسكان و ذلك لان أى عائلة تريد أن تنتقل من منزلها لأى سبب من الأسباب لابد أن تبحث عن منزل قبل انتقالهم من منزلهم الحالي ، اذا تعتبر المساكن الاحتياطية عاملاً هاماً فى استقرار اسعار المساكن و الايجارات ، و لأنه ليس لدينا معلومات كافية عن المساكن الاحتياطية ولا عن نسبة العائلات التي تقوم بتغيير مساكنها لأى سبب من الأسباب ، فكانت أفضل طريقة لتحديد نسبة الاحتياطي من الوحدات السكنية هى الرجوع الى خبرة بعض الدول التي سبقت فى احتساب هذه النسبة ففى بعض الدول المتقدمة يتم تقديرها بـ ٨٪ من عدد المساكن المأهولة بالسكان و هذه

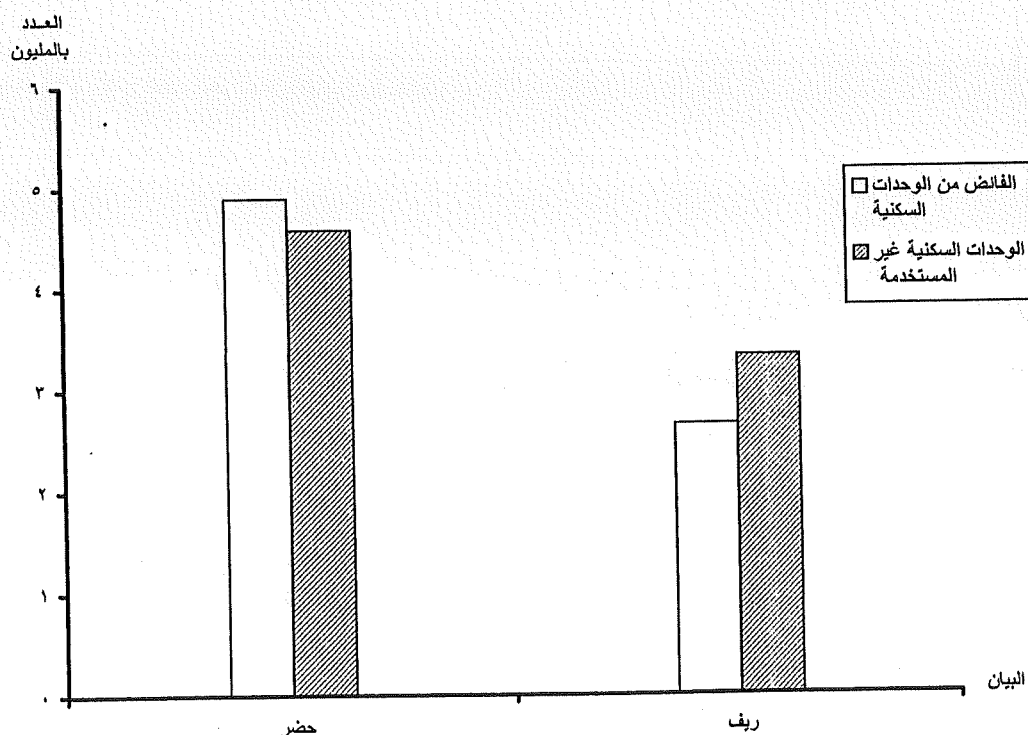
جدول رقم (١)

عدد الوحدات السكنية ونسبة الفائض على مستوى حضر / ريف الجمهورية

وعدد الوحدات غير المستخدمة (المغلقة - الخالية) تعداد ٢٠٠٦

البيان	عدد الأسر	عدد الوحدات السكنية	الفائض من الوحدات السكنية		الوحدات السكنية غير المستخدمة (المغلقة - الخالية)	
			العدد	% لعدد الوحدات السكنية	العدد	% لعدد الوحدات السكنية
حضر	٧٧٥١٥١٢	١٢٦٤٧٩٧٠	٤٨٩٦٤٥٨	٣٨,٧	٤٥٨٠١٢٢	٣٦,٢
ريف	٩٥١٤٠٥٥	١٢١٦٧٣٤٢	٢٦٥٣٢٨٧	٢١,٨	٣٣١٨٩٦٢	٢٧,٣
إجمالي الجمهورية	١٧٢٦٥٥٦٧	٢٤٨١٥٣١٢	٧٥٤٩٧٤٥	٣٠,٤	٧٨٩٩٠٨٤	٣١,٨

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء، نتائج التعداد العام للسكان والإسكان والمنشآت، ٢٠٠٦، إجمالي الجمهورية، مارس ٢٠٠٧، ص ١٠٠.



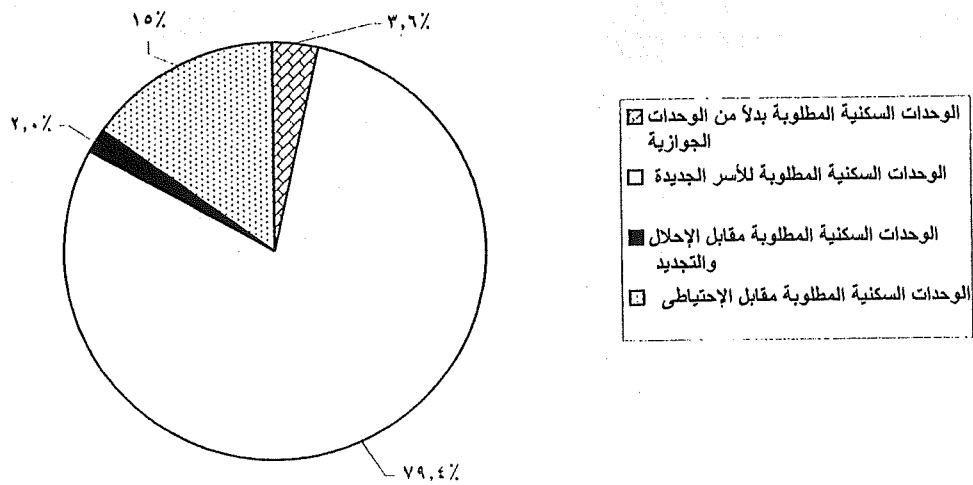
جدول رقم (٢)

الاحتياجات الحالية من الوحدات السكنية لإجمالي الجمهورية عام ٢٠٠٦ (الطلب)

م	البيان	الوحدات السكنية *	
		العدد	% للإجمالي
١	الوحدات السكنية المطلوبة بدلاً من الوحدات ذات المستوى المتدنى (المساكن الجوازية)	٣١٦٩٩٥	٣,٦
٢	الوحدات السكنية المطلوبة للأسر الجديدة (الشباب في سن الزواج) **	٦٩٩٧١٧٤	٧٩,٤
٣	الوحدات السكنية المطلوبة مقابل الإحلال والتجديد (١% من إجمالي عدد الأسر القائمة)	١٧٢٦٥٦	٢,٠
٤	الوحدات السكنية المطلوبة مقابل الإحتياطي (٨% من عدد الوحدات السكنية المأهولة بالسكان)	١٣٢٧٩٣٩	١٥,٠
	الإجمالي	٨٨١٤٧٦٤	١٠٠

* بيانات محسوبة من نتائج التعداد العام للسكان والإسكان والمنشآت ٢٠٠٦ .

** الجهاز المركزي للتعبئة العامة والإحصاء ، نتائج التعداد العام للسكان والإسكان والمنشآت عام ٢٠٠٦ .



(١) نسبة متعارف عليها دوليا و تتمثل في:

الوحدات السكنية المطلوبة مقابل الاحتياطي = عدد الوحدات السكنية المخصصة للسكن / المخصصة للسكن والعمل $\times 8\% = 16525589 + 73644 \times 8\% = 1327929$ وحدة.
ويوضح الجدول رقم (١)، (٢) الاحتياجات الحالية من الوحدات السكنية لإجمالي الجمهورية عام ٢٠٠٦ وفقاً للبنود السابق ذكرها.

ثانياً: الفجوة الاسكانية:

١- على مستوى إجمالي الجمهورية:

تعرف الفجوة الإسكانية بأنها الفرق بين المعروض من الوحدات السكنية والمطلوب منها ، و حتى يتمكن من تقدير الفجوة الإسكانية لابد من توضيح حجم المعروض من الوحدات السكنية الشاغرة أو غير المستخدمة (المغلقة + الخالية) و ذلك لسد جزء من الاحتياجات الحالية من الوحدات السكنية.

و يوضح الجدول رقم (٣) المعروض من الوحدات السكنية لإجمالي الجمهورية عام ٢٠٠٦ ، و منه نجد أن إجمالي عدد الوحدات السكنية المعروض منه بلغ حوالى ٧,٨٩٩ مليون وحدة منها ٢,١٢٢ مليون وحدة مغلقة يخص الحضر منها ٥٥,١ % و الريف ٤٤,٩ % من إجمالي الوحدات المغلقة.

أما بالنسبة لعدد الوحدات السكنية الخالية فقد بلغت نحو ٥,٢٦٦ مليون وحدة يخص الحضر منها ٥٩% و الريف ٤١% من إجمالي الوحدات الخالية.

و يوضح جدول رقم (٤) تقدير الفجوة الإسكانية لإجمالي الجمهورية عام ٢٠٠٦ ، و منه يتضح وجود عجز من الوحدات السكنية قدره حوالى ٩١٦ الف وحدة و هذا الرقم لا يمثل الفجوة الفعلية لأن المعروض من الوحدات السكنية يتضمن وحدات سكنية مغلقة قدرها حوالى ٢,١٢٢ مليون وحدة قد لا تدخل معظمها بالفعل فى سوق العقارات المطروحة للبيع أو الإيجار حيث أنها مملوكة أو مؤجرة لأسر من الممكن أن تعرضها أو من الممكن أن تحتفظ لنفسها مثل الوحدات السكنية الموجودة بالمحافظات السياحية (الإسكندرية - مطروح - البحر الاحمر - الخ) او تحتفظ بها لتكون مسكناً لأحد أفراد الأسرة (الأبناء - الأخوة - أحد الأقارب) و لهذا فأنها قد لا تغطى إلا جزءاً قليلاً من الطلب.

ومن حيث الفائض من الوحدات السكنية فلا يوجد إلا بإقليم القاهرة الكبرى فقط (بمحافظة الجيزة القليوبية) ويرجع السبب فى ذلك لاهتمام الدولة بهذا الإقليم وتوفير الوحدات السكنية به ، وذلك بإنشاء عدد من المدن الجديدة به ، وذلك لأن محافظات هذا الاقليم من المحافظات المستوعبة للهجرة الداخلية من المحافظات الأخرى ، لذلك فإن تكديس أكبر عدد من السكان يوجد بهذا الاقليم حيث بلغ عدد سكانه ١٨,٢٩٦ مليون نسمة يمثل ٢٥,٢١ % من اجمالى سكان الجمهورية ، وقيم بهذا الإقليم ٤,٦٤٢ مليون أسرة تمثل ٢٦,٨٨ % من اجمالى الأسر بالجمهورية ، (جدول رقم ٣) ، كما يوجد بهذا الاقليم ايضا ١,٩٢٠ مليون من الأسر الجديدة (الشباب فى سن الزواج) و يمثلون ٢٧,٤٤ % من اجمالى الأسر الجديدة لإجمالى الجمهورية.

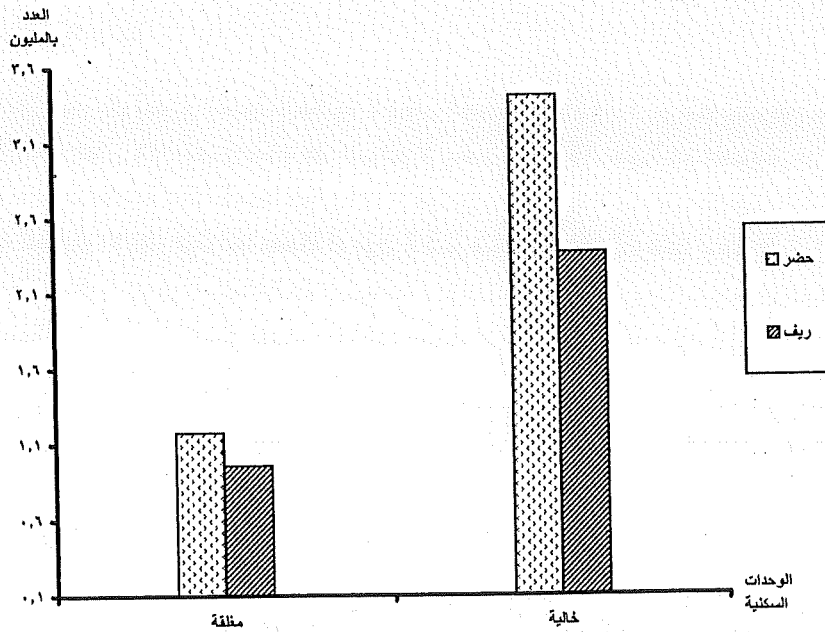
(١) جامعة الدول ، الأمانة العامة ، الشؤون الاقتصادية ، الأمانة الفنية لمجلس وزراء الإسكان والتعمير العرب ، عام ٢٠٠٤ ، ص ١٧ ، ١٨ .

جدول رقم (٣)

المعرض من الوحدات السكنية لإجمالي الجمهورية عام ٢٠٠٦

الوحدات السكنية						البيان
مغلقة		خالية		جملة		
عدد	%	عدد	%	عدد	%	
١١٧٥٤١٢	٥٥,١	٣٤٠٤٧١٠	٥٩	٤٥٨٠١٢٢	٥٨	حضر
٩٥٧٦٠٧	٤٤,٩	٢٣٦١٣٥٥	٤١	٣٣١٨٩٦٢	٤٢	ريف
٢١٣٣٠١٩	١٠٠	٥٧٦٦٠٦٥	١٠٠	٧٨٩٩٠٨٤	١٠٠	إجمالي الجمهورية

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، نتائج التعداد العام للسكان والإسكان والمنشآت ، عام ٢٠٠٦ ، ص ١٠٠ .

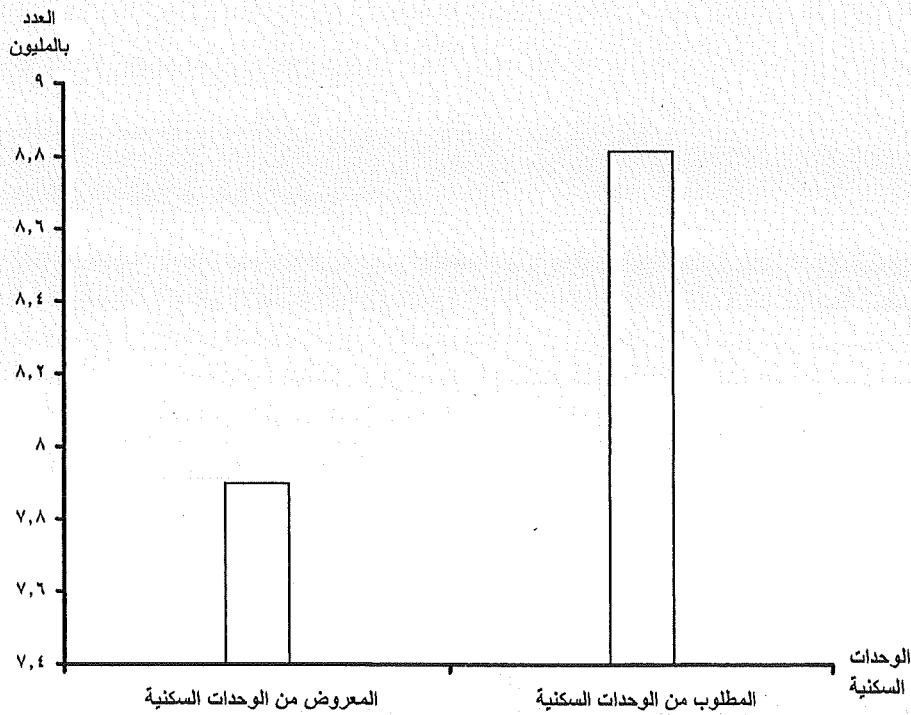


جدول رقم (٤)

تقدير الفجوة الإسكانية لإجمالي الجمهورية عام ٢٠٠٦

البيان	عدد الوحدات السكنية
المعرض من الوحدات السكنية	٧٨٩٩٠٨٤
المطلوب من الوحدات السكنية	٨٨١٤٧٦٤
الفجوة الإسكانية	(٩١٥٦٨٠)

بيانات محسوبة من الجدولين السابقين رقمي (٢) ، (٣) .



عناصر تفاقم المشكلة الإسكانية في مصر

أولاً: موقع الإسكان في قائمة الأولويات القومية.

ثانياً: المتغيرات في الهيكل الاقتصادي و أثرها على تفاقم المشكلة الإسكانية.

ثالثاً: ارتفاع تكلفة المساكن و أثرها على مشكلة الاسكان.

عناصر تفاقم المشكلة الاسكانية في مصر:

يعانى المجتمع المصرى خاصة الذين يعيشون فى المدن الكبرى و لفترة غير قصيرة من عجز شديد فى المساكن المتاحة لمقابلة النمو السكانى ، مع تدهور مستويات الإسكان القائم و عدم توافقه مع البيئة المحيطة به و مع ازدياد حدة العجز و صعوبة العلاج تفاقم المشكلة.

تتركز عناصر المشكلة الإسكانية فى الآتى :

أولاً: موقع الإسكان فى قائمة الأولويات القومية.

ثانياً: المتغيرات فى الهيكل الاقتصادى المصرى و أثرها على تفاقم مشكلة الاسكان.

ثالثاً: ارتفاع تكلفة انشاء المساكن و أثرها على مشكلة الاسكان.

كما سيتضح فان هذه العناصر قد تضافرت و تآثرت كل منها بالأخرى لتتفاقم مشكلة الإسكان فتصل إلى وضعها الراهن.

أولاً: موقع الإسكان فى قائمة الأولويات القومية:

و فيما يلى سوف نوضح أهم الأسباب التى أدت إلى تراجع الأهمية النسبية للإسكان:

١- انخفاض معدلات الاستثمار فى قطاع التشييد و البناء:

لقد تناقص نصيب الإسكان تدريجياً فى اجمالى الاستثمارات و سوف يوضح الجدول رقم (٥) الاستثمار فى قطاع التشييد و البناء و نسبته لاجمالى الاستثمارات خلال الفترة (٢٠٠٧/٢٠٠٨ - ٢٠٠٤/٢٠٠٥) و منه نلاحظ تضائل نصيب قطاع التشييد و البناء من اجمالى الاستثمارات حيث بلغت نسبته ١,٥٪ من اجمالى الاستثمارات فى عام ٢٠٠٤/٢٠٠٥ و زادت زيادة طفيفة فى عام ٢٠٠٥/٢٠٠٦ فبلغت ٢٪ من اجمالى الاستثمارات ثم تناقصت هذه النسبة مرة أخرى إلى أن بلغت ١,٧٪ من اجمالى الاستثمارات فى كل من عامى ٢٠٠٧/٢٠٠٨ ، ٢٠٠٦/٢٠٠٧.

٢- ضعف مساهمة الحكومة فى حل مشكلة السكان :

يوضح الجدول رقم (٦) حجم الاسكان الاقتصادى خلال الفترة (٢٠٠٧ - ١٩٥٢) ومنة يتضح بداية ظهور الاسكان الاقتصادى (منخفضة التكاليف) ومنذ عام ١٩٥٢ حتى عام ١٩٦٠ فى حضر مصر حيث بلغ عدد الوحدات السكنية ٢٢٠٠٠ وحدة وخلال الفترة (١٩٦٠ / ١٩٦٦ - ١٩٧٦ / ٧٧) بدا ظهور الاسكان الاقتصادى فى الريف ايضا حيث بلغ عدد الوحدات السكنية به ١٣٠١٢٥ وحدة فى حين بلغ عدد هذه الوحدات فى الحضر ٢٩٨٠٢٠ وحدة خلال نفس الفترة.

بلغ عدد الوحدات التى يتم تنفيذها ٢٤٦,٤ ألف وحدة سكنية معظمها من الإسكان الاقتصادى وخاصة الإسكان الاقتصادى بالمناطق الريفية حيث بلغ ٢٢٠ ألف وحدة سكنية تمثل ٦٣,٥٪ من اجمالى الوحدات التى لم يتم تنفيذها ، ويمثل الإسكان الاقتصادى بالحضر ٢٤,٥٪ من اجمالى الوحدات .

جدول رقم (٥)

تطور الاستثمار في قطاع التشييد والبناء ونسبته لإجمالي الاستثمارات

خلال الفترة (٢٠٠٧ / ٢٠٠٦ = ٢٠٠٥ / ٢٠٠٤)

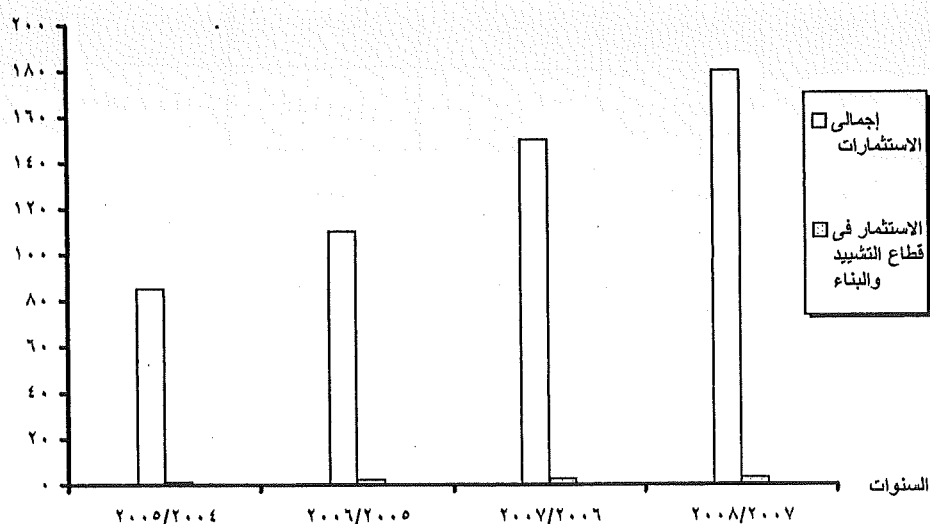
بالمليار جنيه مصري

السنة	إجمالي الاستثمارات	الاستثمار في قطاع التشييد والبناء	% لإجمالي الاستثمارات
٢٠٠٥/٢٠٠٤	٨٥,٠	١,٣	١,٥
٢٠٠٦/٢٠٠٥	١١٠,٠	٢,٢	٢,٠
٢٠٠٧/٢٠٠٦	١٥٠,٠	٢,٥	١,٧
* ٢٠٠٨/٢٠٠٧	١٨٠,٠	٢,٠	١,٧

* متوقع

المصدر: وزارة الدولة للتنمية الاقتصادية , الخطة الخمسية للتنمية الاقتصادية والاجتماعية الخامسة العام الثالث، ص ١٥٤ , العام الرابع ، ص ١٨٤ ، والخطة السادسة لعامها الأول ٢٠٠٧ / ٢٠٠٨ ، ص ٧٥ .

الاستثمارات
بالمليار جنيه



جدول رقم (٦)

حجم الإسكان الإقتصادي خلال الفترة (١٩٥٢ - ٢٠٠٧)

الفترة الزمنية	عدد الوحدات السكنية من الإسكان الإقتصادي (الشعبي)		جملة	% للإجمالي
	حضر	ريف		
١٩٥٢ - ١٩٦٠	٣٢٠٠٠	-	٣٢٠٠٠	١,١
٦٠ / ١٩٦١ - ٦٦ / ١٩٦٧	١٣١٣٤١	٨٠٩٦٣	٢١٢٣٠٤	٧,٢
٦٧ / ١٩٦٨ - ٧١ / ١٩٧٢	١٠٤٥٠٧	٣٨٦١٩	١٤٣١٢٦	٤,٩
٧٢ / ١٩٧٣ - ٧٦ / ١٩٧٧	٦٢٢٢٢	١٠٥٥٣	٧٢٧٧٥	٢,٥
٧٧ / ١٩٧٨ - ٨١ / ١٩٨٢	٣٨١٥٩٨	-	٣٨١٥٩٨	١٣
٨٢ / ١٩٨٣ - ٨٦ / ١٩٨٧	٤٦٤٢٨٦	-	٤٦٤٢٨٦	١٥,٨
٨٧ / ١٩٨٨ - ٩١ / ١٩٩٢	٣٢٤٨٨٧	-	٣٢٤٨٨٧	١١,١
٩٢ / ١٩٩٣ - ٩٦ / ١٩٩٧	٩٨٩٤٠	-	٩٨٩٤٠	٣,٤
٩٧ / ١٩٩٨ - ٠١ / ٢٠٠٢	٢٥٣٦٧٤	-	٢٥٣٦٧٤	٨,٧
٠٢ / ٢٠٠٣ - ٠٦ / ٢٠٠٧	٥٨٥٧٠٠	٣٦٠٠٠٠	٩٤٥٧٠٠	٣٢,٣
الأجمالي	٢٤٣٩١٥٥	٤٩٠١٣٥	٢٩٢٩٢٩٠	١٠٠

المصدر: (١) الجهاز المركزي للتعبئة العامة والإحصاء , الكتاب الإحصائي السنوي , أعداد مختلفة , (١٩٥٢ - ٢٠٠١)

(٢) وزارة الدولة للتنمية الإقتصادية , الخطة الخمسية للتنمية الإقتصادية والإجتماعية

. (٢٠٠٧ / ٢٠٠٨ - ٢٠١١ / ٢٠١٢) وخطة عامها الأول ٢٠٠٧ / ٢٠٠٨ ص ٤٧ .

ثانياً : المتغيرات في الهيكل الاقتصادي المصري وأثرها على تفاقم المشكلة الإسكانية :

سوف نستعرض بشيء من الإيجاز أهم المتغيرات في الهيكل الاقتصادي المصري التي كان لها أثر بالغ في مشكلة الإسكان :

- التضخم الذي حدث بعد حرب ١٩٧٣ .
- التوسع في الأنشطة غير المنتجة لوحدة سكنية ثم الاستيراد والتصدير جعل الناس تتصرف عن الاستثمار في الإسكان .
- اختلال هيكل توزيع الدخل .

يوضح الجدول رقم (٧) متوسط المخصص للانفاق على الإسكان من فئات الدخل عام ٢٠٠٤/ ٢٠٠٥ ومنه نلاحظ أن هذا المخصص للانفاق على الإسكان في فئات الدخل أقل من ٦٠٠٠ جنية ٣٨٩ جنيهاً في السنة أي ٣٢,٤ جنيهاً شهرياً في الحضر ، أما الفئة أقل من ١٧٠٠٠ جنيهاً بلغ المخصص ٤٧٨ جنيهاً سنوياً أي ٣٩,٨ جنيهاً شهرياً . بينما بلغت هذه المخصصات في الريف لنفس الفئات ٣٩٢ جنيهاً سنوياً ، ٣٢,٧ جنيهاً شهرياً ، ٣٧٢ جنيهاً سنوياً ، ٣١,٠ جنيهاً شهرياً على التوالي .

أما اجمالي الجمهورية يبلغ المخصص للانفاق على الإسكان لنفس الفئات (أقل من ٦٠٠٠ جنيهاً سنوياً وأقل من ١٧٠٠٠ جنية سنوياً على التوالي) ٣٩١ جنيهاً سنوياً ، ٣٢,٦ جنيهاً شهرياً ٤٣٦ جنيهاً سنوياً ، ٣٦,٣ جنيهاً شهرياً .

ثالثاً : ارتفاع تكلفة المساكن وأثرها على مشكلة الإسكان :

يعتبر ارتفاع تكاليف البناء عنصراً أساسياً من عناصر المشكلة الإسكانية فهو يؤثر بشكل مباشر على عرض المساكن والطلب عليها فارتفاع التكاليف يعني عجز صغار المستثمرين الذين اعتادوا استثمار أموالهم في الإسكان المتوسط أو الشعبي عن الاستثمار في هذا المجال وبالتالي يفقد الإسكان الاقتصادي مموليه من القطاع الخاص وهناك عوامل تسهم في ارتفاع تكلفة البناء منها ارتفاع أسعار أراضى البناء وكذلك أسعار مواد البناء وهو ما نتناوله بإيجاز .

١ - ارتفاع أسعار الأراضى :

ولقد أثر ارتفاع أسعار الأراضى وعدم توافرها على ازدياد حدة مشكلة الإسكان وزيادة تكلفة المسكن وهذا يؤدي بدون شك الى التأثير على ارتفاع أثمان المساكن بشكل لا يتناسب مع دخل العديد من الفئات الاجتماعية محدودة الدخل بل أيضاً ومتوسطى الدخل .

٢ - ارتفاع أسعار مواد البناء :

نتيجة لزيادة أسعار الوقود والطاقة وبعض السياسات الاحتكارية للمصنّفين والتجار .

٣ - ارتفاع الأجور العينية للعمالة وندرتها :

لقد تعرضت سوق العمالة بقطاع التشييد والبناء لزيادة كبيرة في الطلب عليها سواء داخل مصر أو بالدول العربية وتكون هذه العمالة محدودة فقد زادت أجورها .

٤ - ارتفاع تكاليف البناء :

عجز القطاع العام عن القيام بتنفيذ عمليات البناء قد أدى إلى دخول القطاع الخاص من الباطن واستحوازه على الجزء الأكبر من الاستثمارات في هذا المجال .

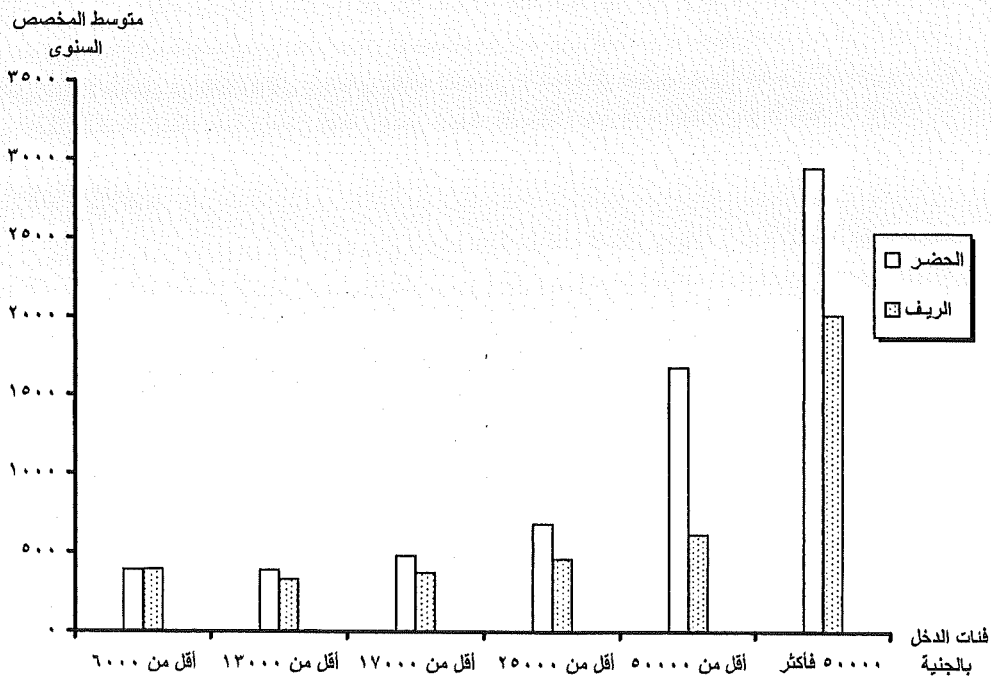
جدول رقم (٧)

متوسط المخصص للإنفاق على الإسكان من فئات الدخل عام ٢٠٠٤ / ٢٠٠٥

القيمة بالجنية

إجمالي الجمهورية		الريف		الحضر		متوسط المخصص
الشهرى	السنوى	الشهرى	السنوى	الشهرى	السنوى	فئات الدخل بالجنية
٣٢,٦	٣٩١	٣٢,٧	٣٩٢	٣٢,٤	٣٨٩	أقل من ٦٠٠٠
٢٩,٣	٣٥٢	٢٧,٤	٣٢٩	٣٢,٢	٣٨٦	أقل من ١٣٠٠٠
٣٦,٣	٤٣٦	٣١,٠	٣٧٢	٣٩,٨	٤٧٨	أقل من ١٧٠٠٠
٥٢,٠	٦٢٤	٣٨,٤	٤٦١	٥٦,٨	٦٨٢	أقل من ٢٥٠٠٠
١٢٩,٢	١٥٥٠	٥١,١	٦١٣	١٣٩,٥	١٦٧٤	أقل من ٥٠٠٠٠
٢٤٧,١	٢٩٦٥	١٦٧,٧	٢٠١٢	٢٤٥,٥	٢٩٤٦	٥٠٠٠٠ فأكثر

المصدر : الجهاز المركزى للتعبئة العامة والإحصاء , بحث الدخل والإنفاق والإستهلاك (بالعينة) ٢٠٠٥ / ٢٠٠٤ , المجلد الرابع , ص ١٨٧ , ١٩٠ , ١٩٣ .



الاحتياجات المستقبلية من الوحدات السكنية

تحدد هذه الاحتياجات سواء كانت حالية أو مستقبلية من خلال نتائج التعدادات السكانية والتي تجرى كل عشرة سنوات وهى تتوقف على معدل النمو السكانى ، وعدد أفراد الأسرة والزيادة السكانية السنوية وكلها مؤشرات يتم استنتاجها من التعدادات السكانية طبقاً لمعادلات ومنتجات ثم التعارف عليها عالمياً وديموجرافياً .

أولاً : منهجية تقدير الاحتياجات المستقبلية من الوحدات السكنية :

يتم تقدير الاحتياجات بناء على تقدير الزيادة فى عدد الأسر وهى تشمل ثلاثة بنود كما سبق وذكرنا فى الفصل الأول وهى :

- ١ - الطلب الناتج عن النمو السكانى بمعنى أن كل أسرة تحتاج لوحدة سكنية .
- ٢ - الطلب بهدف الاحتياطى و المتعارف عليه دولياً بحيث تكون المساكن المطلوبة للاحتياطى فى حدود ٨ % من المساكن اللازمة للنمو السكانى .
- ٣ - المساكن اللازمة للإحلال و التجديد و المتعارف عليه دولياً بأن تكون فى حدود ١ % من المساكن اللازمة للنمو السكانى .

ثانياً : تقدير معدل النمو السكانى و متوسط عدد أفراد الأسرة فى الفترة من ٢٠٢٢ - ٢٠٠٧ :

يتم تحديد الطلب على الإسكان من خلال التعرف على معدل النمو السكانى حتى يمكن التعرف على عدد الأسر التى تحتاج الى وحدات سكنية و لهذا نتناول المؤشرات التالية :

١ . معدل النمو السكانى لاجمالى الجمهورية :

أسفر برنامج تنظيم الأسرة فى مصر عن الانحدار نسبى فى معدل النمو السكانى و يوضح الجدول رقم (٨) هذا الانحدار و تغيره لاجمالى سكان الجمهورية و سوف يستمر هذا الانحدار لعدد من السنوات القادمة نظراً لاستمرار مجهودات مراكز تنظيم الأسرة و الوعى السكانى .

والانحدار الذى حدث فى هذا المعدل بين تعدادى ٢٠٠٦ ، ١٩٩٦ ليصل معدل النمو التقديرى فى الفترة (٢٠١٧ - ٢٠٠٦) إلى ١,٩٨ % ، وأيضاً تم تقدير معدل النمو فى الفترة من (٢٠٢٢ - ٢٠١٧) ليكون ١,٩٦ % .

٢ - التغير فى عدد الأسر و متوسط حجم الأسرة :

يحدث النمو السكانى تغيراً فى عدد الأسر ويرتبط هذا التغير بعاملين أساسيين هما :

أ - معدل النمو الذى يحسب بين تعدادين متتاليين كل عشرة سنوات .

ب - متوسط عدد أفراد الأسرة و الذى يوضحه التعداد السكانى .

يلاحظ من خلال الأربع تعدادات السابقة استمرار انخفاض كل من النمو السكانى و متوسط عدد أفراد الأسر من تعداد لآخر و هذا ما يوضحه الجدول رقم (٩) الذى يتناول أعداد السكان و عدد الأسر و متوسط حجم الأسرة من واقع تعدادى ٢٠٠٦ ، ١٩٩٦ .

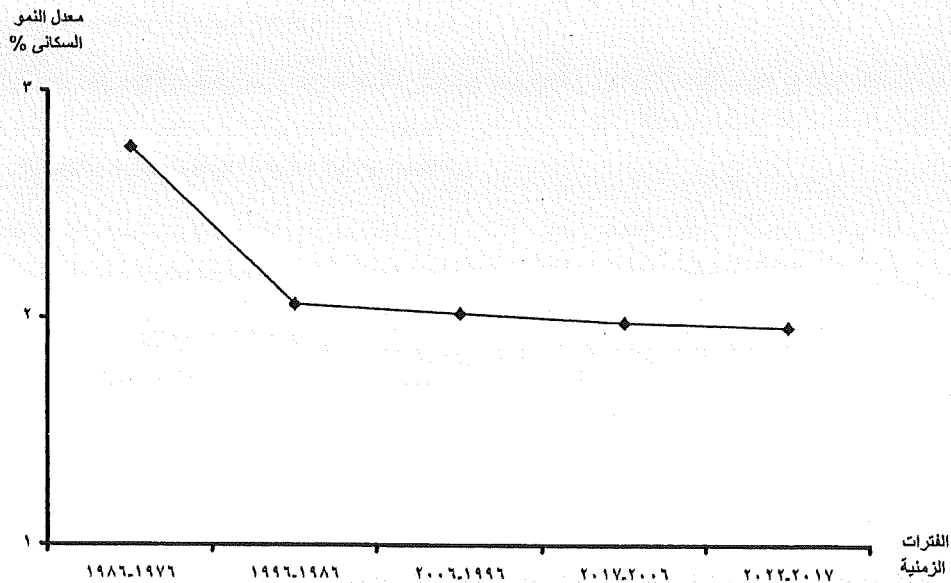
جدول رقم (٨)

أعداد السكان من واقع بيانات التعدادات الفعلية * وتقدير معدل النمو السكاني
لإجمالي الجمهورية خلال الفترة (١٩٧٦ - ٢٠٢٢).

سنوات التعداد	أعداد السكان*
١٩٧٦	٣٦٦٢٦٢٠٤
١٩٨٦	٤٨٢٥٤٢٣٨
١٩٩٦	٥٩٣١٢٩١٤
٢٠٠٦	٧٢٥٧٩٠٣٠
سنوات مابين التعدادات	معدل النمو %
١٩٨٦-١٩٧٦	٢,٧٥
١٩٩٦-١٩٨٦	٢,٠٦
٢٠٠٦-١٩٩٦	٢,٠٢
٢٠١٧ - ٢٠٠٦ (تقديري)	١,٩٨
٢٠٢٢ - ٢٠١٧ (تقديري)	١,٩٦

* السكان داخل الجمهورية

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، التعداد العام للسكان والإسكان والمنشآت عام ١٩٧٦ ، ١٩٨٦ ، ١٩٩٦ ، ونتائج تعداد ٢٠٠٦ .



٢ - متوسط حجم الأسرة :

يعتبر من المؤشرات الهامة التي يمكن بواسطتها تحديد عدد الأسر في مرحلة مستقبلية ويحسب أيضا طبقا لبيانات حقيقية للتعديلات السابقة لهذه المرحلة حيث يوضح التعداد الإجمالي الحقيقي لتعداد السكان وعدد الأسر المقابلة لهذا الإجمالي وبالتالي يمكن تحديد متوسط حجم الأسرة والذي يقدر بقسمة إجمالي السكان على عدد الأسر ونظرا لأنه لا يتغير كثيرا بين عام وآخر فيمكن استخدامه في مرحلة لاحقة لتحديد عدد الأسر لاي عام قادم ، ويوضح الجدول رقم (١٠) تقدير متوسط عدد أفراد الأسرة خلال الفترة (١٩٨٦-٢٠٢٢) على مستوى الجمهورية ، ومما يتبين استمرار انخفاض عدد أفراد الأسرة من ٤,٩ فرد عام ١٩٨٦ إلى ٤,٧ فرد عام ١٩٩٦ ، ٤,٢ فرد عام ٢٠٠٦ ومن المتوقع أن يستمر الانخفاض ليصل متوسط عدد أفراد الأسرة ٣,٩ فردا في الفترة (٢٠٢٢-٢٠٠٧) ، نظرا للتغيرات الاقتصادية وتحول معظم الريف إلى حضر خلال الفترة المذكورة ويعتقد أن هذا المتوسط ان لم ينخفض أو يظل ثابتا على الأقل فأنه لن يرتفع نظرا لما ستفرضه الظروف الاقتصادية والاجتماعية القادمة والمتغيرات الديموجرافية على الشعب المصري والتي تتلخص في ارتفاع تكاليف المعيشة مع زيادة غير ملائمة في الدخل لمقابلة هذه التكاليف.

ثالثا: تقدير الاحتياجات المستقبلية من الوحدات السكنية خلال الفترة (٢٠٢٢-٢٠٠٧) على مستوى محافظات وأقاليم الجمهورية:

- ١- تقدير الاحتياجات المستقبلية من الوحدات السكنية على مستوى الجمهورية:
يوضح الجدول رقم (١١) تقدير الاحتياجات المستقبلية من الوحدات السكنية على مستوى الجمهورية خلال الفترة (٢٠٢٢ = ٢٠٠٧) ولحساب الاحتياجات المستقبلية من الوحدات السكنية تم الأخذ في الاعتبار بعض العوامل التي تتحكم في تقدير التغير لعدد الأسر والتي منها مايلي:
 - ١- معدل النمو الحالي وما سيطرأ عليه خلال الفترة (٢٠٢٢-٢٠٠٧) .
 - ٢- متوسط عدد أفراد الأسر المتوقع خلال الخمسة عشر عاما القادمة.
 - ٣- معدل التزاوج.
 - ٤- تقدير عدد السكان سنويا (٢٠٢٢-٢٠٠٧) .
 - ٥- تقدير عدد الأسر الجدد سنويا (٢٠٢٢-٢٠٠٧) .ونظرا لأنه من الصعب وضع حلا شاملا لمشكلة السكان خلال المرحلة القادمة فسوف نتجاهل حاليا معدل التزاوج حتى يتم تسكين جميع الأسر وحتى يصبح في إمكان كل أسرة الحصول على مسكن ملائم يبدأ بعده النظر إلى تخفيض معدل التزاوج وخصوصا أنه بدراسة متوسط عدد أفراد الأسرة في العشر سنوات السابقة اتضح أنه في تناقص مستمر بما يعنى أن معدل التزاوج يحل تلقائيا .
- ويوضح الجدول رقم (١٢) تقدير الاحتياجات المستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية خلال الفترة (٢٠٢٢ = ٢٠٠٧) ومما يتبين أن عدد الوحدات السكنية المطلوبة من عام ٢٠٠٧ حتى عام ٢٠٢٢ عدد ٧,٥٣٢,٩١٩ وحدة سكنية .

٢- تقدير الاحتياجات المستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية خلال الفترة (٢٠٢٢-٢٠٠٧)

يشمل هذا التحليل تقدير الاحتياجات من الوحدات السكنية على مستوى محافظات وأقاليم الجمهورية ، وتشمل هذه التقديرات ثلاث مكونات للاحتياجات وهى الطلب الناتج عن نمو السكان والطلب بهدف الاحلال والتجديد بنسبة ١٪ من الرصيد القائم والطلب بهدف الاحتياطى بنسبة ٨٪ من المساكن التى تنشأ ،

ويتضح من الجدول رقم (١٢) أن المطلوب توفيره يبلغ حوالى ٧,٥ مليون وحدة سكنية على مستوى أقاليم ومحافظات الجمهورية خلال الفترة (٢٠٢٢ = ٢٠٠٧) منها ٦,٩ مليون وحدة لنمو السكان ٦٩ ألف وحدة للاحلال والتجديد ونحو ٥٥٢ ألف وحدة للاحتياطى وطبقا للتقديرات نجد أن اقليم القاهرة الكبرى يحتاج الى أكبر عدد من الوحدات السكنية حيث تقدر بحوالى ١,٨ مليون وحدة تقريبا بنسبة ٢٤٪ من اجمالي الوحدات المطلوبة ليلية اقليم

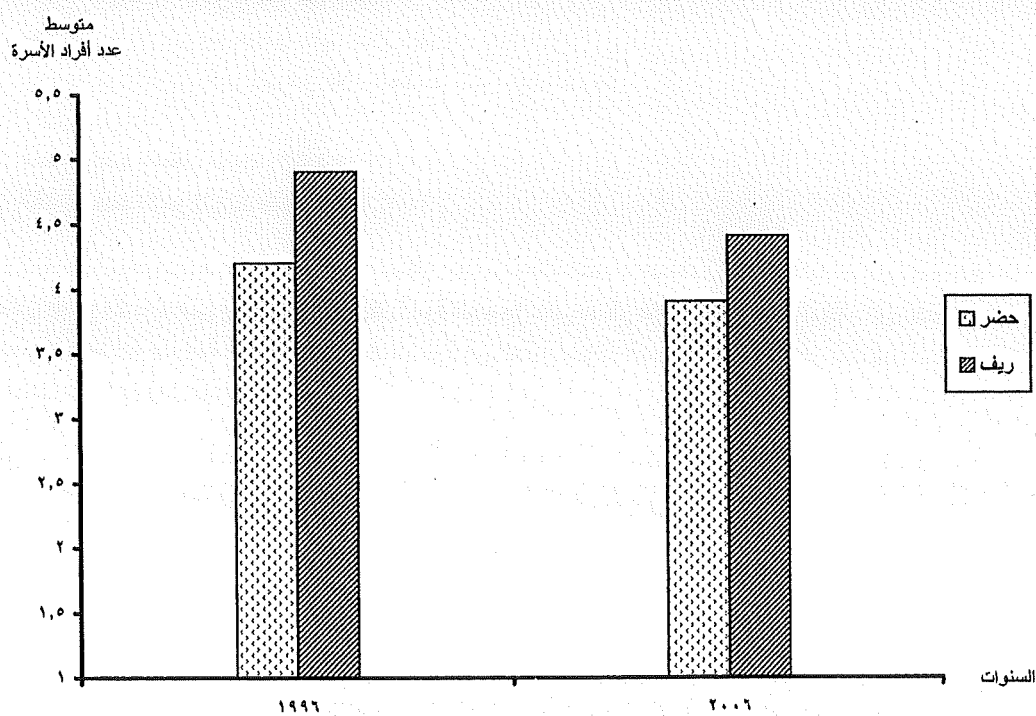
جدول رقم (٩)

أعداد السكان وعدد الأسر ومتوسط حجم الأسرة من واقع تعدادي ٢٠٠٦ ، ١٩٩٦

البيان	تعداد ١٩٩٦			تعداد ٢٠٠٦			الزيادة في عدد الأسر ١٩٩٦ - ٢٠٠٦
	عدد السكان*	عدد الأسر	متوسط عدد أفراد الأسرة	عدد السكان*	عدد الأسر	متوسط عدد أفراد الأسرة	
حضر	٢٥٢٨٦٣٣٥	٥٨٣٩٨٧٧	٤,٣	٣٠٩٤٩٦٨٩	٧٧٥١٥١٢	٣,٩	١٩١١٦٣٥
ريف	٣٤٠٢٦٥٧٩	٦٨٦٢٧٢٣	٤,٩	٤١٦٢٩٣٤١	٩٥١٤٠٥٥	٤,٤	٢٦٥١٣٣٢
جملة	٥٩٣١٢٩١٤	١٢٧٠٢٦٠٠	٤,٧	٧٢٥٧٩٠٣٠	١٧٢٦٥٥٦٧	٤,٢	٤٥٦٢٩٦٧

* السكان داخل الجمهورية

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء، نتائج التعداد العام للسكان والإسكان والمنشآت ، مرجع سابق ، ص ١٤٢.



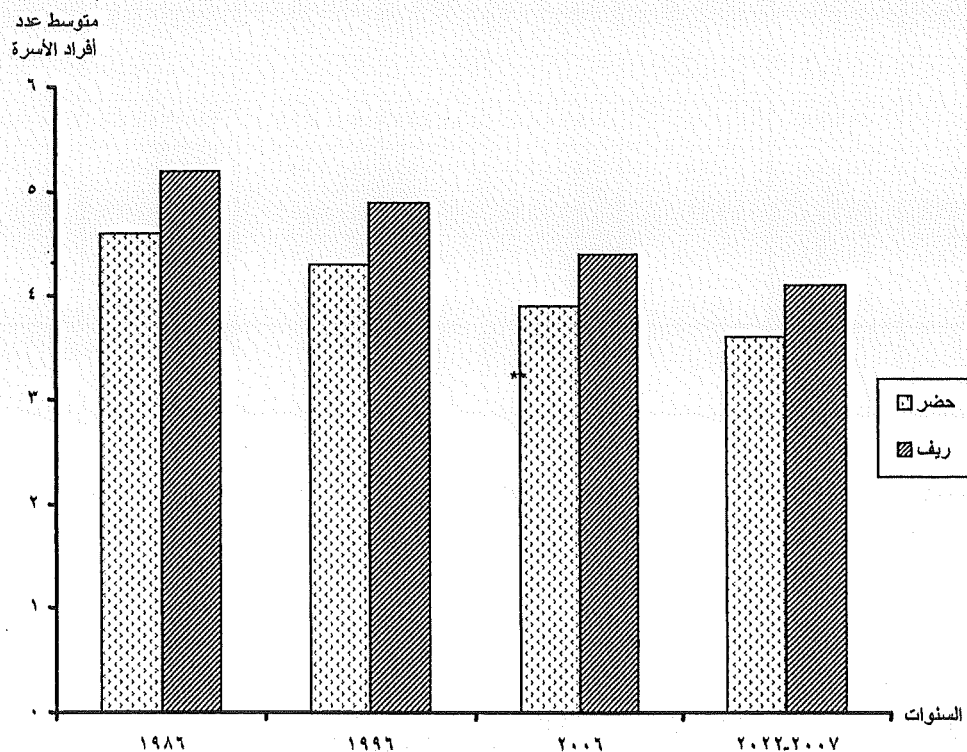
جدول رقم (١٠)

تقدير متوسط عدد أفراد الأسرة خلال الفترة ٢٠٢٢ - ١٩٨٦ على مستوى الجمهورية

البيان	متوسط عدد أفراد الأسرة			
	٢٠٠٧-٢٠٢٢ *	٢٠٠٦	١٩٩٦	١٩٨٦
حضر	٣,٦	٣,٩	٤,٣	٤,٦
ريف	٤,١	٤,٤	٤,٩	٥,٢
جملة	٣,٩	٤,٢	٤,٧	٤,٩

* يفرض ثبات متوسط عدد أفراد الأسرة خلال الفترة من ٢٠٢٢-٢٠٠٧ .

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، نتائج التعداد العام للسكان والإسكان والمنشآت ، مرجع سابق ، ص ١٤٢ .



جدول رقم (١١)

تقدير الاحتياجات المستقبلية من الوحدات السكنية على مستوى الجمهورية خلال الفترة (٢٠٠٧ - ٢٠٢٢)

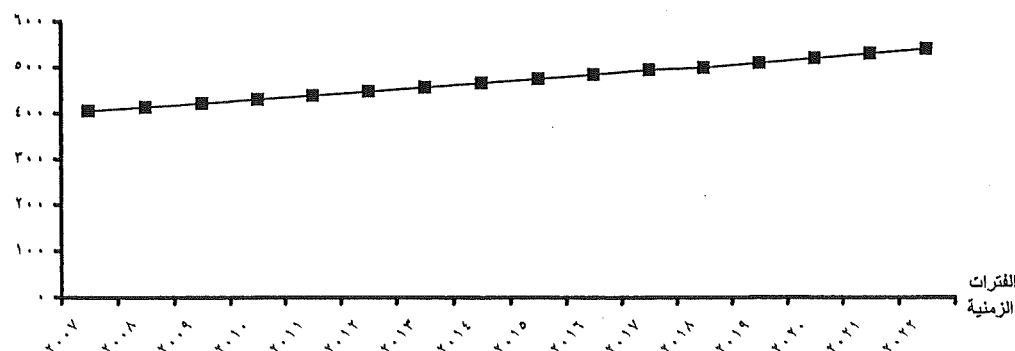
السنوات	عدد السكان المتوقع *	الزيادة في عدد السكان	الزيادة في عدد الأسر	عدد الوحدات المطلوبة لنمو السكان	عدد الوحدات المطلوبة للإحلال والتجديد	عدد الوحدات المطلوبة للإحتياطي	إجمالي عدد الوحدات المطلوبة
٢٠٠٦	٧٢٥٧٩٠٣٠**	-	-	-	-	-	-
٢٠٠٧	٧٤٠٣٠٤١٦	١٤٥١٣٨٦	٣٧٢١٥٠	٣٧٢١٥٠	٣٧٢١٥٠	٢٩٧٧٢	٤٠٥٦٤٣
٢٠٠٨	٧٥٥١٠٨٢٦	١٤٨٠٤١٠	٣٧٩٥٩٢	٣٧٩٥٩٢	٣٧٩٥٩٢	٣٠٣٦٧	٤١٣٧٥٥
٢٠٠٩	٧٧٠٢٠٨٤٠	١٥١٠٠١٤	٣٨٧١٨٣	٣٨٧١٨٣	٣٨٧١٨٣	٣٠٩٧٥	٤٢٢٠٣٠
٢٠١٠	٧٨٥٦١٠٥٠	١٥٤٠٢١٠	٣٩٤٩٢٦	٣٩٤٩٢٦	٣٩٤٩٢٦	٣١٥٩٤	٤٣٠٤٦٩
٢٠١١	٨٠١٣٢٠٦٠	١٥٧١٠١٠	٤٠٢٨٢٣	٤٠٢٨٢٣	٤٠٢٨٢٣	٣٢٢٢٦	٤٣٩٠٧٧
٢٠١٢	٨١٧٣٤٤٨٦	١٦٠٢٤٢٦	٤١٠٨٧٨	٤١٠٨٧٨	٤١٠٨٧٨	٤١٠٩	٤٤٧٨٥٧
٢٠١٣	٨٣٣٦٨٩٥٧	١٦٣٤٤٧١	٤١٩٠٩٥	٤١٩٠٩٥	٤١٩٠٩٥	٤١٩١	٤٥٦٨١٤
٢٠١٤	٨٥٠٣٦١١٣	١٦٦٧١٥٦	٤٢٧٤٧٦	٤٢٧٤٧٦	٤٢٧٤٧٦	٤٢٧٥	٤٦٥٩٤٩
٢٠١٥	٨٦٧٣٦٦٠٧	١٧٠٠٤٩٤	٤٣٦٠٢٤	٤٣٦٠٢٤	٤٣٦٠٢٤	٤٣٦٠	٤٧٥٢٦٦
٢٠١٦	٨٨٤٧١١٠٧	١٧٣٤٥٠٠	٤٤٤٧٤٤	٤٤٤٧٤٤	٤٤٤٧٤٤	٤٤٤٧	٤٨٤٧٧١
٢٠١٧	٩٠٢٤٠٢٩٢	١٧٦٩١٨٥	٤٥٣٦٣٧	٤٥٣٦٣٧	٤٥٣٦٣٧	٤٥٣٦	٤٩٤٤٦٤
٢٠١٨	٩٢٠٢٦٤٤٩	١٧٨٦١٥٧	٤٥٧٩٨٩	٤٥٧٩٨٩	٤٥٧٩٨٩	٤٥٨٠	٤٩٩٢٠٨
٢٠١٩	٩٣٨٤٧٩٦٠	١٨٢١٥١١	٤٦٧٠٥٤	٤٦٧٠٥٤	٤٦٧٠٥٤	٤٦٧١	٥٠٩٠٨٩
٢٠٢٠	٩٥٧٠٥٥٢٥	١٨٥٧٥٦٥	٤٧٦٢٩٩	٤٧٦٢٩٩	٤٧٦٢٩٩	٤٧٦٣	٥١٩١٦٦
٢٠٢١	٩٧٥٩٩٨٥٧	١٨٩٤٣٣٢	٤٨٥٧٢٦	٤٨٥٧٢٦	٤٨٥٧٢٦	٤٨٥٧	٥٢٩٤٤١
٢٠٢٢	٩٩٥٣١٦٨٤	١٩٣١٨٢٧	٤٩٥٣٤٠	٤٩٥٣٤٠	٤٩٥٣٤٠	٤٩٥٣	٥٣٩٩٢٠
الإجمالي العام من ٢٠٠٧ - ٢٠٢٢		-	-	-	-	٥٥٢٨٧٥	٧٥٣٢٩١٩

* السكان داخل الجمهورية

** فعلى طبقاً نتائج تعداد ٢٠٠٦

بيانات تقديرية محسوبة .

الوحدات السكنية
المطلوبة بالآلاف



الدلتا الذى يحتاج الى حوالى ١,٢ مليون وحدة ينسبة ١٧,٣٪ ثم إقليم القناة ويحتاج الى ١,٢ مليون وحدة بنسبة ١٦ ٪ وأقل الأقاليم طلبا للوحدات السكنية هو إقليم أسيوط حيث يحتاج الى حوالى ٢٥٤ ألف وحدة بنسبة ٤,٧ ٪ من اجمالى الوحدات المطلوبة .

جدير بالذكر أن الدولة قامت بإنشاء قرى جديدة واستصلاح مساحات جديدة لإيجاد فرص عمل للمواطنين فى مجال الزراعة والصناعة وتعمير الساحل الشمالى الملاصق لمدينة الإسكندرية وخصوصاً التوسع فى مدينة برج العرب الصناعية سوف يجعل محافظة الإسكندرية وما يتبعها من مدن سياحية بالساحل منطقة جذب سكانى فى الفترة القادمة وهذا سيرفع معدل النمو السكانى ليصل إلى ٢ ٪ فى الفترة (٢٠٢٢ = ٢٠٠٧) ، وكذلك محافظة البحيرة حيث يتم التوسع المستمر فى المناطق الصحراوية الغربية من المحافظة مما سيرفع معدل النمو السكانى خلال نفس الفترة القادمة إلى ١,٨ ٪ . كما قامت الدولة بدراسة المحافظات التى لها ظهير صحراوى حيث تم استصلاح وبيع هذا الظهير مما سيجعل المناطق الصحراوية بهذه المحافظات تتحول إلى مناطق زراعية وبالتالي أصبحت مناطق جذب لأبناء هذه المحافظات الذين تركوها إلى محافظات كانت جاذبة لهم ليعودوا إلى محافظتهم وتعمير المناطق الصحراوية الجديدة . وقد وضعت الدولة فى خطتها خلال الخمس سنوات القادمة بناء ٤٠٠ قرية فى الظهير الصحراوى لقرى الصعيد لتستوعب ٥ مليون نسمة فى هذه المناطق الجديدة وتخلق ٢٨٠ ألف فرصة عمل . ولهذه الأسباب فإنه يتوقع أن معدل النمو السكانى فى جميع محافظات الصعيد لن ينخفض وإن لم يرتفع فسوف يظل ثابتاً خلال الخمسة عشر عاماً القادمة على أقل تقدير .

٢ - تقدير الاحتياجات الحالية والمستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية حتى عام ٢٠٢٢ :

يوضح الجدول رقم (١٢) تقدير اجمالى الاحتياجات الحالية والمستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية حتى عام ٢٠٢٢ ، ومنه يتضح تقدير المحصلة النهائية لإجمالى الاحتياجات الحالية والمستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية حيث إتضح من تقييم الاحتياجات الحالية لكل محافظات الجمهورية أن هناك فائض وحدات فى كثير من المحافظات .

بينما يوجد عجز فى محافظات أخرى . وقد تم أخذ هذا فى الاعتبار عند حساب الاحتياجات المستقبلية حيث أضيفت الاحتياجات الحالية للمستقبلية فى المحافظات التى بها عجز حالياً (٢٠٠٦) بينما تم خصم فائض أعداد الوحدات الحالية من أعداد الاحتياجات المستقبلية فى المحافظات التى بها فائض حالياً (٢٠٠٦) .

وقد أظهرت النتيجة النهائية للخصم والإضافة أن اجمالى الاحتياجات الحالية والمستقبلية من الوحدات السكنية حتى عام ٢٠٢٢ تقدر ٨,٤ مليون وحدة منها ٩١٦ ألف وحدة حتى عام ٢٠٠٦ ، ٧,٥ مليون وحدة خلال الفترة ٢٠٠٧ = ٢٠٢٢) . وذلك لمقابلة الزيادة فى النمو السكانى وبالتالي أعداد الأسر خلال هذه الفترة طبقاً لمعدلات النمو المحسوبة وتقدير عدد السكان .

جدول رقم (١٢)

تقدير الاحتياجات المستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية

خلال الفترة (٢٠٢٢-٢٠٠٧)

المحافظات	عدد الوحدات المطلوبة لزيادة الأسر	عدد الوحدات المطلوبة للأحلال والتجديد	عدد الوحدات المطلوبة للإحتياطي	إجمالي عدد الوحدات المطلوبة
القاهرة	٢٥٤٥٢٠	٢٥٤٥	٢٠٢٦٢	٢٧٧٤٢٧
الجيزة	٨٧٦٢٤٧	٨٧٦٢	٧٠١٠١	٩٥٥١١٠
القليوبية	٥٢٠١٠١	٥١٩٩	٤١٦٠٩	٥٦٦٩٠٩
إجمالي إقليم القاهرة الكبرى*	١٦٥٠٨٦٨	١٦٥٠٦	١٢٢٠٧٢	١٧٩٩٤٤٦
الإسكندرية	٤٢٩٠٥٠	٤٢٩٠	٢٤٢٢٥	٤٦٧٦٦٥
البحيرة	٤٢٧٢٢٤	٤٢٧١	٢٤١٧٩	٤٦٥٦٧٤
مطروح	٦٠٩٠٠	٦٠٩	٤٨٧١	٦٦٢٨٠
إجمالي إقليم الإسكندرية	٩١٧١٧٤	٩١٢٠	٧٢٢٧٥	٩٩٩٧١٩
الشرقية	٦٢١٩٩٦	٦٢٢٠	٤٩٧٦١	٦٧٧٩٧٧
بور سعيد	٥٢٧٨٢	٥٢٩	٤٢٠٢	٥٨٦٢٣
الإسماعيلية	١٤٢٤٥٠	١٤٢٥	١١٢٩٦	١٥٥٢٧١
السويس	٥١٠٨٠	٥١٢	٤٠٨٥	٥٥٦٧٧
شمال سيناء	٥١٧٣٧	٥١٧	٤١٢٨	٥٦٣٩٢
جنوب سيناء	١٥٦٢٠١	١٥٦٢	١٢٤٩٥	١٧٠٢٥٨
إجمالي إقليم القناة	١٠٧٧٢٤٦	١٠٧٧٥	٨٦١٧٧	١١٧٤١٩٨
المنوفية	١٩١٩٧٢	١٩٢٠	١٥٢٥٧	٢٠٩٢٤٩
دمياط	٩٦٧٥٠	٩٦٨	٧٧٤٠	١٠٥٤٥٨
الغربية	٢٠٨٣٦٢	٢٠٨٤	٢٤٦٦٩	٢٣٦١١٥
كفر الشيخ	١٩٩٥٦٢	١٩٩٥	١٥٩٦٦	٢١٧٥٢٤
الدقهلية	٤٠٥٨٢٠	٤٠٥٩	٢٢٤٦٦	٤٤٢٣٤٥
إجمالي إقليم الدلتا	١٢٠٢٤٦٧	١٢٠٢٦	٩٦١٩٨	١٢١٠٦٩١

* إقليم القاهرة الكبرى قبل تعديل المحافظات كان يشمل محافظات القاهرة والجيزة والقليوبية وبعد تعديل المحافظات أصبح يشمل نفس المساحة بتقسيم جديد وهو محافظة القاهرة والجيزة والقليوبية والسادس من أكتوبر وحلوان .

(تابع) جدول رقم (١٢)

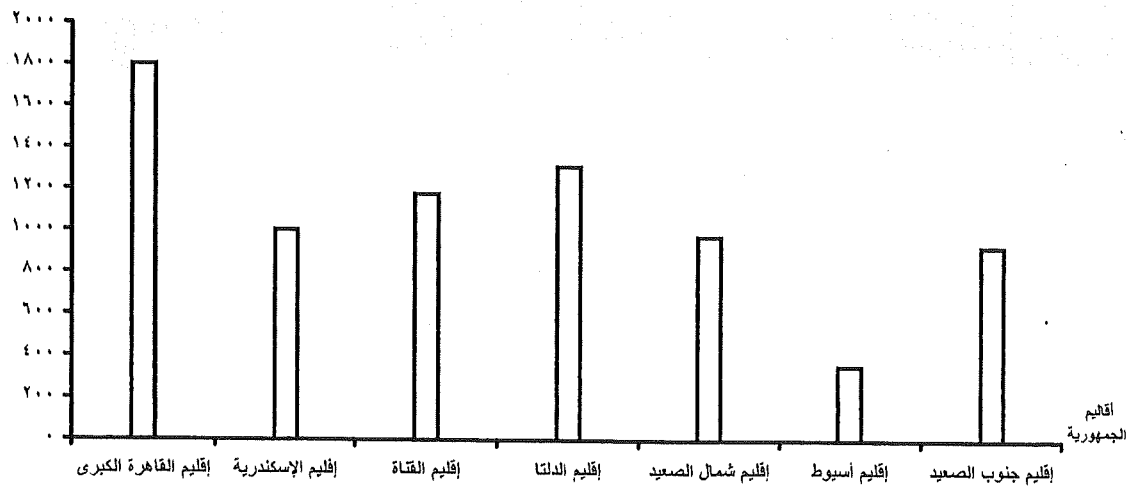
تقدير الاحتياجات المستقبلية من الوحدات السكنية بمحافظات وأقاليم الجمهورية

خلال الفترة (٢٠٠٧ - ٢٠٢٢)

المحافظات	عدد الوحدات المطلوبة لزيادة الأسر	عدد الوحدات المطلوبة للإحلال والتجديد	عدد الوحدات المطلوبة للإحتياطي	إجمالي عدد الوحدات المطلوبة
بنى سويف	٢٠٢٩٣٤	٢٠٣٠	١٦٢٣٦	٢٢١٢٠٠
المنيا	٤٣٤٦٣٠	٤٣٤٨	٣٤٧٧٢	٤٧٢٧٥٠
الفيوم	٢٥٣٨٧٩	٢٥٣٩	٢٠٣١١	٢٧٦٧٢٩
إجمالي إقليم شمال الصعيد	٨٩١٤٤٣	٨٩١٧	٧١٣١٩	٩٧١٦٧٩
أسيوط	٢٩٧٧٨٠	٢٩٧٧	٢٣٨٢٣	٣٢٤٥٨٠
الوادى الجديد	٢٦٧٢٨	٢٦٧	٢١٣٩	٢٩١٤٤
إجمالي إقليم أسيوط	٣٢٤٥١٨	٣٢٤٤	٢٥٩٦٢	٣٥٣٧٢٤
سوهاج	٢٦٩١٩٨	٢٦٩٢	٢١٥٣٤	٢٩٢٤٢٤
قنا	٢٩٠٩٢٢	٢٩٠٩	٢٣٢٧٣	٣١٧١٠٤
الأقصر	٤١٤٥١	٤١٤	٣٣١٥	٤٥١٨٠
أسوان	١١٦٩٤٥	١١٦٨	٩٣٥٤	١٢٧٤٦٧
البحر الأحمر	١٢٨٧٠٤	١٢٨٧	١٠٢٩٦	١٤٠٢٨٧
إجمالي إقليم جنوب الصعيد	٨٤٧٢٢٠	٨٤٧٠	٦٧٧٧٢	٩٢٣٤٦٢
إجمالي الجمهورية	٦٩١٠٩٣٦	٦٩١٠٨	٥٥٢٨٧٥	٧٥٢٢٩١٩

بيانات تقديرية محسوبة .

إجمالي عدد الوحدات
المطلوبة (بالآلاف)



جدول رقم (١٢)

تقدير إجمالي الاحتياجات الحالية والمستقبلية من الوحدات السكنية بمحافظة

وأقاليم الجمهورية حتى عام ٢٠٢٢

المحافظات	الاحتياجات* الحالية من الوحدات السكنية عام ٢٠٠٦ (الفجوة الإسكانية)	الاحتياجات المستقبلية من الوحدات السكنية (٢٠٠٧ = ٢٠٢٢)	إجمالي عدد الوحدات الحالية والمستقبلية من الوحدات السكنية
القاهرة	(١٣٦٨٩٤)	٢٧٧٤٢٧	٤١٤٣٢١
الجيزة	٥٩٢٤١	٩٥٥١١٠	٨٩٥٨٦٩
القليوبية	٩٢٥٩٠	٥٦٦٩٠٩	٤٧٤٣١٩
إجمالي اقليم القاهرة الكبرى	١٤٩٣٧	١٧٩٩٤٤٦	١٧٨٤٥٠٩
الإسكندرية	١٥٣١٥٢	٤٦٧٦٦٥	٢١٤٥١٢
البحيرة	(٢١٧٠٧٢)	٤٦٥٦٧٤	٦٨٢٧٤٦
مطروح	٢٦٦١١	٦٦٢٨٠	٢٩٧٦٩
إجمالي اقليم الاسكندرية	(٢٧٣٠٨)	٩٩٩٧١٩	١٠٣٧٠٢٧
الشرقية	(٢٨١٩٢)	٦٧٧٩٧٧	٧١٦١٦٩
بور سعيد	(٢٧٢٦٨)	٥٨٦٢٣	٨٥٩٩١
الإسماعيلية	(١٣٤٧٧)	١٥٥٢٧١	١٦٨٧٤٨
السويس	(٦٢٩٣)	٥٥٦٧٧	٦١٩٧٠
شمال سيناء	٦٦٦٩	٥٦٣٩٢	٤٩٧٢٢
جنوب سيناء	(١٢٠٤٤)	١٧٠٢٥٨	١٨٢٣٠٢
إجمالي اقليم القناة	(٩٠٧٠٥)	١١٧٤١٩٨	١٢٦٤٩٠٣
المنوفية	(٣١٨٨٥)	٢٠٩٢٤٩	٢٤١١٢٤
دمياط	٢٨٧٩٠	١٠٥٤٥٨	٦٦٦٦٨
الغربية	(٤٤٦٢٥)	٢٣٦١١٥	٢٨٠٧٤٠
كفر الشيخ	(٨٦٥٩٦)	٢١٧٥٢٤	٣٠٤١٢٠
الدقهلية	(٩٨٨)	٤٤٢٣٤٥	٤٤٢٣٣٣
إجمالي اقليم الدلتا	(١٢٥٣٠٤)	١٣١٠٦٩١	١٤٣٥٩٩٥

(تابع) جدول رقم (١٢)

تقدير إجمالي الاحتياجات الحالية والمستقبلية من الوحدات السكنية بمحافظة

وأقاليم الجمهورية حتى عام ٢٠٢٢

المحافظات	الإحتياجات الحالية* من الوحدات السكنية عام ٢٠٠٦ (الفجوة الإسكانية)	الإحتياجات المستقبلية من الوحدات السكنية (٢٠٠٧ = ٢٠٢٢)	إجمالي عدد الوحدات الحالية والمستقبلية من الوحدات السكنية
بنى سويف	(٧٨٤٧٨)	٢٢١٢٠٠	٢٩٩٦٧٨
المنيا	(١٧٥٤٨٧)	٤٧٢٧٥٠	٦٤٩٢٣٧
الفيوم	(٩٥٣٥٩)	٢٧٦٧٢٩	٢٧٢٠٨٨
إجمالي إقليم شمال الصعيد	(٣٤٩٢٢٤)	٩٧١٦٧٩	١٣٢١٠٠٢
أسيوط	(١٠١٢٥٥)	٢٢٤٥٨٠	٤٢٥٩٣٥
الوادى الجديد	(٢٤٤٠)	٢٩١٤٤	٢١٥٨٤
إجمالي إقليم أسيوط	(١٠٣٧٩٥)	٢٥٣٧٢٤	٤٥٧٥١٩
سوهاج	(٥٨٦٨٢)	٢٩٣٤٢٤	٢٥٢١٠٧
قنا	(١٢٠٥٣٦)	٢١٧١٠٤	٤٣٧٦٤٠
الأقصر	(٦٧١٦)	٤٥١٨٠	٥١٨٩٦
أسوان	(٧٢١٨٥)	١٢٧٤٦٧	١٩٩٦٥٢
البحر الأحمر	٢٣٩٢٩	١٤٠٢٨٧	١٠٦٣٤٨
إجمالي إقليم جنوب الصعيد	(٢٢٤١٨١)	٩٢٣٤٦٢	١١٤٧٦٤٢
إجمالي الجمهورية	(٩١٥٦٨٠)	٧٥٢٢٩١٩	٨٤٤٨٥٩٩

* بيانات محسوبة من واقع نتائج التعداد العام للسكان والاسكان والمنشآت ، اجمالى الجمهورية ، عام ٢٠٠٦ ، و جدول رقم ١٢ .

التوصيات

فضلا عن ما تم ذكره سابقاً عن الوضع الحالى للإسكان فى مصر (الفصل الأول) والتبايع الدولة عدة سياسات واجراءات تسهم فى توفير السكن الملائم لطالبيه ومع تفاقم المشكلة الإسكانية فإن مواجهة ذلك ينصب حول بناء الخطط المستقبلية الخاصة بالإسكان والتعمير للعمل على تصحيح مسار النشاط الإقتصادى الموجه لقطاع الإسكان وتكثيف وترشيد الأدوار التى تسهم فى التخفيف من حدة المشكلة وذلك على النحو التالى :

أولاً : استراتيجيات عامة ورئيسية :

- ١ - التوجه نحو توفير السكن الإقتصادى للأسر محدودة الدخل والأولوية لسكانى الخيام والمقابر وأيضاً المتزوجين حديثاً وسكانى المساكن المتدهورة .
- ٢ - تحديد الإحتياجات وحصر الإمكانات حتى يمكن تحديد الأهداف ووضع الإستراتيجية وفى ضوء ذلك توضع الخطة الإسكانية .
- ٣ - التوسع فى إنشاء مدن جديدة فى الظهير الصحراوى للمدن القديمة مع توفير الإسكان الإقتصادى لخلق مجتمعات جديدة .
- ٤ - ربط المناطق الجديدة بوسائل نقل وشبكة طرق مع الوادى القديم لتشجيع الإقامة فى تلك المدن وإستغلال المباني الموجودة بها .
- ٥ - وضع سياسة جديدة وفرض رسوم مالية على الوحدات المغلقة والخالية بصفة دائمة لإجبار أصحابها على إستغلالها بالبيع أوالتأجير .
- ٦ - إعداد مخططات عامة للمدن والقرى وإعادة تعمير وصيانة المناطق المتدهورة وزيادة إرتفاعات بعض المباني فى الحدود المسموح بها .
- ٧ - بناء الشراكة بين القطاعين العام والخاص بهدف توفير الإسكان للفئات محدودة الدخل وأيضاً تذليل العقبات أمام القطاع الخاص لزيادة إستثماراته فى بناء الوحدات الإقتصادية :

ثانياً : الإستثمارات :

- ١ - زيادة نسبة الإستثمارات المخصصة لقطاع الإسكان فى الخطة القومية لتوفير السكن الملائم بالسعر المناسب
- ٢ - توجيه الإستثمارات والمساعدات المخصصة لمشروع لله ابنى بيتك لله إلى المناطق المستصلحة وزيادة عدد قطع الأراضى ومساحة القطعة لتستوعب أعداداً كبيرة من الشباب محدودى الدخل .
- ٣ - التشجيع على تخصيص جزء أكبر من الموارد الاستثمارية لصالح الإسكان وأيضاً تخصيص المعونات التى تقدمها الدول الصديقة لمصر لتمويل مشروعات الإسكان والبنية التحتية .
- ٤ - ترشيد برامج التمويل العقارى وتطويرها لتناسب كل فئات الدخل وليس فقط فئات الدخل المتوسط والمرتفع ، ويمكن أن تتضمن مخططات التمويل العقارى الجديدة أنظمة الشراء من أجل الإيجار والملكية المشتركة .
- ٥ - تقليل سعر الفائدة على القروض العقارية بحيث لا تزيد عن فائدة الإيداع بأكثر من ١٪ لتشجيع عملية البناء .

ثالثاً : الأراضى :

١ - وضع سياسة خاصة بالأراضى تتيح إمكانية توفير أراضى معدة ومجهزة للبناء بأسعار ميسرة لتشجيع القطاع الخاص للدخول فى مجال بناء المساكن الإقتصادية .

٢ - يجب أن تعطى للأفراد أو الشركات أراضى مناسبة للبناء بأسعار ميسرة وذلك للبناء وليس للإحتفاظ بها ويمكن منح حق الإنتفاع بالأراضى فى حالة البناء عليها لمدة زمنى طويل يعادل العمر الافتراضى للمبنى مقابل رسوم محددة .

رابعاً : مواد البناء :

١ - زيادة معدلات الإستثمار المخصصة لصناعة مواد البناء حتى يمكن توفيرها بالقدر الذى يكفى السوق المحلية ويقلل من تكلفة إنشاء الوحدات الجديدة .

٢ - تتدخل الدولة لمنع الإحتكار فى مواد البناء خاصة حديد التسليح والأسمنت الذى وصل إلى أسعار تؤدى إلى إرتفاع تكلفة البناء .

خامساً : العمالة المدربة :

١ - التوسع فى إنشاء مراكز التدريب على التشييد والبناء وتشجيع الشباب على الدخول فى هذه المراكز لتوفير العمالة المدربة .

٢ - تنظيم هجرة العمالة للخارج ويعنى ذلك أن يؤخذ فى الاعتبار إحتياجات خطة التنمية الإقتصادية من عمالة البناء والتشييد ثم السماح بهجرة ما يزيد عن حاجة الخطة .

سادساً : التشريعات :

وضع تشريعات وقوانين محددة لقطاع الإسكان لأن كثرة التشريعات والقوانين أدى إلى عدم التشجيع فى مجال الإستثمار الإسكانى خاصة المتوسط والاقتصادى إلى جانب عدم الاهتمام بصيانة العقارات القائمة مما أدى إلى تدهور الثروة العقارية ، كما أن التشريعات المتعاقبة المتصلة بالإسكان لم تقض على المشكلة بل أدت إلى تفاقمها .

سابعاً : البحث العلمى :

تولى الجهات المختصة بالبحث العلمى الاهتمام بمجال بحوث البناء بأن تدعم وتشجع مراكز البحوث ذات الصلة لإعداد دراسات مفيدة حول إبتكار مواد بناء جديدة تعتمد على الموارد المحلية غير باهظة التكاليف .

ثامناً : مقترحات للدراسة :

١ - السماح بالبناء على كامل مساحة الأرض المخصصة للبناء دون تحديد نسبة كما هو يحدث حالياً فى عديد من المناطق على أن يخصص نسبة من العقار للتأجير .

٢ - السماح بهدم المباني القديمة والخالية من السكان (غير الأثرية ولا تدخل فى المحميات) وإعادة بناءها على أن يتعهد المالك بتأجير نسبة يتفق عليها .

وبهذين الاقتراحين يمكن أن نرى ولو تدريجياً - لافتة " شقة للإيجار " .

المراجع

أولاً : التقارير :

- ١ - جامعة الدول العربية ، الأمانة العامة ، الشئون الاقتصادية ، تقرير وتوصيات إجتماع أجهزة الإحصاء بالدول العربية ، ٢٠٠٤ .
- ٢ - مجلس الشورى ، مشكلة الإسكان فى مصر ، تقرير رقم (٢٠) ، ١٩٨٢ .
- ٣ - وزارة الدولة للتنمية الاقتصادية ، الخطة الخمسية للتنمية الاقتصادية والاجتماعية (٢٠٠٢/٢٠٠٣) - (٢٠٠٦/٢٠٠٧) ، إبريل ٢٠٠٢ .
- ٤ - وزارة الدولة للتنمية الاقتصادية ، الخطة الخمسية للتنمية الاقتصادية والاجتماعية (٢٠٠١/٢٠٠٢) - (٢٠٠٧/٢٠٠٨) وخطة عامها الأول (٢٠٠٧/٢٠٠٨) ، إبريل ٢٠٠٧
- ٥ - طه محمد عبد المطلب ، الظهير الصحراوى واتساع رحاب التنمية ، جريدة الأهرام الصادرة فى ٢٠٠٧/١/١٠ .

ثانياً : الدراسات والبحوث :

- ١ - حازم كمال وآخرون ، برنامج الإسكان المدعم فى مصر والدروس المستفادة ، برنامج الدعم الفنى لإصلاح السياسات ، فبراير ٢٠٠٧ .
- ٢ - الوكالة الأمريكية للتنمية الدولية ، دراسة الطلب على الإسكان فى جمهورية مصر العربية ، يوليو ٢٠٠٦ .
- ٣ - الوكالة الأمريكية للتنمية الدولية ، دراسة الطلب على الإسكان فى القاهرة الكبرى ، يونيو ٢٠٠٧ .
- ٤ - وزارة الإسكان والمرافق والمجتمعات العمرانية ، الهيئة العامة للتخطيط العمرانى ، المرصد الحضرى الوطنى ، دراسة استرشادية حول تجربة مصر فى تقدير الاحتياجات الإسكانية بأستخدام مؤشرات ، أكتوبر ٢٠٠٥ .

ثالثاً : الكتب :

- ١ - أحمد خالد علام وآخرون ، مشكلة الإسكان فى مصر ، مكتبة دار نهضة مصر ، القاهرة ، يناير ٢٠٠٢ .
- ٢ - سناء الخولى ، أزمة السكن ومشاكل الشباب ، دار المعرفة الجامعية ، القاهرة ، ٢٠٠٢ .
- رابعاً : الإحصاءات الرسمية للجهاز المركزى للتعبئة العامة والإحصاء :
- ١ - التعداد العام للسكان والإسكان والمنشآت للأعوام ١٩٧٦ ، ١٩٨٦ ، ١٩٩٦ .
- ٢ - نتائج التعداد العام للسكان والإسكان والمنشآت ٢٠٠٦ ، إجمالى الجمهورية ، مارس ٢٠٠٧ .
- ٣ - النشرة السنوية لأسعار المواد والمنتجات الصناعية ، أعداد مختلفة .
- ٤ - النشرة السنوية لبحث العمالة بالعينة بجمهورية مصر العربية ، أعداد مختلفة .
- ٥ - الكتاب الإحصائى السنوى ، أعداد مختلفة .
- ٦ - بحث الدخل والإنفاق والإستهلاك ، ٢٠٠٤/٢٠٠٥ .

السكان والغذاء في مصر

فيليب فهمى فانوس

مقدمة:

يمثل الغذاء الحاجة الأساسية الأولى التي تسعى الدول والحكومات إلى توفيرها لشعوبها، فمن يتحكم في غذاء الإنسان قادر على التحكم في فكره وتعطيل عقله والسيطرة على ملكاته وقدراته، فالغذاء سلاح سياسى يستخدم ببراعة في تذويب مقاومة الشعوب النامية والفقيرة وإخضاعها لسياسة الدول التي تمسك بمفاتيح الغلال في العالم، فالأمن الغذائى ضرورة سياسية واقتصادية لتأمين حياة الشعوب.

وتعتبر مصر إحدى الدول التي لا يفى إنتاجها من الغذاء بشكل عام احتياجات السكان المتزايدة، حيث يتزايد السكان بمعدلات أعلى من معدلات الزيادة في الإنتاج الزراعى، الأمر الذى أدى لوجود فجوة غذائية تتزايد حداثها عاماً بعد آخر وانعكس أثرها على عجز الميزان التجارى وأيضاً ميزان المدفوعات، حيث أدى الانخفاض في معدلات نمو الإنتاج الزراعى عن معدلات الطلب على الغذاء إلى زيادة الواردات من بعض السلع الغذائية الأساسية، وأيضاً أدى انخفاض معدلات النمو في الإنتاج الزراعى عن معدلات نمو الطلب على الغذاء إلى تراجع نصيب الفرد من الرقعة الزراعية بحيث تعجز هذه الرقعة عن توفير الاحتياجات الغذائية للفرد.

١- أهمية الدراسة :

لم يعد الغذاء مجرد سلعة تتداول وتدخل تحت قانون العرض والطلب وإنما يعتبر فى مقدمة القضايا التي تهتم بها الدولة سواء من خلال إنتاجها المحلى أو من خلال اللجوء إلى الاستيراد من المصادر الخارجية لسد الفجوة الغذائية.

تأتى أهمية موضوع الغذاء مع التزايد المستمر في أعداد السكان وما يتطلبه ذلك من زيادة الانتاج من الغذاء بنفس المعدل حتى لا تتزايد حدة الفجوة الغذائية والتي يتم محاولة سدها عن طريق الواردات من السلع الغذائية وما سيتبع ذلك من عجز في الميزان التجارى. ولا شك أن استيراد الغذاء وخاصة من السلع الاستراتيجية هو أمر محفوف بالمخاطر بل ويؤثر على الأمن القومى وخاصة مع تغير السياسات الزراعية للدول المنتجة للغذاء بصفة عامة والتي تضع قيوداً على بعض السلع الغذائية الهامة التي تصدرها أو تساعد بها الدول النامية، بل ويزداد الأمر صعوبة حينما تلجأ هذه الدول المصدره للغذاء إلى استخدامه كأداة ضغط على الدول المستوردة مما يجد من المساحة المتاحة لحرية قرارها السياسى حيث أصبح الغذاء سلعة سياسية ذات طابع استراتيجى.

من كل ما تقدم تأتى أهمية هذه الدراسة التي تعرض حاجة الدولة إلى تحقيق قدر معقول من الأمن الغذائى باعتباره أحد ركائز الأمن القومى نظراً لتأثيره المباشر على صحة الأفراد ونشاطهم اليومى وقدرتهم على المشاركة في العملية الانتاجية، حتى أصبح توفير الغذاء اليومى أحد القضايا التي تشغل الدولة وتضعها في حالة من العمل المستمر تأمينا للمواطنين ولوقعها السياسى على خريطة العالم حيث (من لا يملك قوته لا يملك ارداته).

٢ - أهداف الدراسة :

تهدف الدراسة إلى التعرف على ما يلي :

أ - أبعاد مشكلة الغذاء في مصر وارتباطها المباشر بزيادة السكان .

ب - مصادر إنتاج الغذاء .

ج - الفجوة الغذائية والاكتفاء الناتج .

د - امكانيات تحقيق الأمن الغذائي في مصر .

٣ - منهجية الدراسة :

استخدم في هذه الدراسة المنهج الوصفي التحليلي للبيانات بغرض معرفة العلاقات بين المتغيرات

المختلفة في الإنتاج الزراعي ومدى تشابك هذه العلاقات لتحديد حجم الفجوة الغذائية .

جدير بالذكر أن الدراسة قد تناولت المؤشرات الإحصائية من عام ٢٠٠١ حتى عام ٢٠٠٧ باستثناء الجزء

الخاص بالسكان والمساحة المنزرعة والتي تبدأ مؤشرات من عام ١٩٤٧ حسب سنوات التعداد .

الفصل الأول السكان والمساحة المنزرعة

أولاً : النمو السكاني

مع بداية القرن التاسع عشر قدر عدد السكان بنحو ٢,٥ مليون نسمة ارتفع إلى حوالي عشرة ملايين نسمة في نهايته . وفي المقابل زادت الرقعة الزراعية من ٢,٠٢٢ مليون فدان إلى ٥,٠٨٨ مليون فدان خلال الفترة (١٨٢١ - ١٨٩٧) , ومع بداية القرن العشرين أخذ السكان في الزيادة المعتدلة ثم الزيادة المتسارعة في آخره حيث بلغ عدد السكان عام ١٩١٧ حوالي ١٢,٧ مليون نسمة ارتفع إلى ٢٦,٦ مليون نسمة عام ١٩٧٦ ثم إلى ٧٢,٧ مليون عام ٢٠٠٦ , في الوقت الذي بلغت فيه المساحة المنزرعة حوالي ٨,٤ مليون فدان لينخفض نصيب الفرد من المساحة المنزرعة من ٠,٥٢ فدان عام ١٨٩٧ إلى ٠,١٢ فدان عام ٢٠٠٦ . ويوضح الجدول رقم (١) تطور أعداد السكان خلال السنوات التعدادية من عام ١٩٤٧ - ٢٠٠٦ ومنه يتضح أن :

أ - تضاعف عدد السكان تقريباً خلال أقل من ثلاثين عام (١٩٤٧ - ١٩٧٦) , ثم قفز عدد السكان إلى حوالي ٥٩,٢ مليون نسمة عام ١٩٩٦ . وفي آخر تعداد عام ٢٠٠٦ بلغ عدد السكان في مصر نحو ٧٢,٧ مليون نسمة (باستبعاد أعداد المصريين في الخارج) .

ب - بدراسة نسب الزيادة السنوية نجد أن أعلى نسبة زيادة سنوية سجلت خلال العشر سنوات المحصورة بين تعدادي (١٩٧٦ - ١٩٨٦) وبلغت ٢,٢ % .

وفي الوقت الذي تنادي فيه الدولة بمحاولة وقف الانفجار السكاني . نجد أن معدل النمو مازال أكبر كثيراً من الدول الغنية مما يؤثر دائماً على معدلات التنمية وإيضاً في زيادة المشكلة الغذائية بصفة مستمرة .

ويوضح الجدول رقم (٢) معدل نمو السكان في الفترة مابين التعدادات السكانية (١٩٤٧ - ٢٠٠٦) ومنه يتضح الآتي :

أ - إنجه معدل نمو السكان نحو الإرتفاع خلال التسعة عشر عاماً الواقعة بين تعدادي ١٩٤٧ , ١٩٦٦ , حيث بلغ ٢,٢٤ % خلال الفترة من عام ١٩٤٧ حتى عام ١٩٦٠ ثم ارتفع إلى ٢,٥٢ % خلال الفترة المحصورة بين تعدادي ١٩٦٠ , ١٩٦٦ .

ب - إنخفض معدل النمو السكاني خلال السنوات العشرة المحصورة بين عامي ١٩٦٦ - ١٩٧٦ وبلغ ١,٩٢ % , وقد يعزى ذلك إلى عدوان ١٩٦٧ وحرب الإستنزاف وما صاحبهما من عدم إستقرار الأحوال السكانية وإنخفاض معدلات الخصوبة

ج - عاد معدل النمو السنوي للسكان إلى الإرتفاع حتى بلغ أعلاه خلال الفترة المحصورة بين عامي ١٩٧٦ , ١٩٨٦ وبلغ ٢,٧٥ % وهو أعلى معدل عرفته مصر في تاريخها السكاني الحديث ثم إنخفض خلال الفترة من عام ١٩٨٦ حتى عام ١٩٩٦ وبلغ ٢,٠٨ % ثم إلى ٢,٠٢ % خلال الفترة من عام ١٩٩٦ - ٢٠٠٦ .

جدول رقم (١)

تطور أعداد السكان خلال السنوات التعدادية من عام ١٩٤٧ - ٢٠٠٦

الزيادة السنوية		الزيادة الكلية بين التعدادين		عدد السكان (بالآلاف)	البيان سنوات التعداد
النسبة المئوية للزيادة	العدد (بالآلاف)	النسبة المئوية للزيادة	العدد (بالآلاف)		
=	=	=	٢٠٤٦	١٨٩٦٧	١٩٤٧
٢,٩	٥٤٧,٥	٢٧,٥	٧١١٨	٢٦٠٨٥	١٩٦٠
٢,٦	٦٦٥,٢	١٥,٢	٢٩٩١	٢٠٠٧٦	١٩٦٦
٢,٢	٦٥٥,١	٢١,٨	٦٥٥١	٢٦٦٢٧	١٩٧٦
٢,٢	١١٦٢,٧	٢١,٧	١١٦٢٧	٤٨٢٥٤	١٩٨٦
٢,٢	١١٠٥,٩	٢٢,٩	١١٠٥٩	٥٩٢١٢	١٩٩٦
٢,٢	١٢٤٨,٥	٢٢,٧	١٢٤٨٥	٧٢٧٩٨	٢٠٠٦

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، الكتاب الإحصائي السنوي ٢٠٠٨ ، ديسمبر ٢٠٠٨ ص ٢٧

جدول رقم (٢)

معدل نمو السكان في الفترة ما بين التعدادات السكانية ١٩٤٧ - ٢٠٠٦

الفترة	متوسط معدل النمو السكاني (%)
١٩٦٠ - ١٩٤٧	٢,٢٤
١٩٦٦ - ١٩٦٠	٢,٥٢
١٩٧٦ - ١٩٦٦	١,٩٢
١٩٨٦ - ١٩٧٦	٢,٧٥
١٩٩٦ - ١٩٨٦	٢,٠٨
٢٠٠٦ - ١٩٩٦	٢,٠٢

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، الكتاب الإحصائي السنوي ٢٠٠٨ ، ديسمبر ٢٠٠٨ ص ٣١

يوضح الجدول رقم (٢) والخاص بتوزيع السكان وفقاً للنوع (حضر / ريف) خلال السنوات التعدادية من ١٩٤٧ - ٢٠٠٦ ما يلي :

١ - هناك تباين في توزيع السكان بين الحضر والريف حيث ارتفع عدد سكان الحضر من حوالي ٦.٤ مليون نسمة بنسبة ٢٢,٥ ٪ من جملة سكان مصر عام ١٩٤٧ إلى حوالي ٢١,٢ مليون نسمة بنسبة ٤٤ ٪ من السكان عام ١٩٨٦ , وهذا يعنى أن سكان الحضر قد زاد عددهم أكثر من ثلاث مرات خلال أربعين عاماً , وبلغ هذا العدد ٢٠,٩ مليون نسمة عام ٢٠٠٦ بنسبة ٤٢,٦ ٪ من إجمالي سكان الجمهورية . وهذا التزايد السريع لسكان الحضر يمكن تفسيره بالزيادة الطبيعية , والهجرة من الريف للحضر والضم الإداري . وقد ترتب على التزايد السريع لسكان المدن الكثير من المشكلات التي من أهمها البطالة وزحف المدن الجديدة على حساب الأراضي الزراعية , وكذلك زيادة الطلب على الغذاء لإشباع حاجات الوافدين الجدد إلى المدن وزيادة المناطق العشوائية بالمدينة .

٢ - أما على مستوى النوع فنجد أن نسبة الإناث تناقصت من ٥٠,٥ ٪ من جملة عدد السكان في تعداد ١٩٤٧ إلى ٤٩,١ ٪ (١) في تعداد ١٩٧٦ أما في عام ٢٠٠٦ فبلغت نسبتهم ٤٨,٩ ٪ من جملة عدد السكان بينما تزايدت نسبة الذكور من ٤٩,٥ ٪ عام ١٩٤٧ إلى ٥١,١ ٪ عام ٢٠٠٦ .

ومن المعروف أن نوع الغذاء وكمياته تختلف باختلاف أعداد الذكور والإناث داخل أي أسرة , إما بسبب الاختلافات الفسيولوجية بين النوعين , أو بفعل التفضيل الأسري للذكور عن الإناث .

ثانياً : تطور المساحة المنزرعة:

بلغت مساحة الأراضي المزروعة حوالي ٨,٤ مليون فدان عام ٢٠٠٦ , بينما كانت حوالي ٥,١ مليون فدان عام ١٨٩٧ حيث تزايدت بعدها بمعدلات بطيئة حتى وصلت إلى المساحة المزروعة الحالية .

(١) من الجدول رقم (٤) و الذي يوضح تطور نصيب الفرد من المساحة المزروعة و المحصولية خلال الفترة من عام ١٨٩٧ حتى عام ٢٠٠٦ يتبين أن :

١ - تزايدت المساحة الأرضية المزروعة خلال الفترة (١٨٩٧ - ٢٠٠٦) بنسبة ٦٥,٢ ٪ فقط وترتب على ذلك أن نصيب الفرد من الأرض المزروعة هبط من حوالي نصف فدان عام ١٨٩٧ إلى ٠,١٢ فدان عام ٢٠٠٦ .

ب - السكان في تزايد سريع ففى خلال نفس الفترة تزايد عدد السكان بنسبة ٦٥٠,٦ ٪ أى حوالي عشرة أضعاف الزيادة فى مساحة الارض.

ج - تزايدت المساحة المحصولية خلال الفترة المذكورة بنسبة ١٢٠,٦ ٪ نتيجة التحول من رى الحياض إلى الرى

الدائم , و إنخفض نصيب الفرد من المساحة المحصولية من ٠,٧ فدان عام ١٨٩٧ إلى ٠,٢١ فدان عام ٢٠٠٦ .

و من المتوقع أن يتناقص متوسط نصيب الفرد من الأرض الزراعية بإطراد مع التزايد السريع للسكان بما

يتجاوز كثيراً الزيادة فى المساحة المنزرعة.

جدول رقم (٢)

توزيع السكان وفقاً للنوع (حضر / ريف) خلال السنوات التعدادية من عام ١٩٤٧ - ٢٠٠٦

(عدد السكان بالآلاف)

البيان السنوات	حضر		ريف		الجملة		والريف من جملة السكان	
	ذكور	إناث	ذكور	إناث	ذكور	إناث	حضر	ريف
١٩٤٧	٢٢١٩	٢١٤٤	٦١٧٢	٦٤٢١	٩٢٩٢	٩٥٧٥	٢٢,٥	٦٦,٥
١٩٦٠	٥٠٧١	٤٨٩٤	٨٠٤٧	٨٠٧٢	١٢١١٨	١٢٩٦٧	٢٨,٢	٦١,٨
١٩٦٦	٦١٢٢	٥٩٠١	٩٠٤٤	٨٩٩٩	١٥١٧٦	١٤٩٠٠	٤٠,٠	٦٠,٠
١٩٧٦	٨٢٢٨	٧٨٠٨	١٠٤٢٠	١٠١٧٠	١٨٦٤٨	١٧٩٧٨	٤٢,٨	٥٦,٢
١٩٨٦	١٠٩٠٩	١٠٢٠٧	١٢٨٠٠	١٢٢٢٨	٢٤٧٠٩	٢٢٥٤٥	٤٤,٠	٥٦,٠
١٩٩٦	١٢٩٥٨	١٢٢٢٨	١٧٢٩٤	١٦٦٢٢	٢٠٢٥٢	٢٨٩٦١	٤٢,٠	٥٧,٠
٢٠٠٦	١٥٧٩٤	١٥١٥٦	٢١٢٠٧	٢٠٢٢٢	٢٧١٠١	٢٥٤٧٨	٤٢,٦	٥٧,٤

المصدر: (١) الجهاز المركزي للتعبئة العامة والإحصاء، الكتاب الإحصائي السنوي ٢٠٠٥، (إصدار ٢٠٠٦، ص ٢٣، ٢٤).

النتائج النهائية للتعداد العام للسكان والإسكان والمنشآت ٢٠٠٦، مارس ٢٠٠٧، ص ٧.

جدول رقم (٤)

تطور نصيب الفرد من المساحة المزروعة والمحصولية

خلال الفترة من عام ١٩٩٧ حتى عام ٢٠٠٦

البيان السنوات	عدد السكان (الف نسمة)		المساحة المزروعة			المساحة المحصولية		
	العدد	نسبة التغير %	المساحة (ألف فدان)	نسبة التغير %	نصيب الفرد (فدان)	المساحة (ألف فدان)	نسبة التغير %	نصيب الفرد (فدان)
١٩٩٧	٩,٦٦٩	-	٥,٠٨٨	-	٠,٥٢	٦,٧٦٤	-	٠,٧٠
١٩٠٧	١١,١٩٠	١٥,٧	٥,٤٠٢	٦,٢	٠,٤٨	٧,٦٦٢	١٢,٢	٠,٦٨
١٩١٧	١٢,٧١٨	١٢,٧	٥,٢٩٦	٢,٠	٠,٤٢	٧,٦٨٦	٠,٢	٠,٦٠
١٩٢٧	١٤,١٧٨	١١,٥	٥,٥٤٤	٤,٧	٠,٢٩	٨,٦٦١	١٢,٧	٠,٦١
١٩٣٧	١٥,٩٢١	١٢,٢	٥,٢٨١	٤,٧	٠,٢٢	٨,٢٥٨	٢,٥	٠,٥٢
١٩٤٧	١٨,٩٦٧	١٩,١	٥,٧٦١	٩,١	٠,٢٠	٩,١٦٧	٩,٧	٠,٤٨
١٩٦٠	٢٦,٠٨٥	٢٧,٥	٥,٩٧٤	٢,٧	٠,٢٢	١٠,٢٩٧	١٢,٤	٠,٤٠
١٩٦٦	٢٠,٠٧٦	١٥,٢	٦	٠,٤	٠,٢٠	١٠,٤٠٠	٠,٠٢	٠,٢٥
١٩٧٦	٢٦,٦٢٧	٢١,٨	٦,٢٠٠	٥,٠	٠,١٧	١١,٩٠٠	١٤,٤	٠,٢٢
١٩٨٦	٤٨,٢٥٤	٢١,٧	٦,٠١٩	٤,٥	٠,١٢	١١,٢٦٢	٥,٤	٠,٢٢
١٩٩٦	٥٩,٢١٢	٢٢,٩	٧,٥٨٩	٢٦,١	٠,١٢	١٢,٧١٠	٢١,٧	٠,٢٢
٢٠٠٦	٧٢,٧٩٨	٢٢,٧	٨,٤١١	١٠,٨	٠,١٢	١٥,١٧٦	١٠,٧	٠,٢١

المصدر: (١) الجهاز المركزي للتعبئة العامة والإحصاء، النتائج النهائية للتعداد العام للسكان والإسكان والمنشآت ٢٠٠٦، مارس ٢٠٠٧، ص ١٤٢.

نشرة المساحة المحصولية والمساحات المزروعة في ج.م. ع ١٩٩٦، مارس ١٩٩٨، ص ٧٩، ٥.

(٢) وزارة الزراعة وإستصلاح الأراضي، الإدارة المركزية للإقتصاد الزراعي "دراسة لأهم مؤشرات الإحصاءات الزراعية"، الجزء الثاني ٢٠٠٥ / ٢٠٠٦، ص ٥٨، ٢٤.

(٣) محمد عبد الغنى سعودى، وسيم عبد الحميد، السكان والغذاء في مصر، كتاب الأفرام الإقتصادى، العدد (٧٩)، ١٩٩٤، ص ٢٤.

د- رغم مجهودات استصلاح الأراضي فإن هناك مشكلة النمو العمراني الذي يبتلع الأرض الجيدة ، حيث تقدر المساحة التي تستقطع من الأراضي الزراعية بنحو ٤٥ ألف فدان سنوياً ، بالإضافة إلى نقص أو تدمير الطاقة الحيوية للأرض و التي تؤدي في النهاية إلى التصحر أي أنه إذا استمرت معدلات استصلاح الأراضي و معدلات تأكلها الراهنة ، فستكون جملة المستصلحة مساوية لجملة الفاقد و حينئذ تكون المحصلة صفراً ، و سينعكس ذلك سلباً على توفير الغذاء .

(٢) لمعرفة مدى التوازن أو الاختلال بين الأراضي المزروعة والسكان على مستوى المحافظات المصرية (باستثناء محافظات الحدود) و بدراسة الجدول رقم (٥) ، والخاص بالتوزيع النسبي للسكان ومساحة الأراضي المزروعة في المحافظات المصرية عام ٢٠٠٦ ، يتلاحظ الآتي :

يمكن تقسيم المحافظات إلى ثلاثة مجموعات وهي :

- أ - محافظات يتحقق فيها التوازن تقريباً بين نسبة السكان ونسبة المساحة المزروعة بها مثل محافظات (قنا - الدقهلية - دمياط - المنوفية - بور سعيد - السويس - بنى سويف - أسيوط - المنيا - الأقصر) حيث تضم هذه المحافظات نحو ٢٢,٨ ٪ من السكان ونحو ٢٢,٢ ٪ من المساحة المزروعة .
 - ب - محافظات لا تتوازن فيها المساحة المزروعة مع السكان وهي (القاهرة - الجيزة - الإسكندرية) حيث تضم هذه المحافظات نحو ٢٥,٠ ٪ من السكان وحوالي ٤,٧ ٪ فقط من المساحة المزروعة .
 - ج - محافظات يقل نسبة سكانها إلى جملة سكان مصر عن نسبة مساحة أراضيها المزروعة إلى جملة المساحة المزروعة في مصر مثل محافظة كفر الشيخ حيث تضم ٢,٦ ٪ من السكان يقابله ٧,٢ ٪ من جملة المساحة المزروعة .
- محافظة الشرقية وبها نحو ٧,٤ ٪ من السكان يقابله ٩,٦ ٪ من المساحة المزروعة .
- ولقد ترتب على تقلص نصيب الفرد من المساحة المزروعة وتدهور خصوبة الأراضي الزراعية انخفاض الحجم النسبي للمستغلين بالزراعة . ففي النصف الأول من القرن العشرين (١٨٩٧ - ١٩٤٧) بلغ الحجم النسبي للعاملين بالزراعة نحو ثلثي حجم قوة العمل الكلية ، إلا أن نسبة المستغلين في الزراعة إلى جملة القوة العاملة انخفضت إلى ٤٧,٦ ٪ في تعداد ١٩٧٦ ثم هبطت إلى ٢٩,٠ ٪ في تعداد ١٩٨٦ ، ثم ٢٠,٩ ٪ في تعداد ١٩٩٦ ، وقد ظلت نسبة المستغلين بالزراعة ٢٠,٩ ٪ إلى جملة القوة العاملة عام ٢٠٠٥ ، ويعد ذلك دليلاً على الهجرة السريعة للسكان من الريف إلى الحضر وهجر الريفيين لمهنة الزراعة ، وهو اتجاه سوف يستمر مستقبلاً طالما أن الزراعة لا تحقق الدخل المناسب للعاملين بها .

(١) محمد عبد الغنى سعودى ، وسيم عبد الحميد ، السكان والغذاء في مصر ، كتاب الأهرام الإقتصادى العدد (٧٩) ،

سبتمبر ، ص ٢٦

(٢) محمد عبد الغنى سعودى ، وسيم عبد الحميد ، " السكان والغذاء في مصر " ، مرجع سابق ص ٣٥

جدول رقم (٥)

التوزيع النسبي للسكان ومساحة الأراضي المزروعة بالمحافظات عام ٢٠٠٦

المحافظة		البيان	
الترتيب	% من جملة السكان	الترتيب	% من جملة المساحة
٢٥	١٠,٧	١	٠,٢
١٨	٥,٧	٨	٢,٠
٢٢	٠,٨	٢٠	٠,٦
٢٤	٠,٧	٢١	٠,٢
١٤	١,٢	١٩	٢,٨
١	٦,٥	٥	١٨,٦
٢١	١,٥	١٨	١,٢
٤	٢,٦	١٤	٧,٢
٧	٥,٥	٩	٤,٦
٣	٦,٩	٤	٧,٨
٢	٧,٤	٣	٩,٦
٨	٤,٥	١٢	٤,٤
١٦	٥,٨	٧	٢,٢
—	٦٠,٩	—	٦١,٧
١٥	٨,٦	٢	٢,٥
٦	٢,٥	١٥	٥,١
١٢	٢,١	١٦	٢,٤
٥	٥,٨	٦	٥,٨
—	٢١,٠	—	١٦,٨
٩	٤,٨	١١	٤,١
١١	٥,٢	١٠	٢,٨
١٠	٤,١	١٣	٤,١
٢٣	٠,٦	٢٢	٠,٥
١٧	١,٦	١٧	٢,٢
—	١٦,٢	—	١٤,٧
—	٩٨,٢	—	٩٢,٢
١٩	٠,٣	٢٦	١,٧
١٢	٠,٤	٢٤	٢,٦
٢٧	٠,٤	٢٥	—
٢٠	٠,٥	٢٣	١,٤
٢٦	٠,٢	٢٧	٠,١
—	١,٨	—	٦,٨
—	١٠٠,٠	—	١٠٠,٠

المصدر : (١) الجهاز المركزى للتعبئة العامة والإحصاء ، النتائج النهائية لتعداد العام للسكان والإسكان والمنشآت ٢٠٠٦

(٢) وزارة الزراعة وإستصلاح الأراضى ، الإدارة المركزية للإقتصاد الزراعى ، دراسة لأهم مؤشرات الإحصاءات الزراعية ،

الجزء الثانى ، المحاصيل الصيفية والتبيلية ٢٠٠٥ / ٢٠٠٦ ص ٥٨

الفصل الثانى الواقع الغذائى المصرى

بدراسة الواقع الغذائى المصرى وتقييمه ما يساعد على رسم سياسات الإنتاج المحلى للغذاء بما يكفل تغطية الاحتياجات الاستهلاكية للمواطنين أو سد الفجوة بين الإنتاج والاستهلاك بالاستيراد. مع مراعاة إمكانات التصدير لسداد قيمة هذه الواردات الغذائية المكملة للإنتاج المحلى.

ومن المسلم به أن معدلات الزيادة السكانية تهدد بزيادة حجم مشكلة الغذاء والتغذية، حيث أن الزيادة فى عدد السكان تفوق الزيادة فى إنتاج الغذاء خاصة فى الدول النامية التى تعاني سياساتها الاستثمارية والإنتاجية من قصور واضح فى هذا المجال لأسباب متعددة داخلية وخارجية.

أولاً: مصادر إنتاج الغذاء

يمثل توفير الغذاء الحاجة الأساسية الأولى التى تسعى جميع الحكومات والدول إلى توفيرها لشعوبها، ويعتبر قطاع الزراعة الركيزة الأساسية لتوفير إنتاج الغذاء للإستهلاك المحلى.

وفيما يلى سوف تتناول مصادر إنتاج الغذاء لاهم السلع الغذائية الرئيسية فى مصر :

١- الإنتاج النباتى:

(أ) — مجموعة الحبوب: وتشمل القمح والأرز الأبيض والذرة الشامية (بيضاء، صفراء) والذرة الرفيعة والشعير.

بدراسة الجدول رقم (٦) والخاص بالإنتاج المحلى من الحبوب وكمية المتاح للإستهلاك ومتوسط نصيب الفرد خلال الفترة من عام ٢٠٠١ حتى ٢٠٠٧ يتضح ما يلى:

(١) بلغ الإنتاج المحلى من الحبوب عام ٢٠٠١ حوالى ١٧,٧ (مليون طن) ظل فى الزيادة حتى عام ٢٠٠٥ ليبلغ ٢١,١ (مليون طن) ثم انخفض عام ٢٠٠٧ إلى ٢٠,٠ (مليون طن) حيث بلغت نسبة الزيادة خلال الفترة الزمنية المذكورة حوالى ١٢٪.

(٢) بلغ نسبة الإنتاج المحلى من كمية المتاح للاستخدام حوالى ٦٨٪ عام ٢٠٠١ انخفضت إلى ٦٧,٢٪ عام ٢٠٠٧ والفرق بينهما يغطى عن طريق الاستيراد.

(٣) بلغ متوسط نصيب الفرد من الحبوب عام ٢٠٠١ نحو ٢١٥,٩ كجم/ سنة، وزاد إلى ٢٢٩,٦ كجم/ سنة عام ٢٠٠٧.

(ب) — مجموعة البقوليات: وتشمل الفول الجاف والعدس والفاصوليا الجافة واللوبياء الجافة والبالاى الجافة والحمص والحبلة والترمس، وكلها مصادر أساسية للبروتين النباتى فى غذاء الإنسان والتى تعتمد عليها الطبقات محدودة الدخل.

بدراسة الجدول رقم (٧) والخاص بالإنتاج المحلي من البقوليات وكمية المتاح للاستخدام ومتوسط نصيب الفرد من عام ٢٠٠١ حتى ٢٠٠٧ يتضح ما يلي:

(١) بلغ الإنتاج المحلي للبقوليات عام ٢٠٠١ حوالي ٥٢٦ ألف طن وأخذ في الانخفاض حتى بلغ ٤٠٧ ألف طن عام ٢٠٠٧ بنسبة انخفاض ٢٢,٦٪.

(٢) بلغت كمية المتاح للاستخدام عام ٢٠٠١ حوالي ٨٥٢ ألف طن بزيادة قدرها ٢٢٦ ألف طن عن الإنتاج المحلي الذي يغطي نسبة ٦٢٪ من المتاح للاستخدام حيث يتم تغطية الفرق بالاستيراد، وبلغ متوسط نصيب الفرد لهذا العام ١٠,٤ كجم/سنة.

(٣) انخفض المتاح للاستخدام عام ٢٠٠٧ إلى ٨٠٦ ألف طن حيث غطى الإنتاج المحلي لهذا العام نسبة ٥٠,٦٪ من المتاح للاستخدام وبلغ متوسط نصيب الفرد لهذا العام ٩ كجم/سنة.

٢ - الإنتاج الحيواني:

تعتبر الثروة الحيوانية ذات أهمية غذائية خاصة بالنسبة للإنسان لأنها أهم المصادر للبروتين الحيواني الذي يحتوي على الأحماض الأمينية الأساسية اللازمة لبناء الجسم، بالإضافة إلى أهميتها في مجال الصناعة والزراعة بدراسة الجدول رقم (٨) والخاص بالإنتاج المحلي وكمية المتاح للإستهلاك ومتوسط نصيب الفرد من اللحوم الحمراء واللحوم البيضاء والأسماك يتضح الآتي:

(أ) - اللحوم الحمراء:

(١) بلغ الإنتاج المحلي للحوم الحمراء عام ٢٠٠١ حوالي ٦٩٥ ألف طن ارتفع إلى ٩٢١ ألف طن عام ٢٠٠٧ بنسبة زيادة قدرها ٣٢,٥٪ عن عام ٢٠٠١.

(٢) بلغ المتاح للإستهلاك عام ٢٠٠١ حوالي ٧٩٤ ألف طن ارتفع إلى ١,٢ مليون طن تقريباً عام ٢٠٠٧.

(٣) زاد متوسط نصيب الفرد من اللحوم الحمراء من ١٢,٢ كجم/سنة عام ٢٠٠١ إلى ١٦,٩ كجم/سنة عام ٢٠٠٧ وهو من أقل المتوسطات في العالم.

(ب) - اللحوم البيضاء:

(١) بلغ الإنتاج المحلي من اللحوم البيضاء عام ٢٠٠١ حوالي ٥٧٧,٧ ألف طن، زادت إلى ٨٧٨,٨ ألف طن عام ٢٠٠٧ بنسبة زيادة قدرها ٥٢,١٪ عن عام ٢٠٠١.

(٢) بلغت كمية المتاح للإستهلاك عام ٢٠٠١ حوالي ٥٧٧ ألف طن، ارتفعت إلى ٨٨٨ ألف طن عام ٢٠٠٧.

(٣) زاد متوسط نصيب الفرد من ٨,٨ كجم/سنة عام ٢٠٠١ إلى ١٢,١ كجم/سنة عام ٢٠٠٧.

جدول رقم (٦)
الإنتاج المحلي من الحبوب وكمية المتاحة للإستخدام ومتوسط نصيب الفرد
خلال الفترة من ٢٠٠١ = ٢٠٠٧

(الكمية بالآلف طن)

السنوات	الإنتاج المحلي	كمية المتاحة للإستخدام	توزيع كمية المتاحة للإستخدام					متوسط نصيب الفرد كجم / سنة
			غذاء الحيوان	التقايى	الصناعة	الفاقد	المتبقى لغذاء الانسان	
٢٠٠١	١٧٦٨٠	٢٦٠١٢	٢٩٨٥	٢٤٦	٢٧٢	٧٨٤	٢٠٦٢٦	٢١٥,٩
٢٠٠٢	١٨٢٧٢	٢٨٢٢٢	٣٨٢٤	٢٥٥	٢٦٩	٨٩٤	٢٢٩٧١	٢٤٤,٨
٢٠٠٣	١٨٤٢٧	٢٦٠٠٧	٢٦٣٦	٢٢٧	٢٥١	٨١٦	٢٠٩٧٧	٢٠٨,٦
٢٠٠٤	١٩٢١٥	٢٥٤٨٢	٢٢٤٣	٢٤١	٢٢٥	٨١٩	٢٠٨٥٥	٢٠٠,٨
٢٠٠٥	٢١٠٨٥	٢٠٢٦٢	٤٤٠٢	٢٨٠	٢٩٩	٩٦١	٢٤٢٣٠	٢٤٢,٩
٢٠٠٦	٢٠٨٦٧	٢٠٠٢٥	٤٢٦٩	٢٧٧	٢٩٧	٩٥٤	٢٤٠٢٨	٢٢٢,٩
٢٠٠٧	٢٠٠٢٩	٢٩٨٠٥	٢٩٢٩	٢٨٧	٢٧٤	٩٥٥	٢٤٢٦٠	٢٢٩,٦

المصدر : الجهاز المركزى للتعبئة العامة والإحصاء ، دراسة تطور حركة الإنتاج والتجارة الخارجية والمتاح للإستهلاك من السلع الزراعية ، أعداد مختلفة .

جدول رقم (٧)
الإنتاج المحلي من البقوليات وكمية المتاحة للإستخدام ومتوسط نصيب الفرد
خلال الفترة من ٢٠٠١ = ٢٠٠٧

(الكمية بالآلف طن)

السنوات	الإنتاج المحلي	كمية المتاحة للإستخدام	توزيع كمية المتاحة للإستخدام					متوسط نصيب الفرد كجم / سنة
			غذاء الحيوان	التقايى	الصناعة	الفاقد	المتبقى لغذاء الانسان	
٢٠٠١	٥٢٦	٨٥٢	١٠١	٢٤	-	٤٠	٦٧٧	١٠,٤
٢٠٠٢	٥٠٦	٨٧١	١٠١	٢٤	-	٢٩	٦٩٧	١٠,٥
٢٠٠٣	٤٣٨	٨٢٢	٩٨	١٥	-	٢٧	٦٧٢	٩,٩
٢٠٠٤	٤١٦	٨٢٣	٩٦	١٣	-	٢٨	٦٧٦	٩,٨
٢٠٠٥	٢٧٨	٨٧٤	٩٨	١٤	٠	٤٢	٧٢٠	١٠,٢
٢٠٠٦	٢٥٢	٧٧٤	٨٩	١٣	-	٢٨	٦٢٤	٨,٨
٢٠٠٧	٤٠٧	٨٠٦	٨٨	١٥	-	٤٠	٦٦٢	٩

المصدر : الجهاز المركزى للتعبئة العامة والإحصاء ، دراسة تطور حركة الإنتاج والتجارة الخارجية والمتاح للإستهلاك من السلع الزراعية ، أعداد مختلفة .

جدول رقم (٨)

الإنتاج المحلي وكمية الإنتاج للإستخدام ومتوسط نصيب الفرد من اللعوم العمراء واللعوم البيضاء والأسماك

خلال الفترة من ٢٠٠١ = ٢٠٠٧

(الكمية بالطن)

متوسط نصيب الفرد (كجم / سنة)		كمية الإنتاج للإستخدام				الإنتاج المحلي			البيان السنوات
الأسماك	اللعوم البيضاء	اللعوم العمراء	الأسماك	اللعوم البيضاء	اللعوم العمراء	الأسماك	اللعوم البيضاء	اللعوم العمراء	
١٤.٥	٨.٨	١٢.٣	٩٥٠	٥٧٧	٨	٧٧٣	٥٧٧,٧	٦٩٥	٢٠٠١
١٣.٩	١١.٦	١٤.٣	٩٣٣	٧٧٠.١	٩٥٤	٨٠١	٧٧٢,٣	٨٢١	٢٠٠٢
١٤.٩	١٠.٣	١٣.٧	١٠١٥	٦٩٩.١	٩٢٠	٨٧٦	٧٠٠.١	٨٠٤	٢٠٠٣
١٥.٣	٩.٥	١٣.٤	١٠٦٤	٦٦١.٤	٩٢٧	٨٦٥	٦٦١,١	٨١٩	٢٠٠٤
١٥.٣	١٤.٣	١٤.٩	١٠٧٥	١٠١٥,٣	١٠٥٣	٨٨٩	١٠١٦	٨٥٥	٢٠٠٥
١٦.٢	١١.٣	١٦.٣	١١٧٣	٨٥٥.٦	١١٧٨	٩٧١	٧٩٥,٣	٨٨٠	٢٠٠٦
١٦.٦	١٢.١	١٦.٩	١٢٦٥	٨٨٨.١	١٢٤٧	١٠٠٨	٨٧٨.٨	٩٢١	٢٠٠٧

المصدر : الجول المركزي للتعينة العامة والإحصاء ، دراسة تطور الإنتاج والتجارة الخارجية والسلع للإستهلاك من السلع الزراعية ، اعداد مختلفة

(٢) - الإنتاج السمكي :

تعتبر الأسماك مصدراً رئيسياً من مصادر البروتين الحيواني عالية القيمة الغذائية منخفضة الثمن نسبياً مقارنة بمصادر البروتين الحيواني الأخرى، وخاصة اللحوم الحمراء.

وتتملك مصر العديد من المصادر المائية المختلفة للصيد مثل البحر المتوسط والبحر الأحمر والبحيرات والمياه العذبة (١)
(نهر النيل، بحيرة ناصر) تقدر بنحو ٦,٩ (مليون فدان) إلا أن الإنتاج السمكي لا يكفى للإستهلاك المتزايد إذ يبلغ إنتاج الفدان المائى نحو ٢٠٠ كجم سنوياً ومقابل ٨ أطنان فى بعض دول العالم.

وبدراسة الجدول رقم (٨) والمشار اليه سابقا يتضح الآتى :

(١) بلغ الإنتاج المحلى من الأسماك الطازجة والمجمدة عام ٢٠٠١ حوالى ٧٧٢ ألف طن زاد إلى ١٠٠٨ ألف طن عام ٢٠٠٧ بنسبة زيادة ٣٠,٦٪ عن عام ٢٠٠١.

(٢) بلغ المتاح للإستهلاك عام ٢٠٠١ حوالى ٩٤٠ ألف طن، ارتفع إلى حوالى ١,٢ مليون طن عام ٢٠٠٧ بنسبة زيادة قدرها ٢٠,٦٪ عن عام ٢٠٠١.

(٣) بلغ متوسط نصيب الفرد من الأسماك عام ٢٠٠١ حوالى ١٤,٥ كجم/ سنه زاد زيادة طفيفة عام ٢٠٠٧ إلى ١٦,٦؛

ثانياً: الفجوة الغذائية والاكتفاء الذاتى:

إن توفير الغذاء المتوازن والمتكامل للمواطنين يعتبر ضرورة اجتماعية تستهدف تحقيق الحد الأدنى الضرورى لصحة الإنسان، كما يعتبر ضرورة اقتصادية تستهدف بلوغ الأهداف الاقتصادية للدولة باعتبار أن التغذية من أهم عوامل تحقيق الإنتاج وزيادته.

١ - الفجوة الغذائية :

وبدراسة الجدول رقم (٩) والخاص بالإنتاج والاستهلاك والفجوة الغذائية لأهم السلع الزراعية خلال ٢٠٠٢-٢٠٠٧ يتضح الآتى:

(أ) بلغ الإنتاج من محصول القمح لعام ٢٠٠٢ حوالى ٦,٦ مليون طن يقابله إستهلاك قدره ١١,٨ مليون طن بما يعنى وجود فجوة قدرها ٥,٢ مليون طن، ارتفع الإنتاج إلى ٧,٤ مليون طن والاستهلاك إلى ١٢ مليون طن عام ٢٠٠٧ لتصبح هناك فجوة قدرها ٥,٦ مليون طن.

(ب) بلغ الإستهلاك من الذرة الشامية عام ٢٠٠٢ حوالى ٧,١ مليون طن مقابل إنتاج قدره ٦,٤ مليون طن بفجوة قدرها ٠,٧ مليون طن، وانخفض الإستهلاك إلى ٧,٢ مليون طن عام ٢٠٠٧ يقابله إنتاج قدره ٦,٩ مليون طن وعجز قدره ٢٠٧ ألف طن.

(ج) زيادة الإستهلاك عن الإنتاج فى محصول الفول حيث بلغ الإستهلاك عام ٢٠٠٢ حوالى ٥٠٨ ألف طن مقابل إنتاج ٤٠٦ ألف طن وفجوة قدرها ١٠٢ ألف طن وانخفض الإنتاج إلى ٢٠٥ ألف طن عام ٢٠٠٧ والاستهلاك إلى ٤٦٢ ألف طن بفجوة قدرها ١٥٧ ألف طن..

(١) محمد عبد الغنى سعودى ، وسيم عبد الحميد ، " السكان والغذاء فى مصر " ، مرجع سابق ، ص ٥٨

جدول رقم (٩)
الإنتاج والاستهلاك والفجوة الغذائية لأهم السلع الزراعية خلال الفترة من عام ٢٠٠٢ حتى ٢٠٠٧

السلع الزراعية / السنوات	٢٠٠٢			٢٠٠٣			٢٠٠٤			٢٠٠٥			٢٠٠٦			٢٠٠٧		
	الإنتاج	الاستهلاك	الفجوة	الإنتاج	الاستهلاك	الفجوة	الإنتاج	الاستهلاك	الفجوة	الإنتاج	الاستهلاك	الفجوة	الإنتاج	الاستهلاك	الفجوة	الإنتاج	الاستهلاك	الفجوة
القمح (١)	٦١٢٥	١١٧٥٠	٥١٢٥٠	٦٨٤٥	١٠٢٦٥	٣٥٢٠٠	٧١٧٨	١١٠٩٩	٣٩٣٠٠	٨١٤١	١٣٥٧٥	٤٤٣٤٠	٨٢٧٤	١٣٨٥٧	٥٥٨٣٠	٧٢٧٩	١٣٠٢٩	٥٦٥٠٠
الأرز	٤٢١٣	٣٥٦٩	٦٥٤٤	٤٢١٢	٢٤١٣	١٧٩٩	٤٢٨٣	٣٣٩٩	٨٨٧	٤٢٣٦	٣٤٤٣	١١٨٣	٤٦٦١	٣٤٧٩	١١٨٥	٤٢٣٩	٣٤٢٧	١٣١٢
الذرة الشامية	٦٤٣١	٧٥٥٠	١١١٩	٦٢٣٠	٦٥٥٧	٣٣٧	٦٧٢٨	٥٧٧٩	٩٤٩	٧٦٩٨	٨١٣٦	٤٣٨٠	٦٩٥٩	٦٧٧٤	١٤٥	٦٩٣٠	٧٢٣٧	٣٠٧٠٠
الفول	٤٠٩	٥٠٨	١٠٢	٢٤٢	٣٣٣	١٦٧	١٧٠	٥٠٣	١٧٠	٢٨٤	٥١١	٢٣٧	٢٥٧	٤٦٧	٢١٠	٢٠٥	٤٦٢	١٥٧٠٠
العدس	٤	٩٦	٩٢	٢	٦١	٥٨	٣	٨٧	٨٤	٢	١٠٤	١٠٢	١	٧٣	٧٣	٢	٨١	٧٩
البطاطس	١٩٨٥	١٣٨٩	٥٩٦	٢٠٢٩	١٣٥٠	٦٨٩	٢٥٤٧	١٥٧٢	٩٧٥	٣١٦٧	٢٠٤٥	١١٢٣	٣٣١٣	١٤٣٩	٨٧٤	٢٧٦٠	١٧٥١	١٠٠٩
البصل	١٣٧٦	٩١٨	٣٥٨	١٢٤٨	٨٧٥	٣٧٣	١١١٩	١١١٦	٤٠٢٠	١٩٣٦	١٤٣٦	٢٩٣	١٩٠٩	١٣٢٧	٢٨٣	١٥٧٤	١٣٩٦	٣٧٨
الثوم	٢٤٦	٢٢٣	٢٣	٢١٦	٢٤٨	١٨	٢٧٥	٢٥٧	١٨	٤٤٢	٤٤٢	١٤٣٦	٢١٧	٢٠٦	١١	٢٠٩	٢٩٣	١٧
الخضروات الطازجة (٢)	١٣١٥٠	١١٩١٤	١٢٣٦	١٢٥٦٦	١٢٣٩٩	١٢٦٧	١٢٩٢٢	١٢٦١٠	١٢١٢	١٢٩١٩	١٣٤٨٠	١٤٣٩	١٥٤٦	١٤٤٠٥	١٥٤١	١٥١٥٩	١٤١٣٣	١٥٣٧
الفواكه	٣٧٧٠	٣٣٥٧	٤١٣	٣٧٤٨	٣٣٠٦	٤٤٢	٢٨٨٨	٢٣٥٩	٥٣٩	٢٠٢٠	٢٥١٠	٥٢٠	٣٢١٢	٢٦١١	٦٠١	٢١٣٢	٢٥٦١٠	٥٧٥
الفاكهة الطازجة (٣)	٥١٩٨	٤٧٥٠	٤٤٨	٥١٠٣	٤١٧٣	٤٣٠	٥٥٠٦	٥٠١١	٤٤٥	٥٨٩٩	٥٣٧٠	٥٢٩	٩٥٩٦	٥٩٣٣	٦١٢	٦١١٤	٥٩٢٩	٦٨٥
اللحوم الحمراء	٨٢١	٩٥٤	١٣٢	٨٠٤	٩٣٧	١٣٦	٨١٩	٩٢٧	١٠٨	٨١٩	١٠١٧	١٩٨	٨٥٥	١١٥٣	٢٨٨	٩٢١	١٢٤٧	٢٢٦
لحوم الدواجن والطيور	٧٧٢	٧٧٠	٢	٧٠٠	٦٩٩	١	٦٦١	٦٦١	—	٦٦١	٦٦١	١	١٠١٦	١٠٢٦	١٠	٨٧٩	٨٨٨	٩٠
الأسمدة الطازجة والمجمدة	٩٢٣	١١٢	١٠	٨٧٦	١٠١٥	١٣٤	٨٦٥	١٠١٤	١٩٩	٨٦٥	١٥١	١٨٦	٩٧١	١١٧٣	٢٠٤	١٠٠٨	١٢٢٥	٢١٧
البضائع	٦٤٨٤	٦٤٧٤	١٠	٦٢٢٤	٦١٨٨	٤٦	٥٢٤٩	٥٢٤٤	٢٥	٥٢٤٩	٥٢١٢	٢٧	٤٣٥٨	٤٣٥١	٧	٤٣٠٨	٣٦٦٧	٦٤١
المنتجات	٤٢١٠	٤٢١٠	—	٥٢٨٠	٥٢٨٠	—	٤٦٨٢	٤٦٨٢	—	٥٥٠١	٥٥٠١	—	٥٥٠١	٥٥٠١	—	—	—	—

(١) القمح الخام فقط ولا يشمل الدقيق المستورد.

(٢) الخضروات الطازجة تشمل جميع الخضروات عدا المحاصيل الشتوية.

(٣) الفاكهة الطازجة وتشمل جميع الفاكهة عدا المواد والمنتجات والشمع والنقل.

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء، دراسة حركة التجارة الخارجية والسلع المستوردة من السلع الزراعية، أعداد مختلفة.

(د) بلغ إنتاج اللحوم الحمراء عام ٢٠٠٢ حوالي ٨٢١ ألف طن يقابلها إستهلاك قدره ٩٥٤ ألف طن بفجوة قدرها ١٢٢ ألف طن، وزاد الإنتاج عام ٢٠٠٧ إلى ٩٢١ ألف طن والاستهلاك إلى ١٢٤٨ ألف طن بفجوة قدرها ٢٢٦ ألف طن.

(هـ) بلغ الإنتاج من لحوم الدواجن والطيور عام ٢٠٠٢ حوالي ٧٧٢ ألف طن استهلك منه ٧٧٠ ألف طن بفائض ألفين طن، وزاد الإنتاج عام ٢٠٠٧ إلى حوالي ٨٧٩ ألف طن يقابله إستهلاك قدره ٨٨٨ ألف طن بفجوة قدرها ٩ آلاف طن.

(و) الأسماك الطازجة والمجمدة بلغ إنتاجها عام ٢٠٠٢ حوالي ٨٠١ ألف طن يقابله إستهلاك قدره ٩٢٢ ألف طن بفجوة قدرها ١٢٢ ألف طن، وزاد الإنتاج عام ٢٠٠٧ إلى ١٠٠٨ ألف طن يقابله إستهلاك قدره ١,٢ مليون طن بفجوة قدرها ٢١٧ ألف طن.

٢ - تقدير نسبة الاكتفاء الذاتي:

كانت مصر قديماً مصدرة للغلال ومتمتعاً بالاكتفاء الذاتي في معظم إنتاجها الزراعي مع فائض قل أو كثر للتصدير، وألان أصبحت محرومة من الاكتفاء الذاتي في معظم هذا الإنتاج ومستورده للغذاء، بل وتزداد كل يوم عجزاً في الإنتاج اعتماداً على الاستيراد وبالتالي استنزافاً لموارد الاقتصاد القومي من النقد الأجنبي.

والجدول رقم (١٠) الخاص بنسب الاكتفاء الذاتي من أهم السلع الزراعية خلال الفترة من ٢٠٠٢ حتى ٢٠٠٧ يتضح ما يلي:

(أ) - بلغت نسبة الاكتفاء الذاتي من القمح عام ٢٠٠٢ حوالي ٥٢,٢٪ ارتفعت عام ٢٠٠٢ إلى ٦٢,٥٪ ثم تراجعت إلى ٥٢,٥٪ عام ٢٠٠٧. أما نسبة الاكتفاء الذاتي من الأرز الأبيض بلغت أقصاها عام ٢٠٠٥ حوالي ١٢٥,٤٪ ثم انخفضت عام ٢٠٠٧ إلى ١٢٠,٤٪. وبالنسبة للاكتفاء الذاتي من الذرة الشامية فبلغت أقصاها عام ٢٠٠٤ حوالي ٧٢,٩٪ ثم انخفضت إلى ٦٠,٨٪ عام ٢٠٠٧.

(ب) - بلغت نسبة الاكتفاء الذاتي من الفول عام ٢٠٠٢ حوالي ٦٠,١٪ أخذت في الانخفاض حتى بلغت عام ٢٠٠٧ نحو ٥١,٦٪، إذا أن الفول غلة شديدة الحساسية للتقلبات الجوية بالإضافة إلى تعرضه للإصابة الشديدة ببعض الأمراض والآفات مما يؤدي إلى نقص كبير في إنتاجية الفدان.

(ج) - بلغت نسبة الاكتفاء الذاتي من العدس عام ٢٠٠٢ حوالي ٤٪ وأخذت في الانخفاض حتى بلغت عام ٢٠٠٧ نحو ٢,٤٪.

(د) - بلغت نسبة الاكتفاء الذاتي للحوم الحمراء عام ٢٠٠٢ حوالي ٨٦,١٪ وأخذت في الانحدار حتى عام ٢٠٠٧ حيث بلغت ٧٢,٩٪. وبالنسبة للاكتفاء الذاتي من لحوم الدواجن والطيور حيث بلغت عام ٢٠٠٢ حوالي ١٠٠,٣٪ وفي عام ٢٠٠٧ بلغت ٩٩٪.

(هـ) - بلغت نسبة الاكتفاء الذاتي من الأسماك الطازجة والمجمدة عام ٢٠٠٢ حوالي ٨٦,٨٪ أخذت في الانخفاض حتى بلغت ٨٢,٢٪ عام ٢٠٠٧.

جدول رقم (١٠)

نسب الإكتفاء الذاتي من أهم السلع الزراعية خلال الفترة من عام ٢٠٠٢ حتى ٢٠٠٧

%

السنوات	٢٠٠٢	٢٠٠٣	٢٠٠٤	٢٠٠٥	٢٠٠٦	٢٠٠٧
القمح (١)	٥٢,٢	٦٢,٥	٦١,١	٦١,٢	٥٦,٤	٥٢,٥
الأرز الأبيض	١١١,٢	١١٧,٨	١٢١,٧	١٢٥,٤	١٢٦,٥	١٢٠,٤
الذرة الشامية	٥٧,٩	٦٠,٢	٧٢,٩	٦٠,١	٦٤,٧	٦٠,٨
الفاصوليا	٦٠,١	٥٢,٤	٥١,٦	٤٣,٤	٤٣,٧	٥١,٦
العدس	٤,٠	٤,٧	٢,٢	١,٨	١٠,٠	٢,٤
البطاطس	١٠٩,٦	١١٢,٥	١١٦,٤	١١١,٢	١١٥,٤	١١٢,١
البصل	١٢٩,٦	١٢٤,٥	١٢٥,٦	١١٨,٤	١١٤,٤	١١٤,٦
الثوم	٩٨,٨	١٠١,٥	١٠١,١	١٠٠,٤	٩٩,٥	١٠٠,٠
الخضروات الطازجة (٢)	١٠٠,١	١٠٠,٢	١٠٠,٢	١٠٠,٤	١٠٠,٥	١٠٠,٥
المواالح	١٠٥,٧	١٠٧,٢	١١٠,١	١٠٨,٦	١١٠,٧	١١٠,٢
الفاكهة الطازجة (٢)	٩٩,٥	٩٩,٥	٩٩,٩	٩٩,٩	١٠٠	١٠٠,٧
بذر الكتان	٢٧,٥	٥٢,٤	٩٠,٩	٤٥,٨	٢٥,٧	٨٠,٠
بذر عباد الشمس	١٢٠,٧	١٢٢,٢	٩٢,٦	١١٦,٧	١٢٢,٦	١١٦,٠
فول الصويا	٥,٢	١٨,٠	١٦,٧	٤,٢	٢,٩	٢,٢
بذر القطن	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠,٠
بذر السمسم	٤٢,٥	٥٠,٠	٦٠,٧	٨٤,١	٨٢,٤	٨٩,٤
قصب السكر	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠,٠
بنجر السكر	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠,٠
اللحوم الحمراء	٨٦,١	٨٦,٥	٨٨,٢	٨٠,٥	٧٤,٨	٧٢,٩
لحوم الدواجن والطيور	١٠٠,٢	١٠٠,١	١٠٠	١٠٠,٢	٩٠	٩٩,٠
الأسماك الطازجة والمجمدة	٨٦,٨	٨٦,٢	٨١,٢	٨٢,٢	٨٢,٨	٨٢,٢
البيض الطازج	١٠٠,٢	١٠٠,٧	١٠٠,٧	١٠٠,٧	١٠٠,٢	١٠٠,٠
اللبان الحليب الخام	١٠٠	١٠٠	١٠٠	١٠٠	١٠٠	١٠٤,١

(١) القمح الخام فقط ولا يشمل الدقيق المستورد

(٢) الخضروات الطازجة تشمل جميع أنواعها عدا المحاصيل الشتوية

(٣) الفاكهة الطازجة تشمل جميع أنواعها عدا الشمام والبطيخ والمواالح والنقل

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء - دراسة تطور حركة الإنتاج والتجارة الخارجية والتمتع للإستهلاك من السلع الزراعية ، أعداد مختلفة .

الفصل الثالث

مشكلة الغذاء في مصر وامكانيات تحقيق الامن الغذائي

اولا : محددات مشكلة الغذاء في مصر

إذا أردنا أن نوصف قضية الغذاء فإنها توصف في محددات أساسية متمثلة في :

١ - السكان .

٢ - الموارد الطبيعية والزراعية .

٣ - إنتاج الغذاء .

٤ - الفجوة الغذائية بين ما ينتج وما يستهلك من غذاء .

فرغم الجهود التي تبذلها مصر لتضييق الفجوة الغذائية فإنها لا تزال تزداد اتساعا حتى مع وجود عمليات التطوير المستمرة مثل إدخال التكنولوجيا الحديثة وتطوير السلالات وزيادة إنتاجية بعض المحاصيل . إلا أن الزراعة المصرية لم تستطع حتى الآن الحد من اتساع الفجوة الغذائية بالمعدل المطلوب ويرجع ذلك إلى المشاكل المزمنة التي تعاني منها الزراعة المصرية مثل الخلل في التركيب المحصولي وتخلل إنتاجية الأرض الزراعية . كذلك النمط الغذائي السائد بين سكان مصر .

١ - اختلال التركيب المحصولي :

إن المركب المحصولي الأمثل هو الذي يوفر أقصى إنتاج من الغلات الزراعية تكفي للوفاء بإحتياجات الإستهلاك الغذائي الداخلي ومتطلبات الصناعات الزراعية المحلية مع فائض للتصدير . وقد كان لتدخل الدولة لتحديد الرقعة المخصصة لإنتاج بعض المحاصيل ثم إحتكار الدولة للتسويق والتعاوني دون النظر إلى تكاليف الإنتاج الفعلية والعوامل المؤثرة على عملية الإنتاج كل ذلك أدى إلى إتجاه المزارعين إلى التوسع في إنتاج الزروع التي تتعاضد ربحيتها النسبية . و نظراً لاختلاف وجهة نظر المزارع والدولة بشأن الزروع التي يجب أن يتضمنها التركيب المحصولي الأمثل والرقعة الزراعية المخصصة فإن الفجوة بين رغبات المزارعين ورغبة الدولة قد زادت تعقيداً .

٢ - تخلل إنتاجية الأرض الزراعية :

لقد كانت العوامل المتعلقة بالدخل المنخفض والنقص في المهارات الفنية والإدارية والتخلل النسبي في إستعمال الآلات الزراعية من أهم المؤثرات السلبية على التنمية الزراعية ، ومن أهم الأسباب الرئيسية لذلك تفتت مساحات الحيازات الزراعية وصغرها وتبعثرها ، و الواقع أن متوسط إنتاجية الفدان في كثير من المحاصيل في مصر منخفض بالمقياس العالمي

و الجدول رقم (١١) يوضح توزيع الأراضي بالمحافظات وفقاً للقدرة الإنتاجية وسنة الحصر ومنه يتبين أن :

(أ) مساحة الأراضي ذات الرتب الإنتاجية العالية تمثل حوالي ٥,٩ ٪ فقط من جملة الأراضي المزروعة .

(ب) مساحة الأراضي ذات الإنتاجية الجيدة تمثل حوالي ٤٥,١ ٪ .

(ج) مساحة الأراضي ذات الإنتاجية المتوسطة تمثل ٢٨,٨ ٪ .

(د) مساحة الأراضي منخفضة الإنتاجية تمثل ١٠,٢ ٪ .

وهذا يعنى أن حوالى نصف المساحة المزروعة متوسطة أو ضعيفة الإنتاجية . مع تدهور خصوبة التربة فى كثير من الأراضى الزراعية فى السنوات الأخيرة مما أدى إلى تباطؤ معدل نمو الإنتاج الزراعى ويعزى تدهور خصوبة التربة إلى الإسراف فى الري وسوء الصرف وكذلك عدم كفاءة وقاعية الإجراءات الجارية لتحسين التربة وصيانتها .

٢ - ضعف الإنتاج الحيوانى والداجنى فى مصر :

تعانى الثروة الحيوانية والداجنة من انخفاض فى كفاءتها الإنتاجية والتناسلية . حيث يقدر متوسط إنتاجية البقرة من الألبان فى السنة بنحو عشر المتوسط العالمى ، أما إنتاجية الماشية من اللحوم فهى منخفضة للغاية حيث لا يتجاوز متوسط وسوف نستعرض فيما يلى وسائل تحقيق الأمن الغذائى : إنتاجية الرأس من اللحم فى السنة نصف المتوسط العالمى . وقد يعزى ذلك لعدم وجود تخطيط واضح لرفع إنتاجية هذه الحيوانات وكذلك نقص الأعلاف خاصة الخضراء اللازمة لتغذية الحيوانات وذلك لعدم ثبات وتوازن المساحات المزروعة بالأعلاف . وليس الإنتاج الداجنى بأحسن حالاً ، إذ أن سلالات الدواجن المحلية منخفضة الإنتاجية من حيث عدد البيض المنتج فى السنة ومتوسط وزن البيضة والكفاءة التحويلية وسرعة النمو .

٤ - انخفاض حجم الإنتاج السمكى :

رغم أن المساحة المائية (البحار والبحيرات والأنهار) تقدر بنحو ٦,٩ مليون فدان ، إلا أن مصر فقيرة نسبياً فى إنتاجها السمكى . حيث أدى ضعف الاهتمام بالثروة السمكية إلى عدم نموها بالقدر اللازم لسد فجوة نقص البروتينات فى غذاء الإنسان المصرى . وهناك أسباب أخرى لعدم كفاية الإنتاج من هذا المصدر الغذائى الفنى منها انخفاض عدد الصيادين ، وارتفاع تكلفة النقل .

٥ - الفاقد من الإنتاج الزراعى :

هناك نوعان من الفاقد هما^(١) :

(أ) - الفاقد المادى من السلع الغذائية ذاتها ويتم خلال فترة الحصاد وحتى الوصول للمستهلك . ويعتبر هذا الفاقد كبيراً وتبلغ قيمته حوالى ١,٨ مليار دولار .

(ب) - الفاقد فى الموارد الأرضية الزراعية ويتمثل فى خروج مساحات من الأراضى من مجال الإستغلال الزراعى إلى أغراض أخرى كالبناى والطرق وغيرها

والجدول رقم (١٢) يوضح نسبة الفاقد من الإنتاج الزراعى خلال عامى ٢٠٠١ ، ٢٠٠٧ ومنه يتبين أن :

١- نسبة الفاقد المادى تختلف من محصول إلى آخر فقد بلغت نحو ١٠,٨ ٪ من الإنتاج فى البقوليات عام ٢٠٠٧ ، أما المحاصيل النشوية يبلغ الفاقد منها نحو ١٥,٢ ٪ وترتفع هذه النسبة فى الفاكهة حيث تبلغ ٢١,٦ ٪ من إنتاجها عام ٢٠٠٧ ، بينما يبلغ الفاقد فى الحبوب حوالى ٤,٦ ٪ من الإنتاج فى نفس العام . كما يبلغ حوالى ٩ ٪ فى الخضروات الطازجة .

٢- بحساب كميات الفاقد من الإنتاج الزراعى يتبين مدى تأثير الكميات المعروضة للإستخدام بهذه الكميات المفقودة وبالعامل على الإقلال بقدر الإمكان منها يمكن زيادة المتاح للإستهلاك .

(١) رافت محمد بسرى " مشكلة الغذاء فى مصر وإمكانية إستخدام الوسائل والطرق الغير تقليدية لتحقيق الأمن الغذائى " أكاديمية ناصر العسكرية العليا ، كلية الدفاع الوطنى الدورة (٢٣) ، ١٩٩٤ ، ص ٥٨

جدول رقم (١١)

توزيع الأراضي على مستوى الجمهورية وفقاً للقدرة الإنتاجية وسنة الحصر *

الدرجة	المساحة الكلية		% من جملة الأراضي المزروعة
	الفدان	%	
الأراضي المزروعة			
الدرجة الأولى (ذات الإنتاجية العالية)	٢٤٩,٩	٤,٢	٥,٩
الدرجة الثانية (ذات الإنتاجية الجيدة)	٢٦٥٥,٨	٢٢,٠	٤٥,١
الدرجة الثالثة (ذات الإنتاجية المتوسطة)	٢٢٨٤,١	٢٢,٦	٢٨,٨
الدرجة الرابعة (ذات الإنتاجية المحدودة)	٦٠٣,٨	٧,٢	١٠,٢
جملة الأراضي المزروعة	٥٨٩٣,٦	٧١,١	١٠٠
أراضي الدرجة الخامسة			
١ - البور الصالح للزراعة	١٠٤٩,٤	١٢,٧	
٢ - البور المغمورة بالمياه	٥٦٦,٢	٦,٨	
أراضي الدرجة السادسة			
١ - بور غير صالحة للزراعة	٧٨,٢	٠,٩	
٢ - المنافع العامة	٧٠٠	٨,٥	
إجمالي الزمام	٨٢٨٧,٥	١٠٠	-

* سنة الحصر تختلف من محافظة لأخرى .

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، نشرة إحصاءات المساحات المحصولية والإنتاج النباتي ٢٠٠٥ ، يونيو ٢٠٠٧ ، ص ١٨٥ .

جدول رقم (١٢)

نسبة الفاقد من الإنتاج الزراعي خلال عامي ٢٠٠١ ، ٢٠٠٧

(الكمية بالآلاف طن)

البيان	الإنتاج		الفاقد		نسبة الفاقد من الإنتاج	
	٢٠٠١	٢٠٠٧	٢٠٠١	٢٠٠٧	٢٠٠١	٢٠٠٧
الحبوب	١٧٦٨٠	٢٠٠٢٩	٧٨٤	٩٥٥	٤,٤	٤,٦
البقوليات	٥٢٦	٤٠٧	٤٠	٤٠	٧,٦	٩,٨
المحاصيل النشوية	٢٢٩٣	٢٢٦٢	٢٣٠	٢٩٩	١٠	٩,٢
الخضروات الطازجة	١١٧٥٤	١٥٦٥٩	١٠٦٦	١٤٢٦	٩,١	٩,١
الفاكهة *	٩٢٢٦	١١٧٨٠	٨٥٦	١١٣١	٩,٣	٩,٦
المحاصيل السكرية	١٨٦٣٢	٢٢٤٧٢	١٥٦	٢١٤	٠,٨	١

* تشمل (الفاكهة الطازجة - الشمام - البطيخ - الموالح) .

المصدر : الجهاز المركزي للتعبئة العامة والإحصاء ، دراسة تطور حركة الإنتاج والتجارة الخارجية والمتاح للإستهلاك من السلع الزراعية ، أعداد مختلفة .

٦ - التطور السلبي لسلوكيات الفلاح المصري :

إن تسرب العمالة من الريف أدى إلى العجز في هذه العمالة وبالتالي إلى انخفاض في إنتاج الأراضي الزراعية مما أثر سلباً على إنتاج الغذاء لأنه كلما انخفضت نسبة السكان الزراعيين التقليديين دون تقدم واضح في تكنولوجيا الاستغلال والإنتاج الزراعي كلما انخفض الإنتاج الغذائي وازدادت بالتالي مخاطر الاعتماد على الخارج للتزود بالغذاء .^(١)

وإذا تتبعنا تطور العمالة الزراعية نجد إنها في تناقص حيث بلغت نسبتها ٥٢,٠ ٪ من جملة عدد المشتغلين عام ١٩٦٠/٥٩ وفى عام ١٩٧٠/٦٩ بلغت نسبتها ٤٨,٩ ٪ ثم انخفضت إلى ٢٩,٠ ٪ عام ١٩٨٢/٨١ ثم توالى الانخفاض حتى بلغت ٢٧,٠ ٪ من جملة المشتغلين عام ٢٠٠٧/٠٦ وعلى العكس تزايدت في بقية القطاعات الخرى خلال نفس الفترة .

٧ - ضعف معدل التنمية في القطاع الزراعي :

انخفض نصيب الزراعة في الاستثمار الكلى من ٢٦ ٪ في الفترة من (٦٠ / ١٩٦١ - ٦٤ / ١٩٦٥) إلى ٢٢ ٪ في الفترة (٦٥ / ١٩٦٦ - ٦٩ / ١٩٧٠) , ثم إلى ١٠,٤ ٪ في الفترة (٧٠ / ١٩٧١ - ٧٤ / ١٩٧٥) , ثم هبط إلى ٧ ٪ في الفترة (٧٥ / ١٩٧٦ - ٧٩ / ١٩٨٠) , وإلى ٧,٧ ٪ في الخطة الخمسية (٨٧ - ١٩٩٢) مما يشير إلى اتجاه طويل الأجل لانخفاض الأهمية النسبية للاستثمار في الزراعة .

٨ - التوسع العمراني والكماش الأرض الزراعية :

كان لاتساع رقعة المدن في مصر على حساب الأراضي الزراعية الجيدة والتي تحاول الدولة المحافظة عليها وزيادتها أثر سلبى على تحقيق الاكتفاء الذاتى من الغذاء , لقد شهد العقدين الأخيرين من القرن السابق إزدياداً مطرداً نحو تحويل شطر كبير من الأرض الزراعية للإستعمالات اللأزراعية سواء في الإسكان أو إقامة المنشآت التصنيعية أو إقامة المرافق العامة .

٩ - الإفراط في الإستيراد وإنتهاج سياسة الدعم في توفير الغذاء :

لقد أدت سياسة الدولة باستيراد الغذاء من الخارج وإنتهاج سياسة الدعم إلى عدم الاهتمام الكافى بالإنتاج المحلى وتحسين الإنتاج بالكيف المطلوبين , الأمر الذى أدى إلى زيادة الاعتماد على معونات الحبوب الغذائية وبالتالي الإهمال فى تطوير وزيادة الإنتاج المحلى ومن ثم زيادة حدة مشكلة الغذاء وإستمرار الفجوة الغذائية بل واتساعها .

١٠ - النمط الغذائي المصري والأمية الغذائية :

على الرغم من التحسن الكمى فى متوسط إستهلاك الإنسان المصري من المواد الغذائية , إلا أن هذا التحسن الكمى لم يقرن بتحسن نوعى . وكذلك فإن المكونات الأساسية لإستهلاك الإنسان المصري تسودها أغذية الطاقة وتقل فيها الأغذية الواقية والبناءة .

(١) معهد التخطيط القومى "التنمية الإقتصادية الزراعية فى مصر "مذكرة داخلية رقم (٣١٤) , يونيو ١٩٧٣ , ص ١٠

(٢) محمد عبد الغنى سعودى , وسيم عبد الحميد , "السكان والغذاء فى مصر " مرجع سابق , ص ٦٩

أى أن هناك اختلالاً فى المركب الغذائى المصرى ونتيجة لذلك فإن كثير من أمراض المصريين هى أمراض نقص التغذية وسوء التغذية والتي تؤدى إلى انخفاض إنتاجية العمل ، وبالتالي انخفاض الناتج القومى .^(١)

أما الأمية الغذائية فتعرف بأنها " عدم تعرف الأم على الأصول الصحيحة لتغذية نفسها أثناء الحمل وكذلك أثناء الرضاعة وعدم تعرفها على القواعد السليمة لتغذية الطفل أثناء فترة الفطام وبعده إلى ما قبل سن المدرسة " . ويمكن القول إن الوجبة المصرية عموماً فيها مبالغة فى تناول النشويات والخضر والفاكهة

ثانياً : امكانات تحقيق الامن الغذائى

تعتبر معدلات الاستهلاك الفردى سواء الحالية او المستقبلية هى المؤشر السليم لقياس مدى تحقيق الامن الغذائى . وعليه فان مواجهة العجز الغذائى تحتاج الى مجموعة تدابير واجراءات تضمن توفير السلع الغذائية .

١ - مفهوم الأمن الغذائى :

(٢)

يشتمل مفهوم الأمن الغذائى بالنسبة للمواطن على ثلاث أركان :

- (أ) - الوفرة أى توافر السلع الغذائية التى يحتاج إليها المواطن فى حياته اليومية فى السوق المحلى .
 - (ب) - الاستقرار أى أن تكون هذه السلع متوافرة طوال الوقت بحيث تكون متاحة عند احتياجه إليها .
 - (جـ) - إمكانية الحصول عليها بحيث يكون دخله كافياً لتمكينه من شراء ما يحتاج إليه من سلع غذائية دون عناء .
- ويلاحظ أن الوفرة يجب أن تكون أساساً من خلال الإنتاج المحلى ومن ثم تسعى جميع الدول للإرتقاء بالكفاية الإنتاجية والتنافسية لزراعتها المحلية لتوفير أكبر قدر من احتياجات مواطنيها بصورة آمنة ومن ثم زيادة نسبة الإكتفاء الذاتى .

٢ - البنية الأساسية المحفزة لتحقيق الأمن الغذائى :

إن توفير بنية أساسية محفزة للتنمية الزراعية المتسارعة لتحقيق الأمن الغذائى أمر بالغ الحيوية ، فإى إستراتيجية للإرتقاء بالزراعة ومن ثم إنتاج الغذاء ، لابد أن تعطى أولوية ثلاث مداخل رئيسية وهى ما يلى :

- (أ) - الإستحواذ على التكنولوجيا باعتبارها الوسيلة الفعالة لتعظيم الإستفادة من الموارد الطبيعية المتاحة المحدودة .
- (ب) - تنمية الموارد البشرية القائمة بعملية التنمية الزراعية .
- (جـ) - المناخ الإقتصادى المحفز للتنمية الزراعية .

٢ - وسائل تحقيق الأمن الغذائى :

لا خلاف على أن المساحة المنزرعة لا تلاحق التزايد المستمر فى حجم السكان ، مما يجعل تحقيق الإكتفاء الذاتى أمراً صعباً ، وبالتالي فإن بناء قاعده إنتاجية لتوفير الأمن الغذائى أصبح أمراً حتمياً ويقتضى ذلك توسيع الإنتاج الزراعى وتعظيمه مع العمل على تصحيح الاختلال فى النمط الغذائى المصرى وترشيد الاستهلاك وسوف نستعرض فيما يلى وسائل تحقيق الأمن الغذائى .:

(١) مجلس الشورى، السياسة الغذائية فى ج ٢٠٠٤ ع ، لجنة الإنتاج الزراعى والرى وإستصلاح الأراضى ، تقرير رقم (٥) ، ١٩٩٢

(٢) محمد السيد عبد السلام ، الأمن الغذائى للوطن العربى ، عالم المعرفة ، الكويت ، ١٩٩٨

أ - التوسع الزراعي الرأسي :

- إن تحقيق الأمن الغذائي يتطلب إحداث ثورة زراعية تنتقل من زراعة تقليدية إلى زراعة عصرية متطورة. ومن خلال استخدام الوسائل التكنولوجية الحديثة التي يؤدي استخدامها إلى زيادة الإنتاج الزراعي وزيادة العائد الإقتصادي . ويمكن تحقيق أفضل استخدام للموارد الزراعية بالوسائل التالية :
- صيانة خصوبة التربة ومعالجة أي تدهور في خواصها .
- ترشيد استخدام مياه الري سواء مياه النيل أو المياه الجوفية إلى جانب منع تلوث المجارى المائية والمعالجة بإعادة استخدام الصرف الصناعي والزراعي .
- ميكنة العمليات الزراعية .
- تخريم تجريف الأراضي الزراعية , ووقف الزحف العمراني على الأراضي الخضراء , ومنع تصحر الأراضي الزراعية .
- زيادة الغلة الفدائية للمحاصيل الزراعية , حيث يمكن رفع متوسط إنتاجية الفدان من القمح والذرة والأرز وقصب السكر إلى أكثر من ضعف المتوسط الحالي .
- إعادة تخطيط التركيب المحصولي بهدف تحقيق أعظم استغلال من الرقعة الزراعية المحددة .

ب - التوسع الزراعي الأفقي :

- يمكن تحقيق ذلك عن طريق استصلاح الأراضي حيث إن زيادة المساحة المنزرعة لا تقل أهميته عن تكثيف إنتاجها فكل منها بعد جوهري من أبعاد التنمية الزراعية , فعمليات الاستصلاح للأراضي لا يجب أن تتوقف حتى استغلال كل قطرة من الماء , وذلك لمواجهة مطالب الأمن الغذائي للسكان الذين يتزايدون بإطراد .

ج - ترشيد الاستهلاك :

- ترشيد الاستهلاك أصبح ضرورة حتمية خاصة وأن معدل التزايد السكاني يفوق معدلات الانتاج , ويمكن ترشيد الاستهلاك بعدة طرق منها :

(أ) عمل برامج توعية وإرشاد المستهلكين لأنماط الغذاء الصحية ومنخفضة التكاليف , والعمل على تقليل نسبة الانفاق على الغذاء بأساليب مختلفة .

(ب) تعديل أنماط الاستهلاك الغذائي غير الرشيدة , والعمل على إعداد نمط غذائي مصرى يحقق إشباع الحاجات الغذائية الأساسية طبقاً للإحتياجات الصحية وعلى أساس المعايير الدولية .

(ج) ترشيد دعم السلع الغذائية , فعلم الترشيح يؤدي لسوء استخدام السلع المدعومة مثل استخدام الخبز كعلف رخيص للحيوان وزيادة أعباء الدولة لتدبير النقد الأجنبي اللازم لإستيراد السلع وبالتالي زيادة أعباء الدعم وتشجيع الاستهلاك على حساب الادخار .

(و) تكثيف برامج التوعية الغذائية عن طريق وسائل الاعلام المختلفة للتأثير على الأنماط السلوكية للمواطنين , وحثهم على تقليل الفاقد وترشيد الاستهلاك الغذائي .

التوصيات

بعد استعراضنا السابق لكل فصول الدراسة فقد خالصنا إلى التوصيات الآتية:

أولاً: في مجال زيادة إنتاج الغذاء:

- ١ - قضية الفاقد في الإنتاج الزراعي تحتاج إلى جهاز أو هيئة خاصة تعمل على دراسة الموضوع بصورة اشمول وتمثل فيها الهيئات العامة المرتبطة بالنشاط ذاته، وأن توضع خطة زمنية محددة لرفع كفاءة هذا القطاع ومتابعة تقييم ما تم إنجازه على مستوى القطاع الزراعي الشامل لتقليل هذا الفقد أو الهدر الاقتصادي.
- ٢ - الاتجاه نحو توسيع الرقعة المنزرعة إلى أقصى حد في نطاق مواردنا المائية الحالية أو التي يمكن تدبيرها من المصادر المختلفة، ومعالجة مياه الصرف الزراعي مع ترشيد الاستخدامات الحالية للمياه.
- ٣ - التركيز على إنتاج محاصيل تحقق البلاد فيها ميزة نسبية وإمكانات مادية وفنية يمكن من خلالها تحقيق حجم كبير من الإنتاج ومستوى جيد من الإنتاجية.
- ٤ - زيادة إنتاج المحاصيل البستانية (الخضر والفاكهة) والتي تتصف بأقل المتوسطات الإنتاجية على الصعيد العالمي، في حين إنه من اليسور رفع إنتاجيتها عن طريق تحسين الأصناف، باستخدام أساليب الهندسة الوراثية والتجهين واستيراد أصناف عالية الإنتاج من الخارج.
- ٥ - مراعاة أهمية التوجه نحو التنسيق بين السياسات الزراعية وخطة التنمية بين الدول العربية - كلما أمكن ذلك - لسد الفجوة الغذائية العربية، خاصة مع توافر مقومات التكامل في هذا المجال من حيث المساحات الزراعية الكبيرة ووفرة رؤوس الأموال، وتوافر الخبرات الفنية، والأيدي العاملة الزراعية.
- ٦ - ضرورة الإهتمام بدقة البيانات والإحصاءات وشمولها وسلامتها، وصدورها بصفة منتظمة، حتى يمكن لوحدة الإنتاج ووحدات الاستهلاك التعرف على الموقف من كل نواحيه بشكل صحيح وواقعي.

ثانياً: في شأن تنظيم إستهلاك الغذاء:

- ٧ - العمل على ضبط معدل النمو السكاني إلى أقل حد ممكن من خلال تكثيف جهود التنمية الحقيقية في كل المجالات، ومن خلال برامج فعالة للتوعية والإرشاد في مجال تنظيم الأسرة.
- ٨ - ترشيد أنماط الإستهلاك لدى المواطنين، وكبح جماح الإستهلاك غير الضروري بالتركيز على برامج إعلام وتوعية المستهلكين بالثقافة والسلوكيات الغذائية السليمة
- ٩ - دراسة حالة الأسواق العالمية وتتبعها بإستخدام الأساليب العلمية والفنية من أجل تنظيم عمليات الإستيراد والجهات التي نستورد منها السلع الغذائية، وتوجيه القطاع الخاص العام والتعاوني إلى أفضل الأسواق التي يمكن شراء السلع الغذائية منها، ومعرفة التوقيت المناسب للإستيراد طبقاً للأساليب الاقتصادية والتجارية، مع تنويع دول الإستيراد خاصة في السلع الإستراتيجية كالقمح والزيوت.

ثالثاً: الدور الحكومي:

١٠ - استمرار الدولة في تبنى نشاط استصلاح الأراضي وتوزيعها على الشباب بأسعار تقسط على آجال طويلة، فرغم العبء المادى الكبير لهذا الأمر فإن له ثمرته المزدوجة من حيث تنمية الإنتاج الزراعى عامة والفدانى خاصة إلى جانب توفير مصدر دخل للشباب المتعطل والراغب فى تكوين الأسر.

١١ - مراجعة المواصفات القياسية القائمة للسلع الفدائية، وتعديلها بما يتوافق مع المواصفات القياسية التى وضعتها منظمة الأغذية والزراعة للأمم المتحدة بمعرفة هيئة الدستور الفدانى، واعتبارها دستوراً يلتزم به المنتجون وتزاول أجهزة الرقابة عملها فى إطارها.

١٢ - العمل على تشجيع صناعة الأعلاف من المنتجات الثانوية للمحاصيل والمخلفات الصناعية لزيادة عرض الأعلاف بما يخفف من توجيه إنتاج الحبوب خاصة القمح للإستهلاك الحيوانى.

رابعاً: الدور الغير الحكومي:

١٣ - إشترك المنظمات غير الحكومية مع الجهات الحكومية المعنية فى تبنى برامج تحسين المستوى الفدانى لجموعات مستهدفة. وذلك مثل التغذية المدرسية وتغذية عمال المصانع والمستشفيات والوحدات الصحية مع الإستعانة بخبراء التغذية فى تحديد أصناف الوجبات ذات المستوى التغذوى الأعلى والتكلفة الأقل على قدر الإمكان، مع تنويعها باستمرار.

١٤ - تشجيع جمعيات حماية المستهلك على القيام ببلورها مع تدريب أعضائها وتعريفهم بأبعاد الدور الذى يمكنهم القيام به لمساعدة أجهزة الرقابة فى ضوء خطة معلومة لمسئولياتهم.

١٥ - مشاركة الجهات غير الحكومية فى التثقيف الفدانى بجانب أجهزة الدولة المعنية مثل وزارات الإعلام والتعليم والصحة والتضامن الإجتماعى، ومن الوسائل الأساسية التى يمكن إستخدامها: المحاضرات المدرسية، والندوات فى الوحدات الإجتماعية، والنوادر ودور العبادة والمصحات فى المدارس والمستشفيات والوحدات الصحية والجمعيات الزراعية وجهات العمل وغيرها.

المراجع

أولاً - التقارير المنشورة :

- ١ - رئاسة الجمهورية ، المجلس القومية المتخصصة ، سياسات الدعم - الأهداف والأشوار والتوجهات المستقبلية ، تقرير المجلس القومى للإنتاج والشنون الإقتصادية ، الدورة (٢٢) ، ٢٠٠٥ / ٢٠٠٦ .
- ٢ - رئاسة الجمهورية ، المجلس القومية المتخصصة ، نمط الإستهلاك الغذائى المصرى وتوقعاته المستقبلية ، تقرير المجلس القومى للإنتاج والشنون الإقتصادية ، الدورة (٢٤) ، ١٩٩٧ / ١٩٩٨ .
- ٣ - مجلس الشورى ، السياسة الغذائية فى جمهورية مصر العربية ، لجنة الإنتاج الزراعى والرى وإستصلاح الأراضى ، تقرير رقم (٥) ، ١٩٩٢ .
- ٤ - المنظمة العربية للتنمية الزراعية ، تقرير أوضاع الأمن الغذائى العربى ، ٢٠٠٢ .
- ٥ - وزارة الدولة للتنمية الإقتصادية ، الخطة الخمسية للتنمية الإقتصادية والإجتماعية (٢٠٠٧ / ٢٠٠٨ - ٢٠١١ / ٢٠١٢) وخطة عامها الأول ، ٢٠٠٧ .

ثانياً - الرسائل العلمية :

- ١ - رأفت محمد يسرى ، مشكلة الغذاء فى مصر وإمكانية إستخدام الوسائل والطرق الغير تقليدية لتحقيق الأمن الغذائى ، أكاديمية ناصر العسكرية العليا ، كلية الدفاع الوطنى ، الدورة (٢٢) ، ١٩٩٤ .

ثالثاً - الكتب :

- ١ - محمد السيد عبد السلام ، الأمن الغذائى للوطن العربى ، عالم المعرفة ، المجلس الوطنى للثقافة والفنون والآداب ، الكويت ، ١٩٩٨ .
- ٢ - محمد عبد الفنى ، وسيم عبد الحميد ، السكان والغذاء فى مصر ، الأهرام الإقتصادى ، العدد (٧٩) ، ١٩٩٤ .
- ٣ - منتدى العالم الثالث ، الزراعة والغذاء فى مصر ، الواقع وسيناريوهات بديلة حتى عام ٢٠٢٠ ، مكتبة مصر ، ٢٠٠١ .

رابعاً - الإحصاءات الرسمية :

- ١ - الجهاز المركزى للتعبئة العامة والإحصاء ، دراسة تطور حركة الانتاج والتجارة الخارجية والمتاح للاستهلاك من السلع الزراعية ، اعداد مختلفة .
- ٢ - الجهاز المركزى للتعبئة العامة والإحصاء ، الكتاب الإحصائى السنوى اعداد مختلفة .
- ٣ - الجهاز المركزى للتعبئة العامة والإحصاء ، نشرة الزمام والملكية الزراعية ٢٠٠٥ ، يونيه ٢٠٠٧ .

٤ - الجهاز المركزي للتعبئة العامة والإحصاء ، إحصاءات المساحات المحصولية والإنتاج النباتي
٢٠٠٥ ، ٢٠٠٧ .

٥ - الجهاز المركزي للتعبئة العامة والإحصاء ، مصر في أرقام ٢٠٠٦ ، يونيو ٢٠٠٧ .

٦ - الجهاز المركزي للتعبئة العامة والإحصاء ، النتائج النهائية للتعداد العام للسكان والإسكان
والمنشآت - مصر ٢٠٠٦ ، مارس ٢٠٠٧ .

٧ - الجهاز المركزي للتعبئة العامة والإحصاء ، إستراتيجية الأمن الغذائي في مصر ،
يونيه ١٩٩٨ .

٨ - وزارة الزراعة واستصلاح الأراضي ، الإدارة المركزية للإقتصاد الزراعي ، أهم مؤشرات
الإحصاءات الزراعية ، الجزء الثاني ، (المحاصيل الصيفية والنييلية) ٢٠٠٥ / ٢٠٠٦ ،
نوفمبر ٢٠٠٧ .

٩ - وزارة الزراعة واستصلاح الأراضي ، قطاع الشئون الإقتصادية ، تقديرات الميزان الغذائي
جمهورية مصر العربية ، ١٩٩١ .

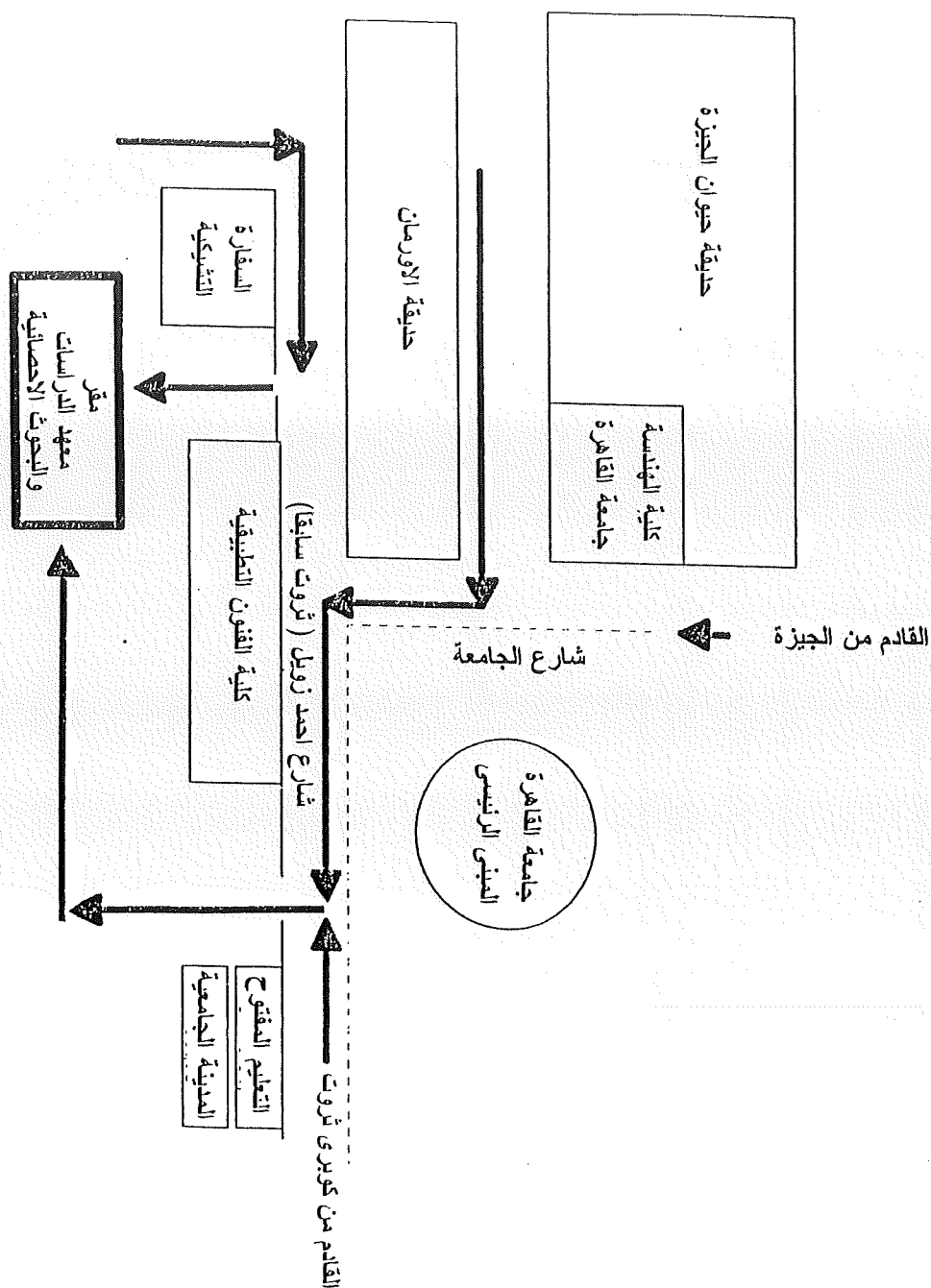
١٠ - منظمة الأغذية والزراعة للأمم المتحدة ، الكتاب الإحصائي السنوي ، المجلد الثاني ،
٢٠٠٥ - ٢٠٠٦ ، ٢٠٠٦ .

١١ - الهيئة العامة للمستشفيات والمعاهد التعليمية ، المعهد القومي للتغذية ، بيانات غير
منشورة .

خامساً - الدوريات والصحف والمجلات:

١ - وزارة الإعلام ، الهيئة العامة للإستعلامات ، مجلة النيل ، العدد (٨٣) ، ٢٠٠٣ .

٢ - معهد التخطيط القومي ، التنمية الإقتصادية الزراعية في مصر ، مذكرة داخلية رقم
(٣١٤) ، ١٩٧٣ .



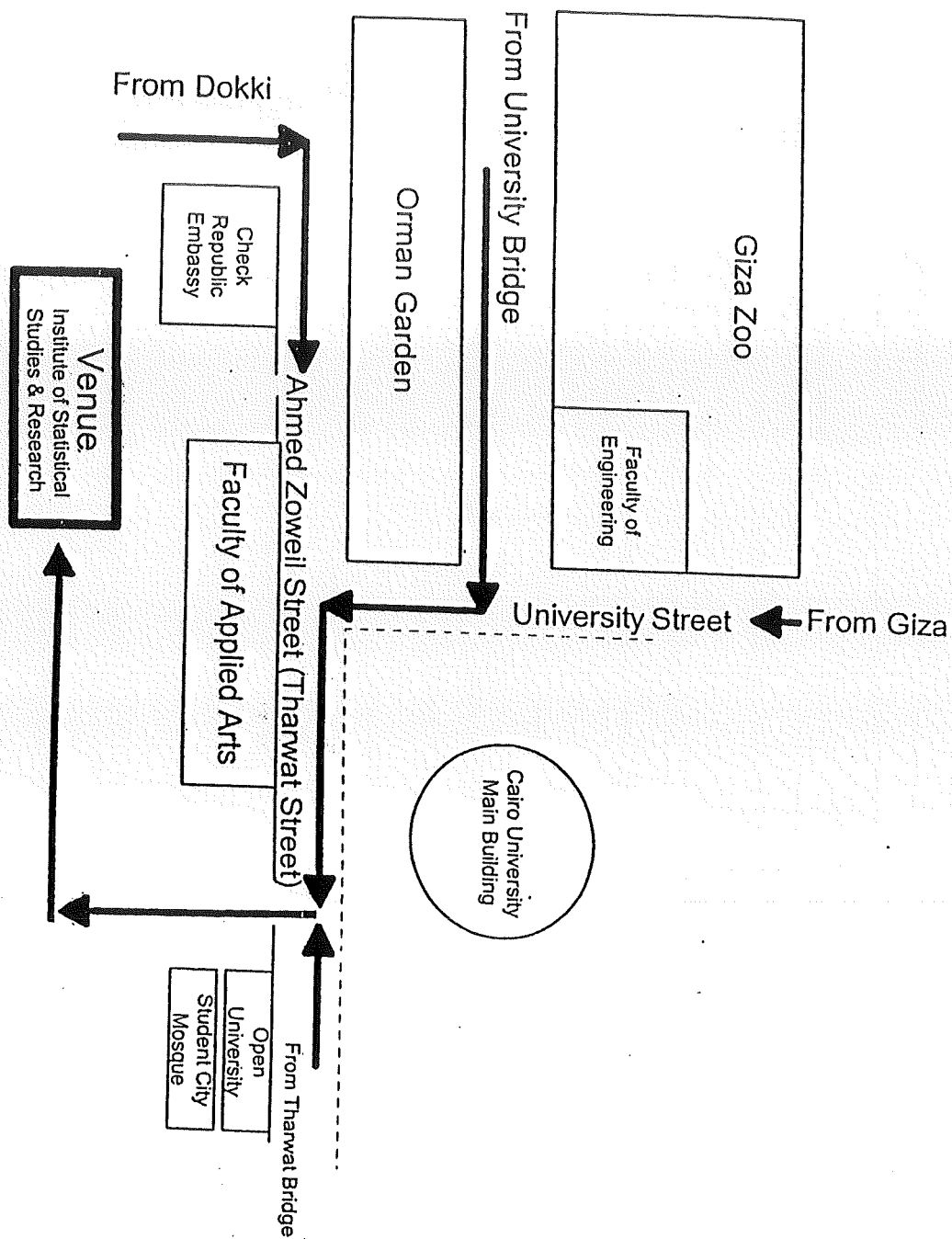


Table (2b) : The true estimated values and the correction estimators(*) of β_0 , β_1 and β_2 , their biases and standard errors for the different correction method with $\sigma_e^2 = 0.25$ and for the given values of n and σ_u^2

		CLS		CSF		MRC		AD		SIMEX	
		c. e.	bias	c. e.	bias	c. e.	bias	c. e.	bias	c. e.	bias
$n=400$, $\sigma_e^2=0.25$, $\sigma_u^2=0.5$	$\hat{\beta}_0$	0.383	-0.621	0.385	-0.619	0.929	0.075	1.064	0.060	0.976	-0.028
	SD	(0.061)		(0.058)		(0.026)		(0.036)		(0.062)	
	$\hat{\beta}_1$	1.614	0.644	2.121	1.151	1.140	-0.170	1.008	0.038	1.039	0.069
	SD	(0.037)		(0.035)		(0.021)		(0.028)		(0.034)	
	$\hat{\beta}_2$	-0.345	0.146	-0.141	0.350	-0.529	0.038	-0.530	-0.039	-0.390	0.101
	SD	(0.025)		(0.026)		(0.015)		(0.021)		(0.016)	
$n=800$, $\sigma_e^2=0.25$, $\sigma_u^2=0.25$	$\hat{\beta}_0$	0.405	-0.587	0.407	-0.585	0.942	0.050	1.045	0.053	0.878	-0.044
	SD	(0.059)		(0.058)		(0.034)		(0.036)		(0.054)	
	$\hat{\beta}_1$	1.480	0.491	1.857	0.868	1.137	-0.148	1.013	0.024	1.127	0.138
	SD	(0.038)		(0.036)		(0.031)		(0.033)		(0.036)	
	$\hat{\beta}_2$	-0.467	0.024	-0.455	0.036	-0.641	0.150	-0.540	-0.049	0.036	0.527
	SD	(0.027)		(0.025)		(0.024)		(0.021)		(0.021)	
$n=800$, $\sigma_e^2=0.25$, $\sigma_u^2=0.5$	$\hat{\beta}_0$	0.384	-0.608	0.387	-0.605	0.893	0.099	1.069	0.077	0.662	-0.330
	SD	(0.051)		(0.057)		(0.039)		(0.038)		(0.049)	
	$\hat{\beta}_1$	1.530	0.541	2.496	1.507	1.324	-0.335	1.024	0.035	0.689	-0.300
	SD	(0.035)		(0.036)		(0.039)		(0.037)		(0.025)	
	$\hat{\beta}_2$	-0.243	0.248	-0.236	0.255	-0.828	0.337	-0.547	-0.056	-0.169	0.322
	SD	(0.020)		(0.024)		(0.034)		(0.022)		(0.011)	

Table (2a) : The true estimated values and the correction estimators(*) of β_0, β_1 and β_2 , their biases and standard errors for the different correction method with $\sigma_e^2 = 0.25$ and for the given values of n and σ_u^2

		CLS		CSF		MRC		AD		SIMEX	
		c. e.	bias	c. e.	bias	c. e.	bias	c. e.	bias	c. e.	bias
$n=200, \sigma_e^2=0.25, \sigma_u^2=0.25$ The true estimated values: $\hat{\beta}_0 = 0.876, \hat{\beta}_1 = 0.528,$ $\hat{\beta}_2 = 0.071$	$\hat{\beta}_0$	0.771	0.105	0.780	-0.096	0.872	0.004	0.879	0.003	0.732	-0.144
	SD	(0.059)		(0.063)		(0.034)		(0.037)		(0.058)	
	$\hat{\beta}_1$	1.140	0.612	1.120	0.592	0.646	-0.118	0.552	0.024	0.582	0.054
	SD	(0.044)		(0.047)		(0.027)		(0.031)		(0.038)	
	$\hat{\beta}_2$	0.146	0.075	0.241	0.170	0.088	-0.017	0.044	-0.027	0.130	0.059
	SD	(0.028)		(0.030)		(0.016)		(0.025)		(0.021)	
$n=200, \sigma_e^2=0.25, \sigma_u^2=0.5$ The true estimated values: $\hat{\beta}_0 = 0.876, \hat{\beta}_1 = 0.528,$ $\hat{\beta}_2 = 0.071$	$\hat{\beta}_0$	0.425	-0.451	0.426	-0.450	0.858	0.018	0.832	-0.044	0.894	0.021
	SD	(0.051)		(0.061)		(0.037)		(0.041)		(0.045)	
	$\hat{\beta}_1$	1.440	0.912	1.230	0.702	1.020	-0.492	0.505	-0.023	0.708	0.180
	SD	(0.034)		(0.039)		(0.031)		(0.033)		(0.023)	
	$\hat{\beta}_2$	0.072	0.001	0.071	0.000	-0.359	0.430	0.099	0.028	0.018	-0.053
	SD	(0.027)		(0.031)		(0.021)		(0.021)		(0.021)	
$n=400, \sigma_e^2=0.25, \sigma_u^2=0.25$ The true estimated values: $\hat{\beta}_0 = 1.004, \hat{\beta}_1 = 0.970,$ $\hat{\beta}_2 = -0.491$	$\hat{\beta}_0$	0.416	-0.588	0.418	-0.586	0.974	0.030	1.042	0.038	0.923	-0.081
	SD	(0.063)		(0.059)		(0.035)		(0.034)		(0.062)	
	$\hat{\beta}_1$	1.420	0.450	1.680	0.710	0.929	0.041	1.004	0.034	1.001	0.031
	SD	(0.037)		(0.034)		(0.029)		(0.031)		(0.036)	
	$\hat{\beta}_2$	-0.465	0.026	-0.453	0.038	-0.457	-0.034	-0.532	-0.041	-0.315	0.176
	SD	(0.028)		(0.027)		(0.020)		(0.021)		(0.019)	

Appendix

Table (1) : Values of $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ (*), and their biases with different values of m for the given values of n, σ_u^2 and σ_e^2

σ_e^2		σ_u^2		n		$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
						e. v.	bias	e. v.	Bias	e. v.	bias	e. v.	bias	e. v.	bias
0.25	0.25	200				0.851	-0.025	0.894	0.018	0.879	0.003	0.868	-0.008	0.868	-0.008
						0.606	0.078	0.581	0.053	0.700	0.172	0.689	0.161	0.656	0.128
						0.102	0.031	0.052	-0.019	0.084	0.013	0.095	0.024	0.105	0.034
0.25	0.5	200				0.825	-0.051	0.887	0.011	0.827	-0.049	0.870	-0.006	0.881	0.005
						0.581	0.053	0.640	0.112	1.160	0.632	1.316	0.788	1.404	0.876
						0.142	0.071	-0.162	-0.091	-0.427	-0.356	-0.631	-0.702	-0.717	-0.788
0.25	0.25	400				0.963	-0.041	0.961	-0.043	0.980	-0.024	0.985	-0.019	0.979	-0.025
						0.935	-0.035	0.924	-0.046	0.881	-0.089	0.934	-0.036	0.971	0.001
						-0.489	0.002	-0.476	0.015	-0.416	0.075	-0.405	0.086	-0.498	-0.007
0.25	0.5	400				0.822	-0.182	0.935	-0.069	0.969	-0.035	0.965	-0.039	0.955	-0.049
						1.247	0.277	1.234	0.264	1.040	0.036	1.072	0.068	1.108	0.104
						-0.445	0.046	-0.627	-0.136	-0.481	0.010	-0.527	-0.036	-0.567	-0.076
0.25	0.25	800				0.906	-0.086	0.929	-0.063	0.951	-0.041	0.962	-0.030	0.964	-0.028
						1.017	0.028	1.132	0.143	1.165	0.176	1.175	0.186	1.198	0.209
						-0.511	0.020	-0.611	-0.120	-0.677	-0.186	-0.696	-0.205	-0.708	-0.217
0.25	0.5	800				0.861	-0.131	0.875	-0.117	0.899	-0.093	0.907	-0.085	0.923	-0.069
						1.035	0.046	1.279	0.290	1.379	0.390	1.440	0.451	1.489	0.500
						-0.564	-0.073	-0.759	-0.268	-0.872	-0.381	-0.939	-0.448	-1.007	-0.516

- [25] Nakamura, T. (1990); " Corrected Score Function for Errors-in-Variables Models: Methodology and Application to Generalized Linear Models " *Biometrika*, Vol. 77, pp. 127-137.
- [26] Olson, K. and Kennedy, C. (2006); " Examination of the Relationship Between Nonresponse and Measurement Error in a Validation Study of Alumni " *Proceeding of the Section on Survey Research Methods, American Statistical Association*, pp. 4181-4188.
- [27] Peytchev, A. (2006); " Estimation of Measurement Error and Identification of Causes: Linking Measurement Error to Nonresponse, Interviewers, and Interviewer Characteristics " *Proceeding of the Section on Survey Research Methods, American Statistical Association*, pp. 3528-3535.
- [28] Prentice, R. (1982); " Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model " *Biometrika*, Vol. 69, pp. 331-342.
- [29] Schafer, D. (1986); " Combing Information on Measurement Error in the Errors-in-Variables Model " *JASA*, Vol. 81, pp. 181-185.
- [30] Schafer, D. (2001); " Semiparametric Maximum Likelihood for Measurement Error Model Regression " *Biometrics*, Vol. 57, pp. 53-61.
- [31] Stefanski, L. (1985); " The Effects of Measurement Error on Parameter Estimation " *Biometrika*, Vol. 72, pp. 583-592.
- [32] Stefanski, L. (1989); " Unbiased Estimation of a Nonlinear Function of a Normal Mean with Application to Measurement Error Models " *Commun. Statist.*, Vol. 18, pp. 4335-4358.
- [33] Stefanski, L. and Carroll, R. (1985); " Covariate Measurement Error in Logistic Regression " *Ann. Statist.* Vol. 13, pp. 1335-1351.
- [34] Stefanski, L. and Carroll, R. (1987); " Conditional Scores and Optimal Scores for Generalized Linear Measurement-Error Models " *Biometrika*, Vol. 74, pp. 703-716.
- [35] Walter, K. and Fuller, W. (1982a); " Estimation of Non-Linear Errors-in-Variables Models " *Ann. Statist.* Vol. 10, pp. 539-548.
- [36] Walter, K. and Fuller, W. (1982b); " Estimation of the Quadratic Errors-in-Variables Model " *Biometrika*, Vol. 69, pp. 175-182.
- [37] Whittemore, A. and Keller, J. (1988); " Approximations for Regression with Covariate Measurement Error " *JASA*, Vol. 83, pp. 1057-1066.

- Survey Research Methods, American Statistical Association, pp. 96-105.
- [13] Feldstein, M. (1974); " Errors in Variables: A Consistent Estimator with Smaller MSE in Finite Samples " JASA, Vol. 69, pp. 990-996.
- [14] Fuller, W. (1980); " Properties of Some Estimators for the Errors-in-Variables Model " Ann. Statist. Vol. 8, pp. 407-422.
- [15] Fuller, W. (1987); " Measurement Error Models " John Wiley and Sons, New York.
- [16] Hill, D. (1992); " Effects of Measurement Error on Occupational Event History Analysis " Proceeding of the Section on Survey Research Methods, American Statistical Association, pp. 572-577.
- [17] Huber, P. (1981); " Robust statistics " New York: Wiley.
- [18] Hutchison, D., Morrison, J. and Felgate, R. (2000); " Bootstrapping the Effect of Measurement Errors in Variables on Regression Results " Proceeding of the Section on Survey Research Methods, American Statistical Association, pp. 569-573.
- [19] Hwang, J. (1986); " Multiplicative Errors-in-Variables Models with Applications to Recent Data Released by the U. S. Department of Energy " JASA, Vol. 81, pp. 680-688.
- [20] Ibrahim, Y. (2005); " Alternative Estimators of the Slope in Simple Linear Regression with Measurement Error (An Empirical Comparison) " Cairo University, Institute of Statistical Studies and Research, The Annual Conference on Statistics, Computer Science and Operation Research, Applied Statistics, pp. 34-54.
- [21] Jang, D. and Eltinge, J. (1997); " Use of Within-PSU Variances and Errors-in-Variables Regression to Assess Stability of a Standard Design-Based Variance Estimator " Proceeding of the Section on Survey Research Methods, American Statistical Association, pp. 255-260.
- [22] Johnston, J. (1972); " Econometric Methods " Second Edition, McGraw-Hill Book Company.
- [23] Kendal, M. and Stuart, A. (1979); " Advanced Theory of Statistics " Second Edition, High Wycombe: Griffin.
- [24] Koutsoyiannis, A. (1973); " Theory of Econometrics " The Macmillan Press LTD.

values of the x_i 's (hetroscedasticity case). A correction method with the hetroscedasticity case will be the subject of a forthcoming study.

References

- [1] Anderson, T. (1984); " The 1982 Wald Memorial Lectures: Estimating Linear Statistical Relationships " Ann. Statist. Vol. 12, pp. 1-45.
- [2] Armstrong, B. (1985); " Measurement Error in the Generalized Linear Model " Comm. Statist., B, Vol. 14, pp. 529-544.
- [3] Bay, J. and Stefanski, L. (2000); " Adjusting Data for Measurement Error to Reduce Bias When Estimating Coefficients of a Quadratic Model " Proceeding of the Section on Survey Research Methods, American Statistical Association, pp. 731-733.
- [4] Biemer, P. and Atkinson, D. (1995); " Estimating Measurement Error Bias and Variance in Two- Phase Samples " Proceeding of the Section on Survey Research Methods, American Statistical Association, pp. 775-780.
- [5] Carroll, R., Ruppert, D. and Stefanski, L. (1995); " Measurement Error in Nonlinear Models " Chapman and Hall, London.
- [6] Carroll, R. and Stefanski, L. (1990); " Approximate Quasilikelihood Estimation in Models with Surrogate Predictors " JASA, Vol. 85, pp. 652-663.
- [7] Chan, L. and Mak, T. (1985); " On the Polynomial Functional Relationship " J. R. Statist. Soc., B, Vol. 47, pp. 510-518.
- [8] Cheng, C. and Schneeweiss, H. (1998); " Polynomial Regression with Errors in the Variables " J. R. Statist. Soc., B, Vol. 60, pp. 189-199.
- [9] Cochran, W. (1968); " Errors of Measurement in Statistics " Technometrics, Vol. 10, pp. 637-666.
- [10] Cook, J. and Stefanski, L. (1994); " Simulation- Extrapolation Estimation in Parametric Measurement Error Models " JASA, Vol. 89, pp. 1314-1328.
- [11] Cramér, H. (1961); " Mathematical Methods of Statistics " Ninth Printing, Princeton University Press.
- [12] Eltinge, J., Yansaneh, I. and Paulin, G. (1994); " Assessment of Reported Differences Between Expenditures and Low Incomes in the U. S. Consumer Expenditure Survey " Proceeding of the Section on

the bias and the standard error. All computation done by using EXCEL and SPSS programs.

As for the two applications presented in this article, a comparison between their results reveals the following; The adjusted data method showed superiority in both applications, followed by the modified regression calibration method.

Also it is noted that the correction methods presented in this article may be classified into three groups; The first group includes the adjusted data and the modified regression calibration methods. The adjusted data method depends on transforming the observed values $w_1, w_2, \dots, w_n$ to $x_1, x_2, \dots, x_n$ under certain constraints. The modified regression calibration method depends on estimating the values of the explanatory variable using the least squares method at different values of m , where m is the number of the replicate measurements of the observed values. The adjusted data method performed well in all cases of the simulation study, since the corrected estimators for this method had smaller biases and standard errors compared than the biases and standard errors of the other corrected estimators, followed by the MRC method. The converse is found for the second group which includes the corrected least squares and corrected score function methods. These methods depend on correcting the moments of the explanatory variable and the cross products of the dependent and explanatory variables. The third group includes the simulation-extrapolation method. This method depends on adding the generated measurement errors to the observed data to create the values of the explanatory variable. Although the corrected estimators for the simulation-extrapolation method had the best performance in the linear regression model (see Ibrahim (2005), these estimators had a moderate performance in most cases and a worst performance in the remainder cases of the simulation study.

Finally, all correction methods assume that the variance of each u_i ; $i=1,2,\dots,n$ is the same for all values of the explanatory variable (homoscedasticity case), but from the medical application presented in section 4, it is noted that the measurements are influenced by many factors. Moreover, the measurements may be taken in different times using more than one instrument by more than one doctor or nurse. So, the values of the measurement error variance σ_u^2 may depend on the

(II) For sample size $n = 400$

1. The MRC estimators have the best performance in the first case, followed by the AD estimators, and vice versa in the second case.
2. The SIMEX estimators have a moderate performance in the two cases.
3. The CLS and the CSF methods do not perform well in the two cases.

(III) For sample size $n = 800$

1. The AD estimators have the best performance in the two cases.
2. The MRC estimators have a moderate performance in the first case and a worst performance in the second case.
3. The SIMEX estimators perform poorly and are worst in the two cases, followed by CLS and CSF estimators respectively.

In general; The AD estimators performed well in all cases, followed by the MRC and SIMEX estimators respectively. The performance of the CLS and the CSF estimators are poor. The MRC and AD estimators have the least standard errors in all cases compared to the standard errors of the estimators of the other methods.

6. Conclusions

This article presented corrected least squares method and an extension of the corrected score function method, suggested by Nakamura (1990), to estimate the parameters of the polynomial regression model when the model (1.3) is applied. A modification of the regression calibration method, suggested by Carroll and Stefanski (1990), is also presented. The adjusted data method, suggested by Jeff and Stefanski (2000), and the simulation-extrapolation method, suggested by Cook and Stefanski (1994), are discussed to estimate the parameters of the polynomial regression model when the explanatory variable is measured with error.

Two applications are introduced to compute the corrected estimators of the parameters of the quadratic regression model, when the model (1.3) is applied, for the different correction methods presented in sections 2 and 3, and to compare these corrected estimators in terms of

1. The performance of the computed estimators from the different correction methods is not the same.
2. The values of the corrected estimators computed for MRC and AD methods are close, and similarly for CLS and CSF methods.
3. The AD estimators are almost superior to other estimators as they have smaller standard errors than that of the other corrected estimators.
4. The SIMEX estimators have the largest standard errors compared with the other estimators.

5. Simulation Study

A simulation study is performed to compare the corrected estimators in terms of the bias and standard error for the used correction methods of the coefficients of the regression model

$$y = 1 + x - 0.5x^2 + e \quad (5.1)$$

with $\sigma_e^2 = 0.25$. The true data $x_1, x_2, \dots, x_n$ were generated from a normal distribution with mean zero and variance one. The values of $y_1, y_2, \dots, y_n$ are computed using (5.1), then the OLS method is used to compute the true estimated values of the parameters. The model (5.1) was simulated with three different sample sizes, $n = 200, 400$ and 800 . For each sample size, the values of the true data were contaminated with one of two levels of measurement error variance $\sigma_u^2 = 0.25$ and $\sigma_u^2 = 0.5$.

From tables (2a) and (2b) in the appendix it is noted that;

(I) For sample size $n = 200$

1. The AD estimators have the best performance because they have the smallest biases compared with the other estimators in the two-variance-cases.
2. The MRC estimators have a good performance in the first case ($\sigma_u^2 = 0.25$) and a worst performance in the second case ($\sigma_u^2 = 0.5$).
3. The SIMEX estimators have a moderate performance in the two cases.
4. The performance of the CSF estimators is poor in the two cases, followed by the CLS estimators.

From the records of the patients received at the Nephrology department-Zone A in New Kasr Al-Aini Teaching Hospital during the years 2007 and 2008, a sample of size $n=100$ patient records is randomly selected. For each patient 5 measurements for systolic blood pressure and the level of the creatinine in blood are taken from the record of the patient. The scatter diagram of the level of the creatinine in blood, y , and the observed systolic blood pressure, w , showed that the relation between the two variables is nonlinear. With correct measurements, assuming that y has a quadratic regression on x . Then, the desired quadratic regression model is

$$y_i = \sum_{j=0}^2 \beta_j x_i^j + e_i \quad i=1, 2, 3, \dots, n \quad (4.1)$$

At first, σ_u^2 was estimated using (1.8), and it is found that $\hat{\sigma}_u^2 = 156.2$.

The naïve estimated quadratic regression model is

$$\hat{y}_i = -67.03854 + 0.92305 w_i - 0.00285 w_i^2$$

(18.306) (0.255) (0.001)

and the estimators of the parameters are found statistically significant at $\alpha = 0.05$

The correction methods presented in sections 2 and 3 are applied to this data set. Table (a) shows the corrected estimators of the parameters of the model (4.1) and their corresponding standard errors.

Table (a): Values of $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$, and their corresponding standard errors (*) for different correction methods

	The correction method				
	CLS	CSF	MRC	AD	SIMEX
$\hat{\beta}_0$	4.480	4.482	-12.864	-13.251	28.373
SD*	(2.643)*	(2.754)	(1.987)	(1.651)	(4.860)
$\hat{\beta}_1$	0.0145	0.035	0.214	0.136	-0.276
SD	(0.042)	(0.050)	(0.038)	(0.037)	(0.065)
$\hat{\beta}_2$	-0.0000117	-0.00001	-0.00057	-0.00001	0.00086
SD	(0.00004)	(0.00004)	(0.00007)	(0.00005)	(0.00021)

From table (a) we find that:

$$\hat{\sigma}_{\hat{\beta}_j} = \frac{1}{Z} \sum_{h=1}^Z \hat{\sigma}_{\hat{\beta}_{jh}} \quad j=0, 1, 2$$

7. If we Consider the values of $\hat{\beta}_{jh}$ as a dependent variable and η_h as an independent variable, then using the values of $\hat{\beta}_{jh}$ and η_h , an extrapolation function is fitted, which can be linear or quadratic according to the best fit.

8. By extrapolating the resultant function to the case where $\eta = -1$, we obtain the SIMEX estimator, which is consistent for β_j , $j=0, 1, 2$.

In the applications that are performed in this paper, the SPSS program was used to generate pseudo errors, new data sets and to calculate the estimators of the parameters of the quadratic regression model and their corresponding standard errors. η was taken the values 0.2, 0.4, 0.6, 0.8, 1.00, 1.2, 1.4, 1.6, 1.8 and 2 respectively. At each value of η the pseudo error vectors were generated 100 times, i. e. $b=100$. The extrapolation function of second degree $f(\beta_j) = a_0 + a_1\eta + a_2\eta^2$; $j=0, 1, 2$ was fitted by the least squares method using the points 0.2, 1 and 2, and extrapolated to the point $\eta = -1$ to compute the corrected estimators of the parameters of the quadratic regression model when the model (1.3) is applied.

4. Medical application

In the medical field, many metabolic measurements are measured using instruments such as the manometer which measures blood pressure, the thermometer which measures temperature. These measurements are influenced by many factors. Most of these factors are related to the patient such as exhaustion, nutrition, smoking, psychological status. The other factors belong to surrounding conditions during taking the measurement and imprecise instrument. So, these measurements are often measured with error.

In this application, the systolic blood pressure is selected as an explanatory variable x , and the level of the creatinine in blood is selected as the dependent variable y .

Due to the measurement errors the observed fallible systolic blood pressure is w where $w = x + u$ and u is the measurement error.

5. Repeating steps (2)–(4) until the estimated values of the parameters β 's converge.
6. Calculating the standard errors of $\hat{\beta}$'s as sample means of the standard errors computed in step (4).

In the presented application in this paper, a macro unit was added to the EXCEL program to find the roots of equation (3.6), and this program was used to apply the Newton-Raphson method. The SPSS program was used to estimate the parameters.

3.3 Simulation-Extrapolation (SIMEX) Method

Simulation-Extrapolation method was presented by Cook and Stefanski (1994). It is performed as follows;

1. In addition to the original data set, a number of data sets, say Z , are generated each with successively larger variance of measurement errors. These levels of the variance of measurement errors are calculated by the formula

$$(1 + \eta_h) \sigma_u^2 \quad h = 1, 2, \dots, z \quad (3.7)$$

where

$$0 < \eta_1 < \eta_2 < \dots < \eta_z$$

2. At the h th level of variance of measurement errors, $h = 1, 2, \dots, Z$, a large number, b , of independent error (pseudo error) vectors

$$[U_{hl}] \sim N(0, (1 + \eta_h) \sigma_u^2) \quad l = 1, 2, \dots, b \quad (3.8)$$

are generated and added to the observed w 's to obtain a new data sets.

3. For the l th new data set, the least squares estimator $\hat{\beta}_{jhl}$ and its standard errors $\hat{\sigma}_{\hat{\beta}_{jhl}}$; $l = 1, 2, \dots, b$; $j = 0, 1, 2$ are computed.

4. Calculating the sample means of the estimated values $\hat{\beta}_{jhl}$ and their standard errors $\hat{\sigma}_{\hat{\beta}_{jhl}}$ as follows;

$$\hat{\beta}_{jh} = \frac{1}{b} \sum_{l=1}^b \hat{\beta}_{jhl} \quad \text{and} \quad \hat{\sigma}_{\hat{\beta}_{jh}} = \frac{1}{b} \sum_{l=1}^b \hat{\sigma}_{\hat{\beta}_{jhl}} \quad j = 0, 1, 2$$

5. Repeating the steps (2) - (4) for all values of η_h ; $h = 1, 2, \dots, z$.
6. Computing the sample mean of the estimated standard errors $\hat{\sigma}_{\hat{\beta}_{jh}}$ as follows;

These estimators define the constraints under which the likelihood function is for the unknown $x_1, x_2, \dots, x_n$. Given that $U \sim N(0, \sigma_u^2)$ and only $w_1, w_2, \dots, w_n$ are observed, the likelihood for the unknown $x_1, x_2, \dots, x_n$ is

$$f(w/x) = \prod_{i=1}^n \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{(w_i - x_i)^2}{2\sigma_u^2}}$$

Since the negative log likelihood is proportional to $\sum_{i=1}^n (w_i - x_i)^2$, then the objective function, to minimize this sum of squares subject to the moment and cross-product constraints, is

$$q = \frac{1}{2} \sum_{i=1}^n (w_i - x_i)^2 + \sum_{r=1}^4 \frac{n}{r} \lambda_r \left[\frac{1}{n} \sum_{i=1}^n x_i^r - m_r \right] + \sum_{r=1}^2 \frac{n}{r} \lambda_{4+r} \left[\frac{1}{n} \sum_{i=1}^n x_i^r y_i - m_{r1} \right] \quad (3.5)$$

where $\lambda_1, \lambda_2, \dots, \lambda_6$ are Lagrange multipliers, and the constants are included to simplify derivatives.

Differentiating q with respect to $x_i, \lambda_1, \lambda_2, \dots, \lambda_6$ and setting the first partial derivative equal to zero gives

$$x_i - w_i + \sum_{r=1}^4 \lambda_r x_i^{r-1} + \sum_{r=1}^2 \lambda_{4+r} x_i^{r-1} y_i = 0 \quad i = 1, 2, \dots, n \quad (3.6)$$

and putting the other partial derivatives in a (6×6) diagonal matrix Λ ,

where the diagonal elements are $n \left[\frac{1}{n} \sum_{i=1}^n x_i - m_1 \right], \frac{n}{2} \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - m_2 \right],$

$\frac{n}{3} \left[\frac{1}{n} \sum_{i=1}^n x_i^3 - m_3 \right], \frac{n}{4} \left[\frac{1}{n} \sum_{i=1}^n x_i^4 - m_4 \right], n \left[\frac{1}{n} \sum_{i=1}^n x_i y_i - m_{11} \right]$ and

$\frac{n}{2} \left[\frac{1}{n} \sum_{i=1}^n x_i^2 y_i - m_{21} \right]$ respectively.

Then the steps of this method are:

1. Choosing initial values of $\lambda_r, r = 1, 2, \dots, 6$.
2. Solving equation (3.6) to estimate the value of x_i .
3. Using Newton-Raphson method to estimate the new values of λ 's.
4. Applying the OLS method to estimate the parameters β 's and their standard errors.

3.2 Adjusted Data (AD) Method

Adjusted data method was presented by Bay and Stefanski (2000) to estimate the parameters of a quadratic regression model. The naïve least squares estimators follow from solving the least squares equations:

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^n w_i + \beta_2 \sum_{i=1}^n w_i^2 &= \sum_{i=1}^n y_i \\ \beta_0 \sum_{i=1}^n w_i + \beta_1 \sum_{i=1}^n w_i^2 + \beta_2 \sum_{i=1}^n w_i^3 &= \sum_{i=1}^n w_i y_i \\ \beta_0 \sum_{i=1}^n w_i^2 + \beta_1 \sum_{i=1}^n w_i^3 + \beta_2 \sum_{i=1}^n w_i^4 &= \sum_{i=1}^n w_i^2 y_i \end{aligned}$$

Though $E(w)$ and $E(wy)$ are unbiased estimators of $E(x)$ and $E(xy)$ respectively, the remaining estimators are all biased since

$$\begin{aligned} E(x^2) &= E(w^2) - \sigma_u^2, \\ E(x^3) &= E(w^3) - 3\sigma_u^2 E(w), \\ E(x^4) &= E(w^4) - 6\sigma_u^2 E(w^2) + 5\sigma_u^4 \text{ and} \\ E(x^2 y) &= E(w^2 y) - \sigma_u^2 E(y) \end{aligned}$$

Thus, the estimated parameters will also be biased. To reduce this bias the observed data $w_1, w_2, \dots, w_n$ are transformed to $x_1^*, x_2^*, \dots, x_n^*$ such that

$$m_r = E(x^{*r}) = E(x^r) \quad r = 1, 2, 3, 4$$

and

$$m_{r1} = E(x^{*r} y) = E(x^r y) \quad r = 1, 2$$

The estimators m_r and m_{r1} are found by using Hermite Polynomials (see Cramér (1961) and Stefanski (1989)) as follows;

$$\begin{aligned} m_1 &= E(x) = E(w) \\ m_2 &= E(x^2) - 1 = E(w^2) - \sigma_u^2 - 1 \\ m_3 &= E(x^3) - 3E(x) = E(w^3) - 3(\sigma_u^2 + 1)E(w) \\ m_4 &= E(x^4) - 6E(x^2) + 3 = E(w^4) - 6(\sigma_u^2 + 1)E(w^2) + \sigma_u^2(5\sigma_u^2 + 6) + 3 \\ m_{11} &= E(xy) = E(wy) \\ m_{21} &= E(x^2 y) = E(w^2 y) - (\sigma_u^2 + 1)E(y) \end{aligned}$$

$$= \text{Var}(x_{im})$$

And

$$\text{Var}(x_{im}) = E(x_{im}^2) - [E(x_{im})]^2 = E(w_{im}^2) - \sigma_u^2 - [E(w_{im})]^2 = \sigma_{w_m}^2 - \sigma_u^2$$

If $\bar{x}_m$ is estimated by $\bar{w}_m$, then equation (3.1) becomes

$$\begin{aligned} E(x_{im} / w_{im}) &= \bar{x}_m - \frac{\sigma_{w_m}^2 - \sigma_u^2}{\sigma_{w_m}^2} \bar{w}_m + \frac{\sigma_{w_m}^2 - \sigma_u^2}{\sigma_{w_m}^2} w_{im} \\ &= \bar{w}_m + \frac{\sigma_{w_m}^2 - \sigma_u^2}{\sigma_{w_m}^2} (w_{im} - \bar{w}_m) \quad i=1, 2, \dots, n \end{aligned} \quad (3.4)$$

where

$$\bar{w}_m = \frac{1}{n} \sum_{i=1}^n w_{im} \quad \text{and} \quad \sigma_{w_m}^2 = \frac{1}{n-1} \sum_{i=1}^n (w_{im} - \bar{w}_m)^2$$

2. The values of y_i and $E(x_{im} / w_{im})$ are used to estimate the parameters of the polynomial regression model (1.1), using the ordinary least squares method, when the model (1.3) is applied.

The modification of this method is sketched as follows;

- Applying the steps (1) and (2) at $m = 1, 2, \dots, 5$, to compute $\hat{\beta}_{jm}$ and $\hat{\sigma}_{\hat{\beta}_{jm}}$; $j = 0, 1, 2$.

A simulation study was performed to show the effect of the values of m on the biases of the corrected estimators of the parameters of the quadratic regression model. Table (1) in the appendix shows some results of this study (section 5 presents the details of this study). From table(1), we find that; when $n = 200$, $\sigma_e^2 = 0.25$ and $\sigma_u^2 = 0.25$ the corrected estimators have minimum biases at $m = 2$; when $n = 400$, $\sigma_e^2 = 0.25$ and $\sigma_u^2 = 0.25$ the corrected estimators have minimum biases at $m = 5$, and so on. So, the optimal value of m differs from case to case.

- Calculating the corrected estimators $\hat{\beta}_j$'s and their corresponding standard errors as follows;

$$\hat{\beta}_j = \frac{1}{5} \sum_{m=1}^5 \hat{\beta}_{jm} \quad \text{and} \quad \hat{\sigma}_{\hat{\beta}_j} = \frac{1}{5} \sum_{m=1}^5 \hat{\sigma}_{\hat{\beta}_{jm}} \quad j = 0, 1, 2$$

3. Alternative Estimators of the Parameters (II)

This section presents modified regression calibration, adjusted data, and simulation-extrapolation estimators, as alternative estimators to the biased and inconsistent least squares estimators, of the parameters of the polynomial regression model, when the models (1.1) and (1.3) are applied.

3.1 The Modified Regression Calibration (MRC) Method

From equation (1.3) we find that the relation between x and w is linear. So, we can regress x on w (see Carroll and Stefanski (1990)) to obtain $\hat{x}_i = \hat{a}_0 + \hat{a}_1 w_i$, using m replicate measurements of w_i , where

$$w_i = \frac{1}{m} \sum_{h=1}^m w_{ih}$$

Carroll and Stefanski (1990) suggested using a suitable number of m , $m = 4$ or 5 . The resultant equation is used to calculate the estimated value of x_i ; $i = 1, 2, 3, \dots, n$, then using the estimated values of x 's and values of y 's to estimate the parameters of the model (1.1) when the model (1.3) is applied. At a given value of m , this method is performed as follows;

1. The least squares method is used to estimate the parameters a_0 and a_1 . The m th estimated value of x_i ; $i = 1, 2, 3, \dots, n$ is

$$E(x_{im} / w_{im}) = \hat{a}_{0m} + \hat{a}_{1m} w_{im} \quad i = 1, 2, \dots, n \quad (3.1)$$

where

$$w_{im} = \frac{1}{m} \sum_{h=1}^m w_{imh}$$

And the m th ordinary least squares estimators of a_0 and a_1 are given by

$$\hat{a}_{0m} = \bar{x}_m - \hat{a}_{1m} \bar{w}_m \quad (3.2)$$

$$\hat{a}_{1m} = \frac{\text{Cov}(x_{im}, w_{im})}{\text{Var}(w_{im})} = \frac{\text{Var}(x_{im})}{\text{Var}(w_{im})} = \frac{\sigma_{w_m}^2 - \sigma_u^2}{\sigma_{w_m}^2} \quad (3.3)$$

Since

$$\text{Cov}(x_{im}, w_{im}) = \text{Cov}(x_{im}, [x_{im} + u_{im}]) = \text{Cov}(x_{im}, x_{im}) + \text{Cov}(x_{im}, u_{im})$$

$$\underline{c} = \begin{bmatrix} \sigma_u^2 \beta_2 \\ \sigma_u^2 \beta_1 + 3\sigma_u^2 \beta_2 \sum_{i=1}^n w_i \\ 6\sigma_u^2 \beta_2 \sum_{i=1}^n w_i^2 - 5n\sigma_u^4 \beta_2 - \sigma_u^2 \sum_{i=1}^n y_i + \sigma_u^2 \beta_0 + 3\sigma_u^2 \beta_1 \sum_{i=1}^n w_i \end{bmatrix}$$

And if the theory of M-estimators (see Huber 1981) is applied to $U^*(\underline{\beta}, \sigma_e^2, w, \underline{y})$, then the unconditional distribution of $\underline{\hat{\beta}}_{csf}$ is asymptotically normal with mean $\underline{\beta}$ and variance

$$\text{var}(\underline{\hat{\beta}}_{csf}) = I(\underline{\beta}, x, \underline{y})^{-1} \cdot C(\underline{\beta}, x, \underline{y}) \cdot I(\underline{\beta}, x, \underline{y})^{-1} \quad (2.25)$$

where

$I(\underline{\beta}, x, \underline{y})$ is the Fisher information matrix, and it is estimated by corrected observed Fisher information $I^*(\underline{\beta}, w, \underline{y})$ which is equal to $\hat{\sigma}_e^{-2} W$.

$C(\underline{\beta}, x, \underline{y})$ is the unconditional variance matrix of $U^*(\underline{\beta}, \sigma_e^2, w, \underline{y})$, and it is estimated as follows;

$$\hat{C}(\underline{\beta}, x, \underline{y}) = \frac{1}{n} \sum_{i=1}^n \left[U(\underline{\hat{\beta}}_{csf}, \hat{\sigma}_e^2, w, \underline{y}) \cdot U(\underline{\hat{\beta}}_{csf}, \hat{\sigma}_e^2, w, \underline{y})' \right] \quad (2.26)$$

Then

$$\text{Var}(\underline{\hat{\beta}}_{csf}) = (\hat{\sigma}_e^{-2} W)^{-1} \cdot \hat{C}(\underline{\beta}, x, \underline{y}) \cdot (\hat{\sigma}_e^{-2} W)^{-1} \quad (2.27)$$

Solving the equation $\frac{\partial}{\partial \sigma_e^2} L^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y)$ we obtain an estimate of σ_e^2

$$\hat{\sigma}_e^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=0}^2 \hat{\beta}_j w_i^j \right)^2 - \sigma_u^2 \hat{\beta}_1^2 - \sigma_u^2 \hat{\beta}_2 \left(6\hat{\beta}_2 \sum_{i=1}^n w_i^2 - 5n\sigma_u^2 \hat{\beta}_2 - 2 \sum_{i=1}^n y_i + 2\hat{\beta}_0 + 6\hat{\beta}_1 \sum_{i=1}^n w_i \right) \quad (2.28)$$

By the same way, we can obtain the corrected estimators of the parameters of the polynomial regression model of degree $k \geq 3$ when the explanatory variable is measured with error.

The estimating equations for the corrected score function estimators are constructed by setting these partial derivatives equal to zero and putting the resultant equations in matrix notation

$$\therefore W \hat{\underline{\beta}}_{csf} = \underline{Y} \quad (2.23)$$

where

$$W = \begin{bmatrix} n & \sum_{i=1}^n w_i & \sum_{i=1}^n w_i^2 - n\sigma_u^2 \\ \sum_{i=1}^n w_i & \sum_{i=1}^n w_i^2 - n\sigma_u^2 & \sum_{i=1}^n w_i^3 - 3n\sigma_u^2 \sum_{i=1}^n w_i \\ \sum_{i=1}^n w_i^2 - n\sigma_u^2 & \sum_{i=1}^n w_i^3 - 3n\sigma_u^2 \sum_{i=1}^n w_i & \sum_{i=1}^n w_i^4 - 6n\sigma_u^2 \sum_{i=1}^n w_i^2 + 5n^2 \sigma_u^4 \end{bmatrix},$$

$$\underline{Y} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n w_i y_i \\ \sum_{i=1}^n w_i^2 y_i - n\sigma_u^2 \sum_{i=1}^n y_i \end{bmatrix}$$

Then

$$\hat{\underline{\beta}}_{csf} = W^{-1} \underline{Y} \quad (2.24)$$

In fact, $U_j^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y); j = 0, 1, 2$ may be rewritten, in matrix notation, as follows;

$$\begin{aligned} U^*(\underline{\beta}, \sigma_e^2, w, \underline{y}) &= \frac{1}{\sigma_e^2} [w' \underline{y} - w' w \underline{\beta}] + \frac{n}{\sigma_e^2} \underline{c} \\ &= U(\underline{\beta}, \sigma_e^2, w, \underline{y}) + \frac{n}{\sigma_e^2} \underline{c} \end{aligned}$$

where

$$w' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ w_1 & w_2 & w_3 & \dots & w_n \\ w_1^2 & w_2^2 & w_3^2 & \dots & w_n^2 \end{bmatrix},$$

where

$$\hat{V} = \frac{1}{n} \sum_{i=1}^n \left[\underline{H}_i - D_i \hat{\underline{\beta}}_{cls} \right] \left[\underline{H}_i - D_i \hat{\underline{\beta}}_{cls} \right]' \quad (2.18)$$

2.2 Extension of the Corrected Score Function (CSF) Method

Corrected Score function Method was presented by Nakamura (1990) for linear regression model. In this subsection an extension of this method is derived for the polynomial regression model. The log likelihood function of the polynomial regression model of second degree, of unknown $x_1, x_2, x_3, \dots, x_n$ and $\beta_0, \beta_1, \beta_2$ is

$$\log L(\beta_0, \beta_1, \beta_2, \sigma_e^2, x, y) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^n \left(y_i - \sum_{j=0}^2 \beta_j x_i^j \right)^2 \quad (2.19)$$

The corrected score function method depends on finding the corrected log function $L^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y)$, which depends on w not x , such that:

$$E[L^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y)] = \log L(\beta_0, \beta_1, \beta_2, \sigma_e^2, x, y) \quad (2.20)$$

If we assume that

$$\text{Cov}(u_i, u_j) = 0 \quad \forall i \neq j \text{ and } i, j = 1, 2, 3, \dots, n \quad (2.21)$$

Then, under assumptions (1.4)-(1.7), (2.21) and using equations (2.4)-(2.8) and (2.10)-(2.12), we find

$$\begin{aligned} L^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y) = & -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^n \left[\left(y_i - \sum_{j=0}^2 \beta_j w_i^j \right)^2 \right] \\ & - \frac{1}{2\sigma_e^2} \left[-n\sigma_u^2 \beta_1^2 - 6n\sigma_u^2 \beta_2^2 \sum_{i=1}^n w_i^2 + 5n^2 \sigma_u^4 \beta_2^2 \right] \\ & - \frac{1}{2\sigma_e^2} \left[2n\sigma_u^2 \beta_2 \sum_{i=1}^n y_i - 2n\sigma_u^2 \beta_0 \beta_2 - 6n\sigma_u^2 \beta_1 \beta_2 \sum_{i=1}^n w_i \right] \end{aligned} \quad (2.22)$$

Then finding the corrected score function $U^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y)$ by differentiating L^* with respect to β_j ; $j = 0, 1, 2$, i. e.

$$U_j^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y) = \frac{\partial}{\partial \beta_j} L^*(\beta_0, \beta_1, \beta_2, \sigma_e^2, w, y) \quad j = 0, 1, 2$$

$$h_{21} = \sum_{i=1}^n x_i^2 y_i = \sum_{i=1}^n w_i^2 y_i - n \sigma_u^2 \sum_{i=1}^n y_i$$

and so on. Then, obviously, $\underline{H} = X' \underline{y}$. The estimating equations for the corrected least squares estimators are constructed by replacing $X'X$ and $X' \underline{y}$ by D and $\underline{H}$ in equations (2.1) respectively to get

$$D \hat{\underline{\beta}}_{cls} = \underline{H} \quad (2.13)$$

$$\therefore \hat{\underline{\beta}}_{cls} = D^{-1} \underline{H} \quad (2.14)$$

The OLS estimator of σ_e^2 in model (1.2) using x is known to be (see Johnston (1972))

$$\hat{\sigma}_{e,ols}^2 = \frac{1}{n} \left[\underline{y}' \underline{y} - \hat{\underline{\beta}}_{ols}' X' \underline{y} \right] \quad (2.15)$$

Replacing $X' \underline{y}$ and $\hat{\underline{\beta}}_{ols}$ by $\underline{H}$ and $\hat{\underline{\beta}}_{cls}$ respectively, in equation (2.15), we obtain the corrected least squares estimator of σ_e^2

$$\begin{aligned} \hat{\sigma}_{e,cls}^2 &= \frac{1}{n} \left[\underline{y}' \underline{y} - \hat{\underline{\beta}}_{cls}' \underline{H} \right] \\ &= \frac{1}{n} \left[\underline{y}' \underline{y} - \underline{H}' D^{-1} \underline{H} \right] \end{aligned} \quad (2.16)$$

Since (see Johnston (1972))

$$\hat{\underline{\beta}}_{ols} - \underline{\beta} = (X'X)^{-1} X' \underline{e}$$

Thus

$$\begin{aligned} Var(\hat{\underline{\beta}}_{ols}) &= E[(X'X)^{-1} X' \underline{e}][X' \underline{e}] \\ &= (X'X)^{-1} E[X' \underline{e} \underline{e}' X] (X'X)^{-1} \end{aligned}$$

Since

$$\begin{aligned} X' \underline{e} &= X' (\underline{y} - X \underline{\beta}) = X' \underline{y} - X' X \underline{\beta} \\ \therefore E(X' \underline{e} \underline{e}' X) &= E(X' \underline{y} - X' X \underline{\beta})(X' \underline{y} - X' X \underline{\beta})' \\ &= E(\underline{H} - D \underline{\beta}_{cls})(\underline{H} - D \underline{\beta}_{cls})' \\ &= \frac{1}{n} \sum_{i=1}^n [\underline{H}_i - D_i \underline{\beta}_{cls}][\underline{H}_i - D_i \underline{\beta}_{cls}]' \end{aligned}$$

Then the variance of $\hat{\underline{\beta}}_{cls}$ is

$$Var(\hat{\underline{\beta}}_{cls}) = D^{-1} \hat{V} D^{-1} \quad (2.17)$$

$$d_{00} = \sum_{i=1}^n x_i^0 = \sum_{i=1}^n w_i^0 = n$$

$$d_{01} = d_{10} = \sum_{i=1}^n x_i = \sum_{i=1}^n w_i$$

$$d_{02} = d_{20} = d_{11} = \sum_{i=1}^n x_i^2 = \sum_{i=1}^n w_i^2 - n\sigma_u^2$$

$$d_{03} = d_{30} = d_{12} = d_{21} = \sum_{i=1}^n x_i^3 = \sum_{i=1}^n w_i^3 - 3n\sigma_u^2 \sum_{i=1}^n w_i - \sum_{i=1}^n u_i^3$$

$$d_{04} = d_{40} = d_{13} = d_{31} = d_{22} = \sum_{i=1}^n x_i^4 =$$

$$\sum_{i=1}^n w_i^4 - 6n\sigma_u^2 \sum_{i=1}^n w_i^2 - 4 \sum_{i=1}^n w_i \sum_{i=1}^n u_i^3 - \sum_{i=1}^n u_i^4 + 6n^2\sigma_u^4$$

and so on. Then, obviously, $D = X'X$.

Similarly,

$$\begin{aligned} \sum_{i=1}^n w_i^r y_i &= \sum_{i=1}^n (x_i + u_i)^r y_i \quad r=0, 1, 2, \dots, k \\ &= \sum_{i=1}^n \left[\sum_{l=0}^r \binom{r}{l} x_i^l \cdot u_i^{r-l} \right] y_i \quad r=0, 1, 2, \dots, k \end{aligned} \quad (2.9)$$

Using equation (2.9) and assumptions (1.4)-(1.7), we get

$$\sum_{i=1}^n w_i^0 y_i = \sum_{i=1}^n x_i^0 y_i = \sum_{i=1}^n y_i \quad (2.10)$$

$$\sum_{i=1}^n w_i y_i = \sum_{i=1}^n x_i y_i \quad (2.11)$$

$$\sum_{i=1}^n w_i^2 y_i = \sum_{i=1}^n x_i^2 y_i + n\sigma_u^2 \sum_{i=1}^n y_i \quad (2.12)$$

and so on. Let us consider the $(k+1) \times 1$ vector $\underline{H}$, where the element $(p, 1)$ is given by h_{p1} ; $p = 0, 1, 2, \dots, k$. The element h_{p1} is calculated

from the observable w_i such that $h_{p1} = \sum_{i=1}^n x_i^p y_i$. Solving equations (2.10), (2.11), (2.12), ... one by one for $x_i^p y_i$ we find that

$$h_{01} = \sum_{i=1}^n x_i^0 y_i = \sum_{i=1}^n w_i^0 y_i = \sum_{i=1}^n y_i$$

$$h_{11} = \sum_{i=1}^n x_i y_i = \sum_{i=1}^n w_i y_i$$

$$(X' X) \hat{\underline{\beta}}_{ols} = X' \underline{y} \quad (2.1)$$

where

$$X' X = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^k \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \dots & \sum_{i=1}^n x_i^{k+1} \\ . & . & . & . & . \\ \sum_{i=1}^n x_i^k & \sum_{i=1}^n x_i^{k+1} & \sum_{i=1}^n x_i^{k+2} & \dots & \sum_{i=1}^n x_i^{2k} \end{bmatrix} \text{ and } X' \underline{y} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ . \\ . \\ \sum_{i=1}^n x_i^k y_i \end{bmatrix} \quad (2.2)$$

From Equation (1.3), we find that

$$\begin{aligned} \sum_{i=1}^n w_i^r &= \sum_{i=1}^n (x_i + u_i)^r & r = 0, 1, 2, \dots, 2k \\ &= \sum_{i=1}^n \left[\sum_{l=0}^r \binom{r}{l} x_i^l \cdot u_i^{r-l} \right] & r = 0, 1, 2, \dots, 2k \end{aligned} \quad (2.3)$$

Using equation (2.3) and assumptions (1.4)-(1.7), we get

$$\sum_{i=1}^n w_i^0 = \sum_{i=1}^n x_i^0 = n \quad (2.4)$$

$$\sum_{i=1}^n w_i = \sum_{i=1}^n x_i \quad (2.5)$$

$$\sum_{i=1}^n w_i^2 = \sum_{i=1}^n x_i^2 + n\sigma_u^2 \quad (2.6)$$

$$\sum_{i=1}^n w_i^3 = \sum_{i=1}^n x_i^3 + 3n\sigma_u^2 \sum_{i=1}^n x_i + \sum_{i=1}^n u_i^3 \quad (2.7)$$

$$\sum_{i=1}^n w_i^4 = \sum_{i=1}^n x_i^4 + 6n\sigma_u^2 \sum_{i=1}^n x_i^2 + 4 \sum_{i=1}^n x_i \sum_{i=1}^n u_i^3 + \sum_{i=1}^n u_i^4 \quad (2.8)$$

and so on. Let us consider the $(k+1) \times (k+1)$ matrix D , where the element (p, q) is given by d_{pq} ; $p, q = 0, 1, 2, \dots, k$ (see Cheng and Schneeweiss (1998)). The element d_{pq} is computed from the observable w_i such that $d_{pq} = \sum_{i=1}^n x_i^{p+q}$. Solving equations (2.4), (2.5), (2.6), ... one by one for x_i^{p+q} we find that

simple linear regression model, when the explanatory variable is measured with error. She recommended to use estimators of the corrected score function and simulation-extrapolation methods as alternative estimators to the least squares estimator (LSE) in application.

Whereas the linear regression model has found extensive treatment in the literature, the polynomial regression model has received only a little attention despite the fact that it is perhaps the most natural extension of the linear regression model. The present research is directed towards introducing corrected estimators, of some correction methods, of the parameters of the polynomial regression model, when the explanatory variable is measured with error, and comparing between these corrected estimators of the parameters of the quadratic regression model when the model (1.3) is applied. So the remainder of the present research is organized as follows: Section 2 presents the corrected least square method and an extension of the corrected score function method to calculate the corrected estimators of the parameters of the polynomial regression model (1.1) assuming that the model (1.3) is applied. Section 3 presents a modification of the regression calibration method, adjusted data and simulation-extrapolation methods to compute the corrected estimators of the parameters, assuming the models (1.1) and (1.3) are applied. In section 4 the correction methods are applied to real data set in the medical field. The simulated data sets are used to compare between the corrected estimators, for these correction methods, in section 5. Conclusions are drawn in section 6.

2. Alternative Corrected Estimators of the Parameters (I)

This section presents the corrected least squares and corrected score function estimators, as alternatives to the biased and inconsistent least squares estimators, of the parameters of the polynomial regression model, when the models (1.1) and (1.3) are applied.

2.1 The corrected Least Squares (CLS) Method

When the true values of $x_i; i = 1, 2, 3, \dots, n$ are known, and the OLS method is used to estimate $\underline{\beta}$, the estimating equations are

The ordinary polynomial regression model is

$$y_i = \sum_{j=0}^k \beta_j x_i^j + e_i \quad i = 1, 2, 3, \dots, n \quad (1.1)$$

In matrix notation

$$\underline{y} = X \underline{\beta} + \underline{e} \quad (1.2)$$

where

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix}, \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ and } \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and e_i 's are independent $N(0, \sigma_e^2)$ and uncorrelated with the x_i 's. The model (1.1) assumes that $x_i \sim N(\mu_x, \sigma_x^2)$ and independent of e_i , and x_i is the true value, i. e. x_i is measured without error. Assume that the true value x_i can not be observed precisely or directly, instead, the imprecise w_i is observed, where

$$w_i = x_i + u_i \quad i = 1, 2, 3, \dots, n \quad (1.3)$$

where

$$E(u_i) = 0 \text{ and } E(u_i^2) = \sigma_u^2 \quad i = 1, 2, 3, \dots, n \quad (1.4)$$

and

$$\text{Cov}(x_i, u_i) = 0 \quad i = 1, 2, 3, \dots, n \quad (1.5)$$

$$\text{Cov}(y_i, u_i) = 0 \quad i = 1, 2, 3, \dots, n \quad (1.6)$$

$$\text{Cov}(e_i, u_i) = 0 \quad i = 1, 2, 3, \dots, n \quad (1.7)$$

If m replicates are available, for each unit i , then the estimated value of σ_u^2 would be (see Carroll et al. (1995))

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \sum_{h=1}^m (w_{ih} - \bar{w}_i)^2}{n(m-1)} \quad (1.8)$$

Ibrahim (2005) compared the estimators of the maximum likelihood, corrected score, simulation-extrapolation, semiparametric likelihood, regression calibration and grouping methods of the slope in

The problem of estimating the parameters in the regression model of a response variable y on an explanatory variable x from observations on (y, w) , where w is the measurement of x , is a case of what has historically been called the errors-in-variables problem. It is well known that the presence of this error in the explanatory variable makes the ordinary least squares (OLS) estimators inconsistent and biased.

There are many suggested methods to estimate the parameters in this case, see Cochran (1968), Koutsoyiannis (1973), Feldstein (1974), Kendall and Stuart (1979) Fuller (1980), Anderson (1984), Hwang (1986), Schafer (1986), Hill (1992), Jang and Eltinge (1997), and Schafer (2001) for the linear model case; Armstrong (1985), Stefanski (1985), Stefanski and Carroll (1987), Whittemore and Keller (1988), Nakamura (1990), and Hutchison et al. (2000) for the generalized linear model case; Walter and Fuller (1982a,b), Chan and Mak (1985), Stefanski (1989), Cheng and Schneeweiss (1998), Bay and Stefanski (2000) for the nonlinear model case; Stefanski and Carroll (1985) for the logistic model; Cook and Stefanski (1994) for linear, nonlinear and logistic models; Prentice (1982) for the Cox regression model case.

Eltinge et al. (1994) presented some explanatory analyses of differences between income and expenditure values reported by single person consumer unites in the U. S. Consumer Expenditure Interview Survey, when the reported income is incorrect.

Biemer and Atkinson (1995) proposed a general estimator for estimating the measurement error bias and its variance for some characteristics such as population mean using two-phase sampling design.

Olson and Kennedy (2006) studied the relation between nonresponse error and measurement error using data from a survey of University of Maryland Alumni. They found that nonresponse bias exists and that measurement error bias exists, but that the nexus of the two is not as clear cut.

Peytchev (2006) presented two-stage model for estimating measurement error. This model is shown to provide the ability to obtain estimates for causes of measurement error bias, such as differences between interviewers or survey protocols, and estimates for how these and other causes affect measurement error variance.

On the Estimation Problem of Parameters of the Polynomial Regression Model When the Explanatory Variable is Measured with Error

Farid Naguib Guirguis*

ABSTRACT

For the polynomial regression model in which the explanatory variable contains measurement error, the correction methods for attenuation estimators of the parameters are considered. The corrected least squares method, an extension of the corrected score function method and a modification of regression calibration method are derived to obtain the corrected estimators of the model parameters. The adjusted data and simulation-extrapolation methods are also presented for the same purpose. A comparative study between the resulted estimators in the case of the quadratic regression model is performed using real medical data as well as a simulation study. The results reveal the superiority of the corrected estimators of adjusted data and modified regression calibration methods over the corrected estimators of the other methods.

1. Introduction

Theory of regression models assumes that the explanatory variables are all measured without error. Unfortunately, most of the actual data are rarely of this quality. The field of errors of measurement is of growing importance since the recent rapid in data-gathering projects in political, economics, social and medical sciences are producing a lot of data containing variables which are obviously difficult to be measure correctly, such as:

- Personal abilities like intelligence score and language ability.
- People's behaviors like opinions, feelings and motivations.
- Metabolic measurements like blood pressure, temperature, pulse and glucose.

The statistical consideration of regression models containing measurement errors began as early as 1877 (see Fuller (1987)).

* Benha University, Faculty of Commerce, Department of Statistics

- [8] Schroeder, J., and Howard, D., (1998); " Multiscale Modeling for Face Detection in Complex Synthetic Aperture Computerized Imagery for Manmade Object Detection " Asilomar'98, Monterey, CA, Nov'98.
- [9] Schroeder, J., (2001); " Automatic Face Detection and Recognition Using Synthetic Aperture Computerized Imagery " Cooperative Research Center for Sensor Signal and Information Processing. SPRI Building, Mawson Lakes Boulevard Mawson Lakes.
- [10] Schroeder, J., and Suvorova, S., (2002); " Automatic Face Classification and Recognition Using a Rotationally Invariant KLT " Special Issue on Defence Applications of Signal Procession, DSP.
- [11] Souror.H.A, (2009); "Using Multivariate Analysis Techniques for Optimization of the Artificial Neural Networks Learning Efficiency", master thesis, Statistics Dep, Benha University.
- [12] Turk, M., and Pentland, A., (1991); "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol. 3, Issue 1, pp.71-86.

Abbreviations

ANN: Artificial neural networks
DA: Discriminate analysis
LDA: Linear discriminate analysis
KLT: Karhunen-Loève transform
KNM: k-nearest mean
kNN: k-nearest neighbor technique
MSE: Mean square error
PCA: Principle components analysis
SSE: Sum square error

In the experiment, we designed a traditional neural network without any modification and designed a modified test neural network based on a previous formula. We found, after calculating results, a new approach which is better than classical approach of increasing neural network learning efficiency by efficiency ratio 97.8%, approximately 5.8 times as of classical approach as well as the time of classical approach 11.3 times the time of the new approach for each image separately. Accordingly, we achieved the main objective of research; it is higher efficiency and lesser time.

Also this technique can increase the possibility of recognizing miss-detections and miss-classification of faces. This technique of face recognition in computerized images opens the way widely for informational and operational control for command centers for all levels, before, during and after operations for collateral continuously events analysis consequently make best decisions. It can use this technique can be used in drugs contention, construction planning and border surveillance. Also this technique can use in iris recognition, detecting petroleum oil wells, water artesian wells, detecting secret tunnels, control underground trains, detecting seismic waves and detecting changes in plate tectonics.

Ψ Where and when it was so decided that the multivariate analysis techniques has already contributed What does not produce much to doubt in increasing the efficiency of the artificial neural networks.

7 References

- [1] DARPA, (1988); " Neural Network Study " Lexington, MA: M.I.T. Lincoln Laboratory AFCEA International Press
- [2] Cooke, T., and Peake, M., (2002); " The Optimal Classification Using a Linear Discriminant for two Point Classes having Known Mean and Covariance " Accepted in Journal of Multivariate Analysis, Vol. 82, Issue 2, pp. 379-394.
- [3] Cooke, T., Redding, N., Schroeder, J., and Zhang, J., (2000); " A Comparison of Selected Features for Face Detection in Synthetic Aperture Computerized Imagery " Special Issue on Defence Applications of Signal Processing, DSP, Vol. 10, Issue 4, pp. 286-296.
- [4] Diamantaras, K., and Kung, S., (1996); " Principal Component Neural Networks: Theory and Applications " John Wiley & Sons.
- [5] Johnson, R., and Wishern, D., (1997); " Applied Multivariate Statistical Analysis " 4th Edition, Prentice Hall, New Jersey, Mardia, Kent, and Bibby.
- [6] Kirby, M., and Sirovich, L., (1990); " Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces " Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 12, Issue 1, pp. 103-108.
- [7] Marcialis, G., and Roli, F., (2002), "Fusion of LDA and PCA for Face Recognition", Department of Electrical and Electronic Engineering, University of Cagliari, Italy, In Proc. of the Workshop on Machine Vision and Perception. 8th Meeting of the Italian Association of Artificial Intelligence (AI*IA), Siena (Italy), September pp10-13.

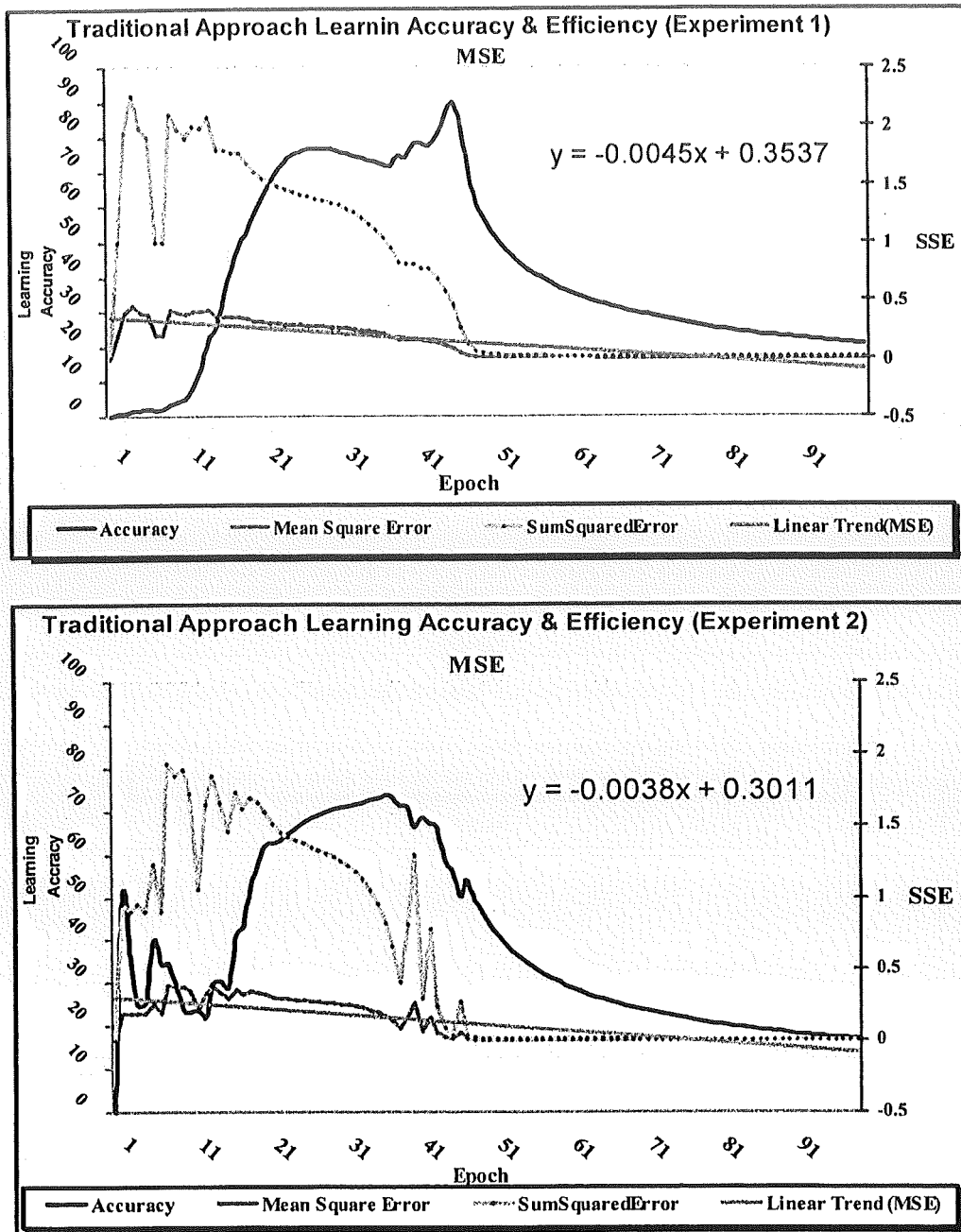
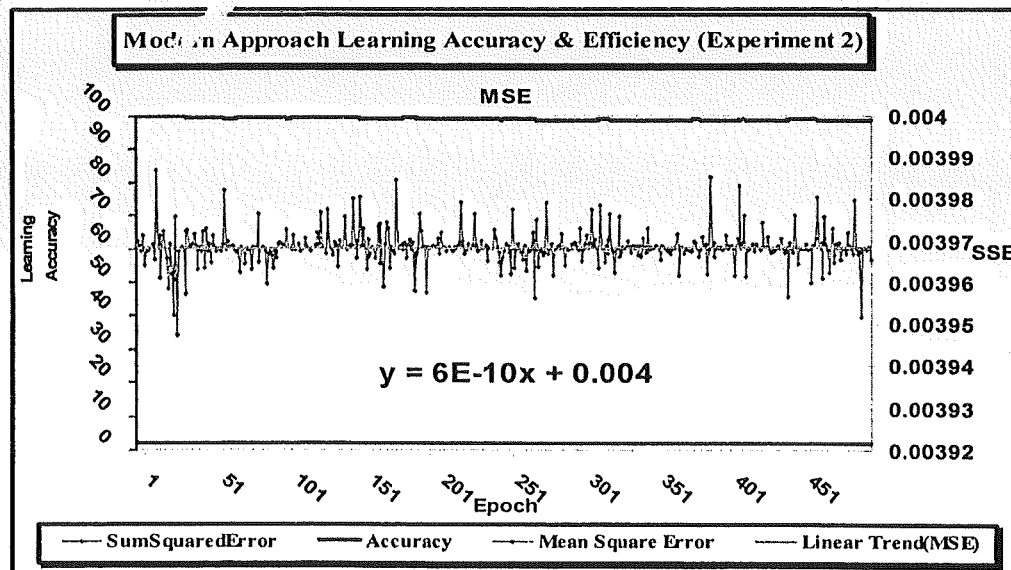
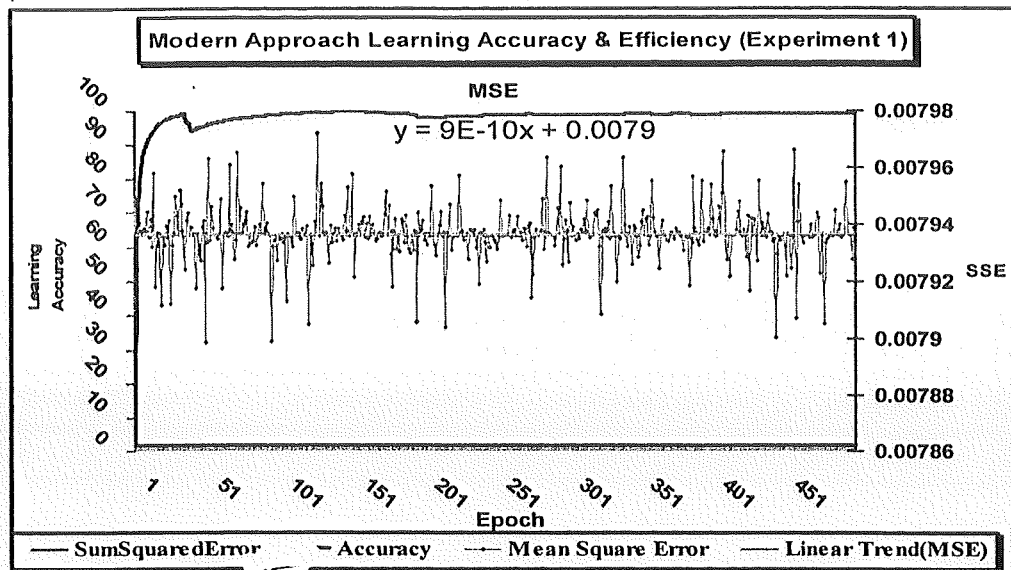


Figure 5 Experiments Results Graphs

6 Conclusions

In this paper, we proposed a new algorithm for face recognition based on technique of neural networks modified learning dataset by (Integration between PCA and LDA) which is developed. This technique consists of two procedures: (I) automatic faces detection (II) automatic faces recognition and identification by this technique.

completion of the learning experiment by classical approach under go 1.8 sec ,consequently the new approach excel at a time also, where the time classical approach 11.3 times the time of the new approach for each image separately.



$MSE = \sum [(OUT - \text{Actual results})^2]$ - MSE is measure of accuracy Mean square error (MSE) is used as a measure of how well the neural net has learnt.

SSE = sum square error is the sum of square of the difference between face output (expected) and the actual output from the ANN , SSE represent Performance function.

Duration measured by decisecond = 10^{-1} second.

We derive from these results that a modern approach is more effective than traditional approach but its duration is longer. This is normal because learning in a modern approach takes many steps to execute a total learning operation but it is more effective.

In the next results graph we find a same context; in traditional approach the learning curve started from bottom and increased to top discretely and alternated quickly to bottom continuously. The explanation of this result is that the learning operation walked without directing parameters. The learning started to grope for in blind area so it initiated moving unsteadily, then found some directing parameters it thought it a correct road of recognition. It did not take long before detecting fuzziness directing parameters and it started fall down quickly or rather steep slope in curve and the level of efficiency 16.8349 % .

Regarding the modern approach graph, it started by higher efficiency because providing it with directing parameters by modified input vector through (PCA+LDA) so it detected some faces; it started increasing in learning propagation so the results improved and it became stable and the level of efficiency 97.8% . From the above results we find that the effectiveness of modern approach of learning is better and higher in spite of its long duration that can be disregarded because of the present computers speeds.

We used sample of 5 grayscale faces images for performance classical approach experiment and counterparts sample of 5 grayscale fuzzy faces images, and also used 3 groups for performance new approach experiment each group contain sample of test images as follows:

- 1- 151 color faces
- 2- 131 color non faces
- 3- 52 color eyes

Although experiment use modern approach under going to duration is 530 d_{sec} to equal 1.5868263473 d_{sec} for each image alone, but this duration is surpass the performance duration 90 dsec of classical approach and equal 1.8 sec for each image alone, This is done by dividing the spent time on the number of images as follows :-

1. 530 d_{sec} and the number of colored images of 334 using the modern approach product would be 0.15868263473 sec for each image separately.
2. 90 d_{sec} and the number of images 5 using the performance classic approach would be 1.8 sec for each image separately.

Despite the apparent difference in time of the learning completion, which bias to favor of classical approach only that the apportionment of the time on the number of images and their quality, we have shown that the time of the completion of the learning experiment by new approach is under go 0.15868263473 sec and time the

5 Experimental Results

Already we completed experiment using software designed and manufactured by researcher called "*AmounRaa Neuro Face*" in desktop computer with processor Core™2 Duo 3.4 GHz , 2 GB DDRAM .The program receive inputs of color and mono images or capture images from motion detect of human faces and others faces, So proceed learning by two approaches (modern (PCA+LDA) and traditional) and calculate mean square error (MSE), sum square error (SSE), efficiency and duration of learning operation.

The feed-forward neural network was trained with the back-propagation learning algorithm during 500 epochs with learning rate 0.9, we performed two different experiments. In the first experiment, we use traditional approach of back-propagation algorithm without any changes. In the second experiment, we use traditional approach with modification of input dataset (vector) as previous concept After doing the experiments on a modern approach of neural networks learning by mixture of some multivariate analysis techniques (PCA+ LDA) and comparing it with traditional method of learning, We noticed the following results:-

Approach Measures	Modern Approach	Traditional Approach
MSE	0.007936579	0.000725832653220399
SSE	2.000018	0.003629163266102
Efficiency Experiment 1	97.4%	27.068 %
Duration 1	600 d_{sec}	90 d_{sec}
Efficiency Experiment 2	97.8%	16.8349 %
Duration 2	530 d_{sec}	90 d_{sec}
Duration (Of Each Image alone)	1.5868263473 d_{sec}	18 d_{sec}
Efficiency Trend line Equation (1)	$y = 2E-07x + 2$	$y = -0.0224x + 1.7686$
Efficiency Trend line Equation(2)	$y = 3E-07x + 2$	$y = -0.0191x + 1.5054$

Table 5.1 Results of experiments that show differences between traditional approach and modern approach

Where:-
$$\text{Efficiency} = \frac{\text{Number of success faces identification}}{\text{total number of faces}} \times 100$$

every iteration in order to minimize the cost function or the mean square error between the output vector and the face vector.

Input for hidden layer is given by

$$net_m = \sum_{z=1}^n x_z w_{mz} \quad (4.1)$$

The units of output vector of hidden layer after passing through the activation function are given by

$$h_m = \frac{1}{1 + \exp(-net_m)} \quad (4.2)$$

In same manner, input for output layer is given by

$$net_k = \sum_{z=1}^m h_z w_{kz} \quad (4.3)$$

and the units of output vector of output layer are given by

$$o_k = \frac{1}{1 + \exp(-net_k)} \quad (4.4)$$

For updating the weights, we need to calculate the error. This can be done by

$$E = \frac{1}{2} \sum_{i=1}^k (o_i - t_i)^2 \quad (4.5)$$

If the error is minimum than a predefined limit, training process will stop; otherwise weights need to be updated. For weights between hidden layer and output layer, the change in weights is given by

$$\Delta w_{ij} = \alpha \delta_i h_j \quad (4.6)$$

Where α is a training rate coefficient that is restricted to the range [0.01, 1.0], h_j is the output of neuron j in the hidden layer, and δ_i can be obtained by

$$\delta_i = (t_i - o_i)(1 - o_i) \quad (4.7)$$

o_i and t_i represents the real output and face output at neuron i in the output layer respectively.

Similarly, the change of the weights between hidden layer and output layer, is given by

$$\Delta w_{ij} = \beta \delta_{Hj} x_i \quad (4.8)$$

Where β is a training rate coefficient that is restricted to the range [0.01, 1.0], x_i is the output of neuron j in the input layer, and δ_{Hj} can be obtained by

$$\delta_{Hj} = x_i (1 - x_i) \sum_{j=1}^k \delta_j w_{ij} \quad (4.9)$$

x_i is the output at neuron i in the input layer, and summation term represents the weighted sum of all δ_j values corresponding to neurons in output layer that obtained in equation (42).

After calculating the weight change in all layers, the weights can simply updated by

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \Delta w_{ij} \quad (4.10) \quad \text{See [11]}$$

2. **Second way**, we obtained the combined distance vector by appending the d^{PCA} vector and d^{LDA} vector:

$$d = \{d_1^{PCA}, \dots, d_N^{PCA}, d_1^{LDA}, \dots, d_N^{LDA}\} \quad (3.2)$$

Where N is the number of images in the face database. If C is the number of the identities, also called classes, an identity C is associated to each couple (d_j^{PCA}, d_j^{LDA}) $j = 1, \dots, N$.

After computing and ordering the combined distance vector d , we can use our two algorithms as follows:

First, the KNN Algorithm:-

The most frequent identity among the first K components of d is selected.

- ❖ If the combined distance vector follows the eq. (3.1), then the algorithm is called "M-KNN" or "Mean-KNN".
- ❖ If it follows the eq. (3.2), then the algorithm is called algorithm "A-KNN" or "Append-KNN".

Second, the NM Algorithm:-

- ❖ We compute a template for each identity in the database.
- ❖ We selected the calculated average image for both PCA and LDA representations. The distance vectors d^{PCA} and d^{LDA} are composed by C components instead of N .
- ❖ These vectors are combined according to eq. (3.1) or (3.2).
- ❖ The identity associated to the smallest combined distance is selected.
- ❖ If the combined distance vector follows the eq. (3.1), then the algorithm is called "M-NM" or "Mean-NM".

If it follows the eq. (3.2), then the algorithm is called algorithm "A-NM" or "Append-NM". *see [7], [10]*

4. Learning Algorithm (Back-propagation)

Learning process in Back-propagation requires providing pairs of input and face vectors.

The output vector o of each input vector is compared with face vector t . In case of difference the weights are adjusted to minimize the difference. Initially random weights and thresholds are assigned to the network. These weights are updated

3. Images are well-aligned frontal view and clearly represent their hosts. They have low output on their own networks. These images will be used as both positive examples for their own networks and negative examples for other networks.
4. Images are not well cut and not can represent their hosts clearly. They have low output on their own networks. If they have high output on some other networks, they will be included as negative examples for other networks. Otherwise they will be ignored.

Once we get these new images, we add them to training examples and retrain the neural networks. Recognition errors will be corrected and the total performance will be improved. While adding some examples from a specific individual will improve the performance of this the own network, it will also influence the performance of other networks. In our experiments, the network training process continues until no significant recognition errors are found. **See [11]**

Integration of LDA and PCA for Face Recognition

Now, we suggest a new approach in which the integration of the two multivariate analysis techniques, PCA representation ("*eigenface*" approach) and LDA representation ("*fisherface*" approach) is considered. The integration of LDA and PCA techniques in the field of the face recognition can be applied as the following steps:

1. Representation of the face according to the PCA and the LDA approaches.
2. The distance vectors d^{PCA} and d^{LDA} from all the N faces in the database are computed.
3. The final decision, these two vectors are combined according to a given combination rule.

Two algorithms for the integration model we considered, these are:

- ❖ K-Nearest Neighbours (KNN algorithm).
- ❖ Nearest Mean (KN algorithm).

The First step, we normalize the distance vectors d^{PCA} and d^{LDA} in order to reduce the range of these distances in the interval [0, 1].

The second step is to compute a combined distance vector d that must contain both PCA and LDA information's. To do this we can use the following two ways:

1. **First way**, we obtained the combined distance vector by computing the mean vector:

$$d = \left\{ \frac{d_1^{PCA} + d_1^{LDA}}{2}, \dots, \frac{d_N^{PCA} + d_N^{LDA}}{2} \right\} \quad (3.1)$$

Where $d_{Mah}^{LDA} = \mu_i C^{-1} \mu_i^T$, $d_{Mah}^{PCA} = \phi_i C^{-1} x_k^T$ Are the Mahalanobis distances.

And $d_{EUC}^{LDA} = \bar{x}_i^2$, $d_{EUC}^{PCA} = \phi_i^2$ are the Euclidian distance.

1. More capable of distinguishing image variation due to identity from variation due to other sources such as illumination and expression.
- Suppose there are C classes
- Let μ_i be the mean vector of the class i , $i = 1, 2, \dots, C$
- Let M_i be the number of samples within class i , $i = 1, 2, \dots, C$

- Let $M = \sum_{i=1}^C M_i$ be the total number of samples, and

Within-class scatter matrix

$$S_w = \sum_{i=1}^C \sum_{j=1}^{M_i} (x_j - \mu_i)(x_j - \mu_i)^T \quad (2.2)$$

Between-class scatter matrix

$$(S_b \text{ has at most rank } C-1) \quad S_b = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T \quad (2.3)$$

$$\mu = \frac{1}{C} \sum_{i=1}^C \mu_i \quad (\text{Mean of entire data set})$$

Note: The types of distances can be calculated as follows:-

✎ $d_{EUC}^{LDA} = \bar{x}_i^2$ is actually Euclidian distance.

✎ $d_{Mah}^{LDA} = \mu_i C^{-1} \mu_i^T$ is actually Mahalanobis distance, which is distance to measure dissimilarity between several groups. See [10]

2.4 Training Set

After the neural network structure is set, the most important thing is to prepare the training examples. In the beginning of the training, we select a number of pictured images from each face that are well aligned frontal view. Any of them can represent their host clearly. All the images here are extracted or cut by the image detection code. These images will be used as positive examples for their own networks and negative examples for other networks.

Here we only deal with images which assume that they are always images. In our tests, most of the time the image detection code can find the image if it exists. The database is not supposed to handle non-image images because in our situation it is unnecessary and it will make network training very difficult.

After the basic neural networks are created, we run them over new images from the individuals in our database. If the image fails to pass the image detection test, it will be ignored. If the image detection code reports image in the image, it will be applied to the image recognition code. We check the recognition result to find more images for training. Here each image will fall into one of the four categories as following:

1. Images have high output on their own networks and low output on other networks. No action here.
2. Images have high output on both their own networks and some other networks. These images will be used as negative examples for other networks.

2.3.1 Methodology

Suppose $X_1, X_2, \dots, X_M$ are $M \times 1$ Vectors

❖ Step 1: $\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i$

❖ Step 2: Subtract the mean $\phi_i = x_i - \bar{x}$

❖ Step 3: From the matrix $A = [\phi_1 \phi_2 \dots \phi_M]$ ($N \times M$ matrix), then

compute $C = \frac{1}{M} \sum_{n=1}^M \phi_n \phi_n^T = AA^T$

(Sample Covariance matrix, $N \times N$, characterizes the scatter of the data)

❖ Step 4: Compute the eigenvalues of $C: \lambda_1 > \lambda_2 > \dots > \lambda_N$

❖ Step 5: Compute the eigenvectors of $C: u_1 > u_2 > \dots > u_N$

Since the C is symmetric, $u_1 > u_2 > \dots > u_N$ from the basis (i.e., any vector x actually $(x_i - \bar{x})$, can be written as linear combination of the eigenvectors):

$$x_i - \bar{x} = b_1 u_1 + b_2 u_2 + \dots + b_N u_N = \sum_{i=1}^N b_i u_i$$

$$b_i = U_i^T (x - \bar{x}) \quad (2.1)$$

❖ Step 6 : (Dimensionality reduction step)

Keep only the terms corresponding to the K largest of eigenvalues

$$x_i - \bar{x} = \sum_{i=1}^K b_i u_i \text{ where } K < N$$

The representation of $x_i - \bar{x}$ into the basis $u_1, u_2, \dots, u_K$ is thus

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix}$$

Note: The types of distances can be calculated as follows:-

$d_{EUC}^{PCA} = \phi_i^2$ is actually Euclidian distance.

$d_{Mah}^{PCA} = \phi_i C^{-1} \phi_i^T$ is actually Mahalanobis distance.

2.3.2 PCA and classification

PCA is not always an optimal dimensionality-reduction procedure for classification purposes.

2.3.3 Second, by using LDA technique

What is the goal of DA?

1. Perform dimensionality reduction "*while preserving as much of the class discriminatory information as possible*".
2. Seeks to find directions along which the classes are best separated.
3. Takes into consideration the scatter within-classes but also the scatter between-classes.

This paper introduces is divided into parts as follows: Input data generation by computerized images is described in **Section two**. Possibility of using modified neural networks for automatic faces pattern recognition is described in **section three**. Presentation a concept of principle components analysis and steps of applying integration between PCA and LDA are sketched in **section four**. Description of automatic face detection is presented in **section five**. Automatic face recognition by new technique is described in **section six**. Conclusions are drawn in **section seven**.

2.1 Principal Components Analysis

In statistics, principal components analysis (PCA) is a technique that can be used to simplify a dataset. It is a linear transformation that chooses a new coordinate system for the data set such that the greatest variance by any projection of the data set comes to lie on the first axis (called the first principal component), the second greatest variance on the second axis, and so on. PCA can be used for reducing dimensionality in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. The idea is that such low-order components often contain the "most important" aspects of the data. However, this is not necessarily the case, depending on the application [5]. PCA is also called the (discrete) Karhunen-Loève transform (or KLT, named after Kari Karhunen and Michel Loève) or the Hotelling transform (in honor of Harold Hotelling). PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance. This advantage, however comes at the price of greater computational requirement if compared, for example, to the discrete cosine transform. Unlike other linear transforms, the PCA does not have a fixed set of basis vectors. Its basis vectors depend on the dataset. See [11]

2.2 Linear Discriminate Analysis

Linear discriminate analysis (LDA) and the related Fisher's linear discriminate are used in statistics to find the linear combination of features which best separate two or more classes of object or event. The resulting combinations may be used as a linear classifier, or more commonly in dimensionality reduction before later classification.

LDA is closely related to ANOVA (analysis of variance) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. In the other two methods however, the dependent variable is a numerical quantity, while for DA it is a categorical variable (i.e. the class label). See [11]

2.3 Face Recognition Using (PCA+LDA)

2.4 First, by using PCA technique

The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the dataset.

In our approach we can use the integration between PCA and LDA for increasing learning efficiency of networks

Modern day computer imaging systems are capable of comparatively high resolution by utilizing synthetic aperture processing methods. The motivating requirement is to employ computer imagery as an aid in finding comparatively small mobile or relocatable faces. The need is to detect faces, cue the analyst, classify/recognize faces, and identify faces if possible. This presentation will address four areas:

- ❖ Face detection
- ❖ Face discrimination
- ❖ Face classification/recognition
- ❖ Performance prediction of recognized Faces

It has been found through experience that the logical and computational problem can be significantly reduced with the detection/discrimination approach prior to classification/recognition/identification. It needs a gross technical support of these basic objectives, such as speckle filtering, super resolution, effects of image compression, coherent change detection, and synthesize skin , segmentation and classification.

We propose an artificial neural network (ANN) architecture for post-processing of faces field data. This architecture is a feed-forward neural network model modified by integration between PCA and LDA, which is learned by using the back-propagation learning algorithm on two features extracted from the data in the vision field, the computerized images supporting this training data. This approach uses data from a calibrated some faces to train the corresponding neural network, which is then used for faces detection over much larger faces. This trail by computerized images shows that the trained network architecture can be effectively used in faces which are distances remote from the calibration faces. It is also effective when tested with the images obtained with digital cameras which have different characteristics from those which were used to collect the network training data, shown in Figure 1. see [11]

New Approach (PCA+LDA) Overview

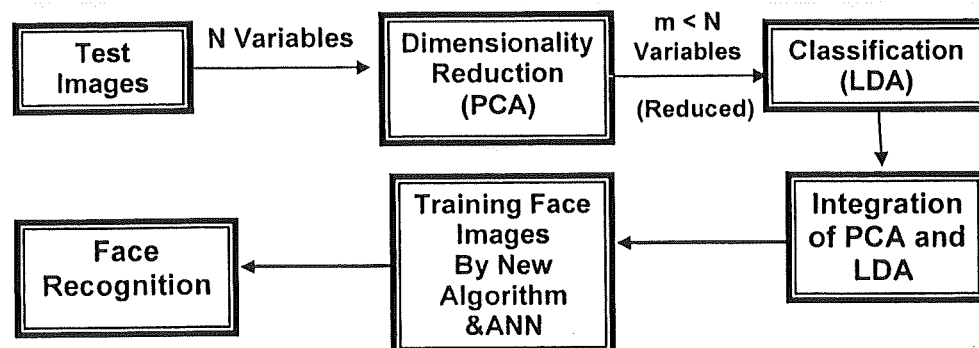


Figure 1 New Approach Overview

Face Recognition Using Integration between Multivariate Analysis Techniques And Artificial Neural Networks

Abd Alfattah.M.Kandel, Sahar Adel Raafat
Fared Naguib Guirguis, Hazem Abd Aghafar Souror

Abstract

Face recognition are used in many aspects applications. But this technique suffer from many important shortcomings of the most important the learning efficiency which negatively affect the reliability of use. This paper presents a modified algorithm for face recognition by using integration between feed-forward neural networks and multivariate analysis techniques are principle components analysis (PCA) and linear discriminate analysis (LDA). Many data preprocessing algorithms used for training a basic feed-forward neural networks showed commonly results perhaps were better than other algorithms, After using integration between a feed-forward neural networks and multivariate analysis techniques (PCA+LDA), the expected result is the best efficient and effective algorithm. The face recognition will be most efficient with this proposed system.

Keywords: Face Recognition, Neural Networks, Pattern Classification, Pattern Recognition, Distances, Feed-Forward, PCA, LDA.

1. Introduction

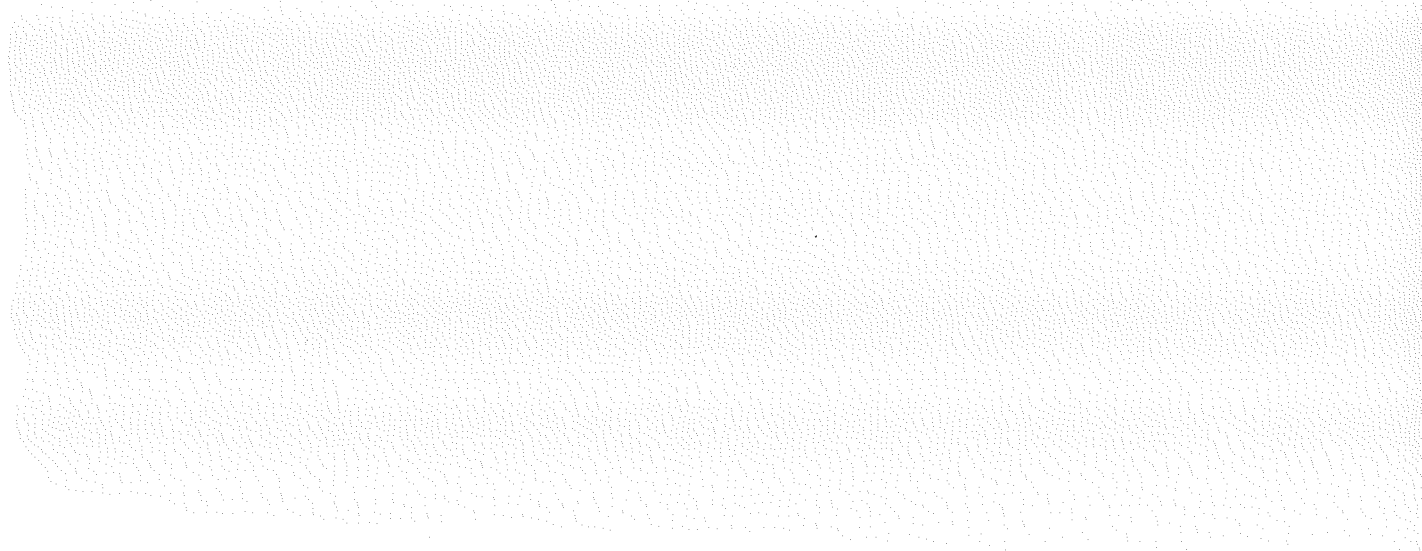
The modern sciences applications use some computerized techniques including face recognition. Face recognition is an approach to the detection and identification of faces. Over the last few years the most researches are not effective enough for face recognition problem due to the inefficiencies of the algorithms used to represent faces. The face representation was performed by using two categories. The First category is global approach or appearance-based, which uses holistic texture features and is applied to the face or specific region of it. The second category is feature-based or component-based, which uses the geometric relationship among the facial features like mouth, nose, and eyes implemented feature-based approach by a geometrical model of a face by 2-D elastic graph. Another example of feature-based was done by independently matching templates of three facial regions (eyes, mouth and nose) and the configuration of the features was unconstrained since the system didn't include geometrical model.

Neural networks are the most famous approach use for face recognition because it includes a property of pattern recognition. But the applicable force of the neural networks depends on learning efficiency.

-
- 1- Professor of Statistics , Benha University.
 - 2- Professor of Statistics , Benha University.
 - 3- Department of Statistics , Benha University.
 - 4- Department of Statistics , Benha University.

There is a very large number of people who are
in the same position as you are in.

There is a very large number of people who are
in the same position as you are in.



There is a very large number of people who are
in the same position as you are in.

7. Nayak, T. and Adeshiyan, S. (2009). A unified framework for analysis and comparison of randomized response surveys of binary characteristics. *Journal of Statistical Planning and Inference* **139**, 2757-2766.
8. Riad, A. (2002). Randomized response study with application. *Unpublished M.S thesis. Mansorra university.*
9. Warner, S. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60**, 63-69.

$$H_A: \pi_s > 0$$

By performing Z-test, we have the calculated $Z = 0.66$ and $Z_{0.01} = 2.326$.

Then the null hypothesis is accepted at $\alpha = 0.01$.

8. Conclusions

In this paper, the mixed two stage randomized response model and its stratification RR model are proposed, which are more confidentiality than Mangat and Singh (1990) and Kim and Elam (2005) models. We determined the optimal sample size and the optimal allocation for the stratified random sample in the case of minimizing the variances of the estimators. An application for the suggested model is presented to estimate the proportion of informal marriage among the student in the university level. There are several issues which support future researches, using the mixed two-stage randomized response model to estimate the proportion of a sensitive attribute like tax evasion, drunken driving, gambling, drug abuse, and bribe taking. Estimation of the proportion of a sensitive attribute from the mixed two-stage randomized response model using Expectation-Maximization (EM) algorithm and comparative studies on costs per sampling unit under the different randomized response techniques.

References

1. Greenberg, B., Abul-El, A., Simmons, W. and Horvitz, D. (1969). The unrelated question randomized response: Theoretical framework. *Journal of the American Statistical Association*, **64**, 529-539.
2. Kim, J. and Warde, W. (2004). A stratified Warner's randomized response model. *J. Statist. Plann. Inference* **120** (1-2), 155-165.
3. Kim, J. and Elam, M. (2005). A two-stage stratified Warner's randomized response model using optimal allocation. *Metrika* **61**, 1-7.
4. Kim, J. and Warde, W. (2004). A mixed randomized response model. *J. Statist. Plann. Inference*, **133**, 1-11.
5. Mangat, N. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika* **77**, 439-442.
6. Moors, J. (1971). Optimization of the unrelated question randomized response model. *J. Amer. Statist. Assoc.* **66**, 627-629.

$$n_1 = 252 \quad \text{and} \quad n_2 = 138$$

7.3 Results

The estimated probabilities are:

$$\hat{\theta}_1 = \frac{76}{252} = 0.3016 \quad \text{and} \quad \hat{\theta}_2 = \frac{40}{138} = 0.2898$$

where $\hat{\theta}_i$ is the estimate of the proportion of "Yes" answers in the sample from the i th stratum, $i = 1, 2$.

Equations (5.2), (5.3), (5.4) and (5.5) are applied, and the results are shown in table (3)

Table (3): Proportions of informal marriage

	Proportion of informal marriage	Variance of proportion of informal marriage
Male	$\hat{\pi}_{s_1} = 0.0340$	$V(\hat{\pi}_{s_1}) = 0.00260$
Female	$\hat{\pi}_{s_2} = 0.0153$	$V(\hat{\pi}_{s_2}) = 0.00488$
Population	$\hat{\pi}_s = 0.0273$	$V(\hat{\pi}_s) = 0.00169$

The results in table (3) reveal the following;

1. The estimated proportion of informal marriage among male students in the sampled population is 0.0340 with variance 0.00260 .
2. The estimated proportion of informal marriage among female students in the sampled population is 0.0153 with variance 0.00488.
3. The estimated proportion of informal marriage among students in the whole sampled population is 0.0273 with variance 0.00169 .
4. The estimated proportion of informal marriage among male students is greater than that of the female students.

7.4 Test of hypothesis

To test

$$H_0: \pi_s = 0$$

versus

students registered at the faculty of commerce, Benha University during the academic year 2008/2009 is $N = 16626$ allocated into $N_1 = 10749$ male and $N_2 = 5877$ females. From previous field experience has indicated that satisfactory results can be achieved with sensitive question proportion equal $0.2(\pm 0.1)$ or $0.8(\pm 0.1)$. So we chose $P_1 = 0.4$ and $P_2 = 0.3$ for the first device, and $M_1 = 0.8$ and $M_2 = 0.7$ for the second device (see for example Riad (2002)).

7.1 The randomized device

There are some questions concerning the problems that some students have. This can not directly be asked because many students consider it personal. One can not blame them for feeling this way, but there is a real need for information about these problems in order to help them. We now have a way that makes it possible for students to give information about these problems without telling any one about their real situation. As mentioned before, for each stratum, there are two stage randomized response model. For the stratum $i; i = 1, 2$ the randomization device R_{1i} consists of three statements; One deals with the problem, the second is completely unrelated and the third is "Go to randomization device R_{2i} ". The randomization device R_{2i} also consists of two statements. One deals with the problem, and the other is the negative sensitive question. Only one of these questions is selected by chance and answered. For the stratum i , the interviewee does not know which of the questions it is. Having two boxes, each box contains 100 cards of identical size which are simply shuffled very well, then the interviewee chooses one card randomly, opens the chosen card, and then reads the question. The interviewee answers only with "yes" or "no" and doesn't tell the interviewer which question is being answered.

7.2 Determining the sample size (choice the sample)

Since there is no prior information on proportion of informal marriage among teenagers (π_{s_i}) and assuming that the cost of conducting the survey is fixed, so we take $(\pi_{s_i} = 0.5)$, standard value $(Z = 2)$ and error $(d = 0.05)$, according to sample size, a sample of $n = 390$ students was required. Using proportional allocation to allocate the total sample size into the two strata, we obtain

$$\beta = \frac{[-4M_i P_i (1 - P_i) \pi_{s_i}] + [M_i P_i (2 - P_i (M_i + 2))]}{n_i [2M_i P_i]^2}$$

Taking partial derivatives and solving the system of equations

$$\begin{aligned} \frac{\partial \Phi}{\partial n_i} &= 0 & i = 1, 2, \dots, k, \\ \frac{\partial \Phi}{\partial \lambda} &= 0 \end{aligned}$$

Using the fact that $n = \sum_{i=1}^k n_i$, it follows that

$$n_i = \frac{w_i \left[(\pi_{s_i} (1 - \pi_{s_i}) + \beta) / c_i \right]^{\frac{1}{2}}}{\sum_{i=1}^k w_i \left[(\pi_{s_i} (1 - \pi_{s_i}) + \beta) / c_i \right]^{\frac{1}{2}}} \quad i = 1, 2, \dots, k \quad (6.3)$$

and

$$n = \frac{(C - c_0) \sum_{i=1}^k w_i \left[(\pi_{s_i} (1 - \pi_{s_i}) + \beta) / c_i \right]^{\frac{1}{2}}}{\sum_{i=1}^k w_i \left[(\pi_{s_i} (1 - \pi_{s_i}) + \beta) / c_i \right]^{\frac{1}{2}}} \quad (6.4)$$

Since $\pi_{s_i}; i = 1, 2, \dots, k$ are unknown, they would be replaced by either a suitable suggested values or by 0.5 if no information is available. In general, informal marriage is perceived to be an illegal and socially undesirable behavior, then, in the next section, we are using a stratified mixed two stage randomized response model to estimate the proportion of the informal marriage among teenagers.

7. Application

This study is designed to estimate the proportions of the informal marriage for male and female students, and the proportion of the informal marriage in the whole population of undergraduate students in the faculty of commerce, Benha University. We assume that the study population consists of two strata (male and female) i.e. $k = 2$. The

6. The optimal sample size, n , and the optimal allocations of the stratified random sampling model.

In stratified sampling it is often of interest to find the values of the sample sizes $n_1, n_2, \dots, n_k$, so that the variances of the obtained estimators are minimum, under fixed cost, or the cost is minimized under the assumption that the variances of the estimators are predetermined. Consider the case of minimizing the variances of the estimators assuming that the cost of conducting the survey is fixed. Suppose that we have a linear cost function given by

$$C = c_0 + c_i \quad (6.1)$$

Where

c_i is the variable cost of interviewing an individual from stratum i and

c_0 is the general fixed cost of the entire survey.

We would like to find the value of $n_1, n_2, \dots, n_k$ which minimize the variance of $\hat{\pi}_s$ subject to a fixed cost. Thus, we have the following optimization problem

$$\min_{n_1, \dots, n_k} \sum_{i=1}^k w_i^2 \left\{ \frac{\pi_{s_i}(1 - \pi_{s_i})}{n_i} + \frac{[-4M_i P_i(1 - P_i)\pi_{s_i}] + [M_i P_i(2 - P_i(M_i + 2))]}{n[2M_i P_i]^2} \right\}$$

subject to

$$C = c_0 + \sum_{i=1}^k c_i n_i,$$

Using the method of Lagrange multipliers, we can equivalently have the unconstrained minimization problem

$$\min_{n_1, \dots, n_k, \lambda} \Phi(n_1, n_2, \dots, n_k, \lambda)$$

Where

$$\Phi(n_1, \dots, n_k, \lambda) = \sum_{i=1}^k w_i^2 \left\{ \frac{\pi_{s_i}(1 - \pi_{s_i})}{n_i} + \frac{[-4M_i P_i(1 - P_i)\pi_{s_i}] + [M_i P_i(2 - P_i(M_i + 2))]}{n_i[2M_i P_i]^2} \right\} + \lambda \left(c_0 + \sum_{i=1}^k c_i n_i - C \right) \quad (6.2)$$

Let

$$\theta_i = P_i \pi_{s_i} + (1 - 2P_i) + P_i [M_i \pi_{s_i} + (1 - M_i)(1 - \pi_{s_i})] \quad i = 1, 2, \dots, k \quad (5.1)$$

where θ_i is the proportion of "Yes" responses and π_{s_i} is the true proportional of respondents with the sensitive trait in the sample from stratum i .

The maximum likelihood estimate of π_{s_i} is

$$\hat{\pi}_{s_i} = \frac{\hat{\theta}_i - P_i(1 - M_i) - (1 - 2P_i)}{2M_iP_i} \quad (5.2)$$

where $\hat{\theta}_i$ is the ML estimate of the proportion of "Yes" answers in the sample from stratum i . Since each n_i has a Binomial (n_i, θ_i) , then $\hat{\pi}_{s_i}$ is an unbiased estimator for π_{s_i} and the variance of $\hat{\pi}_{s_i}$ is

$$Var(\hat{\pi}_{s_i}) = \frac{\pi_{s_i}(1 - \pi_{s_i})}{n_i} + \frac{[-4M_iP_i(1 - P_i)\pi_{s_i}] + [M_iP_i(2 - P_i(M_i + 2))]}{n_i[2M_iP_i]^2} \quad (5.3)$$

Since the selections in different strata are made independently, then the estimators for each individual stratum can be added together to obtain an estimator for the whole population. The maximum likelihood estimator of π_s , the proportion of respondents with the sensitive trait, is

$$\hat{\pi}_s = \sum_{i=1}^k w_i \hat{\pi}_{s_i} = \sum_{i=1}^k w_i \frac{\hat{\theta}_i - P_i(1 - M_i) - (1 - 2P_i)}{2M_iP_i} \quad (5.4)$$

where, $w_i = N_i / N$; $i = 1, 2, \dots, k$, so that $\sum_{i=1}^k w_i = 1$. Then $\hat{\pi}_s$ is an unbiased estimator for π_s . The variance of the estimator $\hat{\pi}_s$ is

$$Var(\hat{\pi}_s) = \sum_{i=1}^k w_i^2 \left[\frac{\pi_{s_i}(1 - \pi_{s_i})}{n_i} + \frac{[-4M_iP_i(1 - P_i)\pi_{s_i}] + [M_iP_i(2 - P_i(M_i + 2))]}{n_i[2M_iP_i]^2} \right] \quad (5.5)$$

stage randomized response model the following observation can be withdrawn:

1- The modified two stage randomized response model is more efficient than Mangat and Singh two stage randomized response model.

2- The mixed two stage randomized response model is more efficient than the modified two stage randomized response model.

Thus, we may conclude that the proposed mixed two stage randomized response model is more efficient than Mangat and Singh two stage randomized response model.

5. A mixed two-stage randomized response model using stratification

Let the population of size N be partitioned into k strata of sizes N_i ; $i = 1, 2, \dots, k$, where $N = \sum_{i=1}^k N_i$. A random sample is then selected by simple random sampling from each stratum. In the first stage of the survey interview, each individual respondent in the sample from stratum i is instructed to use the randomization device R_{1i} which consists of a sensitive question (S) card with probability P_i , its innocuous question card with probability $1 - 2P_i$ and a direction card "Go to the randomization device R_{2i} ", with probability P_i . The respondents in the second stage of stratum i are instructed to use the randomization device R_{2i} which consists of a sensitive question with probability M_i and its negative question with probability $1 - M_i$. The respondent should answer the question with a "Yes" or "No" without reporting which question card he or she has in order to protect the respondent's privacy.

Let

n_i be the number of units in the sample from stratum i ,

n be the total number of units in all samples drawn from all strata,

where $n = \sum_{i=1}^k n_i$.

Under the assumption that the "Yes" and "No" reports are made truthfully and P_i and M_i are set by the researcher, the probability of a "Yes" answer in stratum i for this procedure is

From table(1), it is noted that the proposed model is more efficient than Mangat and Singh (1990) model in most cases, i. e. proposed model is more efficient than Mangat and singh model when $P + M \geq 0.7$.

An efficiency comparison of the mixed two stage randomized response technique and the modified two stage randomized response technique is done.

The variance of the mixed two stage randomized response model is

$$Var(\hat{\pi}_s)_m = \frac{\pi_s(1 - \pi_s)}{n} + \frac{[-4MP(1 - P)\pi_s] + [MP(2 - P(M + 2))]}{n[2MP]^2} \quad (4.4)$$

The percent of $Var(\hat{\pi}_s)/Var(\hat{\pi}_s)_m$ is computed, given the relative efficiency of $(\hat{\pi}_s)_m$ relative to $\hat{\pi}_s$

$$Percent RE = \frac{Var(\hat{\pi}_s)}{Var(\hat{\pi}_s)_m} 100$$

Since n and M did not affect the computation of the percent RE , we used $n = 1000$ and $M = 0.6$. Table (2) Shows the value of the percent relative efficiency for different cases of π_s and P . From table (2), it is noted that the proposed mixed model is more efficient than the modified two stage randomized response model.

Table (2): The efficiency of the mixed two stage randomized response model relative to modified two stage randomized response model

P	π_s					
	0.1	0.2	0.3	0.4	0.5	0.6
0.1	494.0742	590.2771	740.0602	1006.599	1616.379	4451.236
0.2	1548.487	1702.985	1929.929	2280.540	2873.563	4058.739
0.3	4669.857	4836.298	5143.94	5648.413	6465.517	7858.524
0.4	22485.95	22255.58	22666.51	23792.89	25862.07	29386.83

From table (2), it is noted that; The mixed two-stage randomized response model is more efficient than the modified two stage randomized response model.

From previous comparison among Mangat and Singh (1990) model, the modified two stage randomized response model and the mixed two

$$Var(\hat{\pi}_s)_{ma} = \frac{\pi_s(1-\pi_s)}{n} + \frac{(1-P)(1-M)[1-(1-M)(1-P)]}{n[2M-1+2P(1-M)]^2} \quad (4.1)$$

The variance of modified two-stage randomized response model in the special case of the proposed model is

$$Var(\hat{\pi}_s) = \frac{\pi_s(1-\pi_s)}{n} + \frac{[P+M(1-2P)][(1-M)(1-2P)+P]}{n[2P-1+2M(1-2P)]^2} \quad (4.2)$$

Let

$$G = Var(\hat{\pi}_s)_{ma} - \frac{\pi_s(1-\pi_s)}{n}$$

and

$$F = Var(\hat{\pi}_s) - \frac{\pi_s(1-\pi_s)}{n}$$

The ratio G/F is used to compute the relative efficiency (RE) of the proposed model, based on the estimator $\hat{\pi}_s$, with respect to the Mangat and Singh (1990) model, based on the estimator $\hat{\pi}_s$. The value of G/F is

$$G/F = \frac{[P+M(1-2P)][(1-M)(1-2P)+P]}{[(2P-1)+2M(1-2P)]^2} \cdot \frac{[2M-1+2P(1-M)]^2}{(1-P)(1-M)[1-(1-P)(1-M)]} \quad (4.3)$$

The following table shows the percent value of G/F . In these comparisons, it is assumed that: $(P_1 = P_2 = P) = 0.1, 0.2, 0.3, 0.4$ and $M = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

Table (1): The efficiency of the modified two stage randomized response model relative to Mangat and Singh model(1990)

M	$P_1 = P_2 = P$			
	0.1	0.2	0.3	0.4
0.1	90.00594	80.19318	63.55172	24.51691
0.2	80.19318	57.13306	23.90422	10.96866
0.3	63.55172	23.90422	1.523109	407.8818
0.4	24.51691	10.96866	407.8818	5308.333
0.6	323.7956	1019.118	3727.232	23126.32
0.7	235.2621	606.3657	1929.509	10770.73
0.8	231.7374	577.6602	1761.130	9359.26
0.9	295.8491	800.5774	2489.653	13065.43

$X_i = 1$ if the i -th sample element reports a "yes" answer.

$X_i = 0$ if the i -th sample element reports a "no" answer.

$$\therefore \theta = \Pr(X_i = 1) = P\pi_s + (1 - 2P)\pi + P[M\pi_s + (1 - M)(1 - \pi_s)]$$

As $\pi = 1$, we find that

$$\theta = \Pr(X_i = 1) = P\pi_s + (1 - 2P) + P[M\pi_s + (1 - M)(1 - \pi_s)] \quad (3.1)$$

where π is the proportion of "Yes" answers for the innocuous question.

Then, the formula (2.1) can be rewritten as

$$\theta = \pi_s [2MP] + P(1 - M) + (1 - 2P) \quad (3.2)$$

and the probability that the respondent reports a "no" answer

$$\Pr(X_i = 0) = 1 - \theta \quad (3.3)$$

Since X_i is Bernoulli (θ) and $X_1, X_2, \dots, X_n$ are independent and identically distributed i.i.d.. Then $\sum_{i=1}^n X_i = n_1$ is Binominal (n, θ). Thus, the Maximum likelihood (ML) estimator of π_s is

$$\hat{\pi}_s = \frac{\hat{\theta} - P(1 - M) - (1 - 2P)}{2MP} \quad (3.4)$$

where $\hat{\theta}$ is the ML estimate of the proportion of "yes" answer in the sample. Since $n\hat{\theta}$ has a Binomial (n, θ), then the estimator $\hat{\pi}_s$ is unbiased. The variance of $\hat{\pi}_s$ is

$$Var(\hat{\pi}_s)_m = \frac{\pi_s(1 - \pi_s)}{n} + \frac{[-4MP(1 - P)\pi_s] + [MP(2 - P(M + 2))]}{n[2MP]^2} \quad (3.5)$$

4. Efficiency comparison

A numerical comparison is performed to investigate the efficiency of the proposed estimator relative to Mangat and singh (1990) estimator. This comparison is made to determine under which conditions one of them will be superior. The Variance of Mangat and singh (1990) estimator is

Since X_i is Bernoulli (θ) and $X_1, X_2, \dots, X_n$ are i.i.d (independent and identically distributed). Then $\sum_{i=1}^n X_i = n_1$ is Binominal (n, θ). Thus, the Maximum likelihood (ML) estimator of π_s is

$$\hat{\pi}_s = \frac{\hat{\theta} - P_2 - (1 - P_1 - P_2)(1 - M)}{2P_1 - 1 + 2M(1 - P_1 - P_2)} \quad (2.3)$$

where $\hat{\theta}$ is the ML estimate of the proportion of "yes" answer in the sample. Since $n\hat{\theta}$ has a Binomial(n, θ), then the estimator $\hat{\pi}_s$ is unbiased. The variance of $\hat{\pi}_s$ is

$$Var(\hat{\pi}_s) = \frac{\pi_s(1 - \pi_s)}{n} + \frac{[P_1 + M(1 - P_1 - P_2)][((1 - M)(1 - P_1 - P_2) + P_2)]}{n[2P_1 - 1 + 2M(1 - P_1 - P_2)]^2} \quad (2.4)$$

In case of ($P_1 = P_2 = P$) and $P < 0.5$, the variance of $\hat{\pi}_s$ is

$$Var(\hat{\pi}_s) = \frac{\pi_s(1 - \pi_s)}{n} + \frac{[P + M(1 - 2P)][((1 - M)(1 - 2P) + P)]}{n[2P - 1 + 2M(1 - 2P)]^2} \quad (2.5)$$

3. A mixed two-stage randomized response model

A single sample of size n is selected by simple random sampling from the population. Each respondent from the sample is provided with two randomization devices. The randomization device R_1 consists of three statements. The first statement is that "I belong to the sensitive trait group" (with probability P) and the second is "I am a member of the innocuous trait group" (with probability $1 - 2P$) and the third is "Go to randomization device R_2 " (with probability P). The randomization device R_2 consists of two statements, which are "I belong to the sensitive group" and "I do not belong to the sensitive group" with known probabilities M and $1 - M$, respectively. The respondent should answer the question with a "yes" or "no" without reporting which question he or she has in order to protect the respondent's privacy. Under the assumption that these reported yes and no are made truthfully, let

π_s be the true proportional of the population with the sensitive characteristics.

randomized response model and a mixed two-stage randomized response model. The efficiency comparison between proposed models and Mangat and Singh (1990) two-stage RR model, are presented in section 4. A mixed two-stage randomized response model using stratification is given in section 5. The optimal sample size, n , and the optimal allocations of the stratified random sampling model are derived in section 6. An application of the mixed two-stage randomized response model using stratification to estimate the proportion of informal marriage among the students in the university level is given in section 7. Section 8 is devoted for conclusions.

2. A modified two-stage randomized response model

A single sample with size n is selected by simple random sampling from the population. Each respondent from the sample is provided with two randomization devices. The randomization device R_1 consists of three statements. The first statement is that "I belong to the sensitive trait group" (with probability P_1) and the second is "I do not belong to the sensitive group" (with probability P_2) and the third is "Go to randomization device R_2 " (with probability $1 - P_1 - P_2$). The randomization device R_2 also consists of two statements, which are "I belong to the sensitive group" and "I do not belong to the sensitive group" with known probabilities M and $1 - M$, respectively. The respondent should answer the question with a "yes" or "no" without reporting which question he or she has in order to protect the respondent's privacy. Under the assumption that these reported yes and no are made truthfully, let

π_s be the true proportional of the population with the sensitive characteristics.

$X_i = 1$ if the i -th sample element reports a "yes" answer.

$X_i = 0$ if the i -th sample element reports a "no" answer.

$$\therefore \theta = P(X_i = 1) = P_1\pi_s + P_2(1 - \pi_s) + (1 - P_1 - P_2)[M\pi_s + (1 - M)(1 - \pi_s)] \quad (2.1)$$

And the probability that the respondent reports a "no" answer is

$$\Pr(X_i = 0) = 1 - \theta \quad (2.2)$$

different from open or direct surveys was needed to make people comfortable and to encourage truthful answers.

Warner (1965) introduced the randomized response (RR) model as an alternative survey technique for socially undesirable or incriminating behavior questions in order to reduce response error, protect a respondent's privacy, and increase response rates. Warner's model draws respondents using simple random sampling with replacement from the population. It requires the interviewee to give a "Yes" or "No" answer either to the sensitive question or to its negative depending on the outcome of a randomizing device not reported to the interviewer. Greenberg et al. (1969) introduced the theoretical framework for unrelated question randomized response model that is a variation of Warner's (1965) RR model. The unrelated question RR model has one question that asks about a very sensitive trait and a second question that asks about an innocuous (or non-sensitive) trait. Mangat and Singh (1990) proposed a two-stage RR model in which each interviewee (who is selected using simple random sampling) is provided with two randomization devices. The first one consists of two statements: (1) "I belong to the sensitive trait group" (with probability P) and (2) "Go to the second randomization device" (with probability $1-P$). The second randomization device also consists of two statements, which are "I belong to the sensitive group" and "I do not belong to the sensitive group" with known probabilities M and $1-M$, respectively. Kim and Elam (2005) introduced stratified to Mangat and Singh's (1990) two-stage RR model. Nayak and Adeshiyan (2009) developed an approach for comparing RR procedures, taking both respondents' protection and statistical efficiency into account. They showed that Mangat and Singh (1990) compared variances of their estimator and Greenberg et al's (1969) estimator without holding equal respondents' protection. Similarly, they compared the variances of Warner's estimator and their estimator, not taking respondents' protection into account. The present research is directed towards introducing new randomized response procedure, based on Mangat and Singh (1990) two-stage randomized response model, to increase the respondent confidentiality and to make him feel more protective. So The remainder of the present research is organized as follows: sections 2 and 3 introduce a modified two-stage

A mixed two-stage randomized response model: An application to the phenomenon of informal marriage

Sahar Adel Mohamed⁽¹⁾ Farid Naguib Guirguis⁽²⁾
Mahmoud Mansour Mohamed⁽³⁾

Abstract

A new randomized response procedure is proposed based on Mangat and Singh two-stage randomized response model and keeps the respondents' confidentiality. The basic idea of the suggested procedure is to increase the respondent confidentiality and to make him feel more protective, since his response is to be given after using three questions in the first stage not two questions, as Mangat and Singh model and Kim and Elam model suggested in their designs. The proposed model is extended to the stratified sample. The optimal sample size and the optimal allocation for the stratified random sample in the case of minimizing the variances of the estimators, are derived. An application for the suggested model to estimate the proportion of informal marriage among the students in the university level is given.

1. Introduction

Researchers often use sample survey methodology to obtain information about a large aggregate or population by selecting and measuring a sample from that population. Data obtained from surveys are affected by two main sources of error; The first is the sampling error that results from taking a sample instead of enumerating the whole population; The second type of error is the non-sampling error. The main sources of non-sampling error in any survey are non-response bias and response bias. Non-response bias arises from subjects' refusal to respond and response bias arises from giving incorrect responses. When open or direct surveys are about sensitive matters (e.g., tax evasion habits, addiction, drunken driving and abortion), non-response bias and response bias become serious problems because people do not often wish to give correct information. In order to reduce non-response and response bias, a survey technique

⁽¹⁾ Department of Statistics, Mathematics, and Insurance, Faculty of commerce, Benha University

⁽²⁾ Department of Statistics, Mathematics, and Insurance, Faculty of commerce, Benha University

⁽³⁾ Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha university

33. Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). Semiparametric Regression, Cambridge University Press.
34. Sperlich, S., Hardle, W., and Aydinli, G. (2006). The Art of Semiparametrics, Physica-Verlag, Germany.
35. Stone, C.J. (1980). "Optimal rates of convergence for nonparametric estimators", The Annals of Statistics, 8, 6, 1348-1360..
36. Wand, M.P., and Jones, M.C. (1995). Kernel Smoothing, Chapman & Hall, London.
37. Watson, G. S. (1964). "Smooth regression analysis", Sankhya, series A, 26, 359-372.
38. Weisberg, S. and Welsh, A. H. (1994). " Adapting for the missing link", The Annals of Statistics, 22, 1674-1700.
39. Wong, W.H., and Severini, T.A. (1991). "On Maximum Likelihood Estimation in Infinite Dimensional Parameter Spaces", The Annals of Statistics, 19,2, 603-632.
40. Xia, Y., Tong, H., Li, W.K., and Zhu, L.X. (2002). "An adaptive estimation of dimension reduction space", Journal of The Royal Statistical Society, Ser. B, 64, 3, 363-410.
41. Yatchew, A. (2003). Semiparametric Regression for the Applied Econometrician, Cambridge University Press, UK.

11. **Hardle, W., Stoker, T.M.** (1989). "Investigating smooth multiple regression by the method of average derivatives", *Journal of the American Statistical Association*, 84, 408, 986-995.
12. **Hardle, W., Tsybakov, A.B.** (1993). "How sensitive are average derivatives?", *Journal of Econometrics*, 58, no. 1-2, 31-48.
13. **Hastie, T.J., and Tibshirani, R.J.** (1990). *Generalized Additive Models*, Chapman & Hall, London.
14. **Horowitz, J.L.** (1998). *Semiparametric methods in econometrics*, Springer-Verlag, New York.
15. **Horowitz, J.L., Hardle, W.** (1994). "Testing a parametric model against a semiparametric alternative, *Econometric Theory*, 10, 821-848.
16. **Horowitz, J.L., Hardle, W.** (1996). "Direct semiparametric estimation of single-index models with discrete covariates", *Journal of the American Statistical Association*, 91, 436, 1632-1640.
17. **Hristache, M., Juditsky, A., Spokoiny, V.** (2001). "Direct estimation of the index coefficient in single-index model", *The Annals of Statistics*, 29, 3, 595-623.
18. **Ichimura, H.** (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58, 1-2, 71-120.
19. **Klein, R.W., And Spady, R.H.** (1993). "An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 2, 387-421.
20. **Korostelev, A. and Muller, M.** (1995). "Single index models with mixed discrete-continuous explanatory variables", *Discussion paper 26, Humboldt University of Berlin*.
21. **Matzkin, R.** (1992). "Nonparametric and distribution-free estimation of the binary choice and threshold-crossing models", *Econometrica*, 60, 239-270.
22. **Matzkin, R.** (1994). "Restrictions of Economic theory in nonparametric methods", *The Handbook of Econometrics, Vol. IV*, 2524-2559.
23. **McCullagh, P., and Nelder, J.A.** (1989). *Generalized Linear Models*, Chapman & Hall, New York.
24. **Muller, H.-G.** (1984). "Smooth Optimum Kernel Estimators of Densities, Regression Curves, and Models", *The Annals of Statistics*, 12, 2, 766-774.
25. **Muller, M.**, (2004). "Generalized Linear Models", *Handbook of Computational Statistics, Vol. 1*.
26. **Nadaraya, E.A.** (1964). "On estimating regression" , *Theory of probability and its applications*, 10, 186-190.
27. **Nelder, J.A., and Wedderburn, R.W.M.** (1972). "Generalized Linear Models", *Journal of the Royal Statistical Society, A*, 135, 370-384.
28. **Newey, W.K., Stocker, T.M.** (1993). "Efficiency of weighted average derivative estimators and index models", *Econometrica*, 61, 5, 1199-1223.
29. **Powell, J., L.** (1994). "Estimation of Semiparametric Models", *Handbook of Econometric Vol. 4, Ch. 41*.
30. **Powell, J., L., and Stoker, T.M.** (1996). "Optimal bandwidth choice for density-weighted averages", *Econometrica*, 75, 26, 291-316.
31. **Powell, J., L., Stock, J.H., and Stoker, T.M.** (1989). "Semiparametric estimation of index coefficients", *Econometrica*, 57, 6, 1403-1430.
32. **Robinson, P.M.** (1988). "Root-N-Consistent Semiparametric Regression", *Econometrica*, 56, 4, 931-954.

6. Conclusions

This paper considers the main basic idea of the theory of Single-index models. The principal part of this paper is devoted to the estimation of the index coefficients. Four methods have been applied. All considered estimators are root-n consistent and the most have been proven to be asymptotically normal. Theoretically, M-estimator get some very nice properties. Particular, they are efficient and do not require strong assumptions on the distribution of X, contrary to direct methods. In addition to, they supply the bandwidth needed for the nonparametric estimation of the link function. On the other hand, direct estimators are much more easy and fast to compute, since they do not require solving any nonlinear optimization criterion. The simulation study shows that the M-estimators outperform the direct ones. Among the direct estimators, the only one which gives satisfactory results seems to be the weighted average derivative estimator. Once the index coefficients have been estimated, the estimation of the link function is done via standard nonparametric regression methods. The rate of convergence of this estimator is the usual rate of the univariate nonparametric regression.

Bibliography

1. Carroll, R., J., Fan, J., Gijbels, I., and Wand, M., P. (1997). "Generalized Partially Linear Single-Index Models", *Journal of the American Statistical Association*, 92, 477-489.
2. Chen, H. (1988). Convergence Rates for Parametric Components in A partly Linear Model, *The Annals of Statistics*, 16,1, 136-146.
3. Delecroix, M., Hardle, W., Hristache, M. (2002). "Efficient estimation in conditional single-index regression", *Journal of Multivariate analysis*, 86, 2, 213-226.
4. Dobson, A.J. (1990). *An Introduction to Generalized linear Models*, Chapman & Hall, New York.
5. Fan, J.; and Li, Q. (1996). "Consistent model specification test: Omitted variables and semiparametric forms", *Econometrica*, 64, 865-890.
6. Green, P.J., and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, New York.
7. Hardle, W., and Muller, M. (2000). "Multivariate and Semiparametric Kernel Regression", in : M.G. Schimek (Ed.), *Smoothing and Regression. Approaches, computation and Application*, Wiley.
8. Hardle, W., Hall, P., and Ichimura, H. (1993). "Optimal smoothing in single-index models", *The Annals of Statistics*, 21, 1, 157-17800.
9. Hardle, W., Hart, J., Marron, J.S., Tsybakov, A.B. (1992). "Bandwidth choice for average derivative estimation", *Journal of the American Statistical Association*, 87, 717, 218-226.
10. Hardle, W., Muller M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer-Verlag, Berlin, Germany.

$$Y_i = g(X_i^T \beta) + \varepsilon_i,$$

with

$$\begin{aligned} g(x) &= \sin(x), \\ p &= 2, \\ \beta &= (1, 2), \\ \varepsilon_i &\sim N(0, 1), \end{aligned}$$

and

$$X_i \sim N(0, I)$$

The normal distribution of the error term implies that the semiparametric maximum likelihood (SML) exactly amounts to the semiparametric least squares (SLS).

The sample size n is set to be $\{50, 100, 200\}$ with 500 Monte-Carlo replications are drawn in each case. The Nadaraya-Watson estimator is used as a kernel estimator with Epanchnikov kernel.

The following table(1) shows the expectation and the standard deviation of $\hat{\beta}$, for each method.

Estimation Method	n=50		n=100		n=200	
	$E(\hat{\beta})$	$St.Dev.(\hat{\beta})$	$E(\hat{\beta})$	$St.Dev.(\hat{\beta})$	$E(\hat{\beta})$	$St.Dev.(\hat{\beta})$
SLSE	1.69	2.65	2.25	1.16	2.10	0.47
PMLE	1.64	2.46	2.20	1.18	2.11	0.58
ADE	0.84	2.84	1.42	2.51	1.57	3.12
WADE	1.72	2.30	2.07	2.07	2.24	1.04

Table(1): The expectation and the standard deviation of $\hat{\beta}$, using the semiparametric least squares (SLS), Pseudo maximum likelihood (PMLE), average derivative (ADE), and weighted average derivative (WADE).

From Table(1), it is noted that once the sample size grows, the M-estimators (SLSE and PMLE) become clearly better than the direct ones. On the other hand, for small sample size ($n=50$), the WADE estimator seems to be the best estimator with small bias and smallest variance. Nevertheless, this estimator is subject to the curse of dimensionality when p is larger than 2. A surprising feature is that the variance of ADE does not decrease with increasing the sample size n . The method of ADE is numerically unstable, so that no reliable estimation seems to be possible from it when n becomes large. Finally, the SLS method is totally equivalent to PMLE method, it is shown that the first one is slightly better than the latter one, when n is sufficiently large.

$$H_0: E[Y|X=x] = G(x^T\beta) \quad (26)$$

versus

$$H_1: E[Y|X=x] = g(x^T\beta).$$

The link function, G , under H_0 is a known function while the link function, g , under H_1 is an unspecified function.

Specification of the parametric model can be tested by comparing the nonparametric estimate of $E[Y|X^T\beta=v]$ with the parametric fit using the known link G . The test statistic is defined as;

$$T = \sqrt{n} \sum_{i=1}^n w(X_i^T \hat{\beta}) \{Y_i - G(X_i^T \hat{\beta})\} \{\hat{g}_{-i}(X_i^T) - G(X_i^T \hat{\beta})\} \quad (27)$$

where $\hat{g}_{-i}$ is a leave-one-out Nadaraya-Watson estimate for the regression of Y on the estimated index values, h is the bandwidth used in the kernel regression. $w(\cdot)$ is a weight function that downweights extreme observations.

Hardle et al. (2004) showed that under some suitable regularity conditions, the statistic T is asymptotically distributed as a $N(0, \sigma_T^2)$ where σ_T^2 denotes the asymptotic sampling variance of the statistic.

Fan and Li (1996) proposed a consistent test which is based on the analysis of the residuals, with

$$\varepsilon_i = Y_i - g(X_i^T \beta)$$

It is shown that:

$$\begin{aligned} E(\varepsilon_i | X = x_i) &= E(Y_i | X = x_i) - E(g(X_i^T \beta) | X = x_i) \\ &= m(X_i) - g(X_i^T \beta) \end{aligned}$$

which equals zero iff H_0 is true. Hence $E\{[E(\varepsilon_i | X = x_i)]^2\} \geq 0$ and the equality holds only under H_0 . This property will be the basis of the test.

Delecroxi et al. (2004) developed a test which is based on geometrical arguments. They proposed a test statistics which can be built on a measure of the variability of integral averages of the function $m(\cdot)$ over hyper cylinders whose axes are orthogonal to the potential index vector of the model.

5. A simulation Study

The practical performance of different methods of estimating the unknown index vector β will be investigated. The data $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ are independent and drawn from the following model:

$$\begin{aligned}\hat{\beta} &= \hat{J}^{(1)} - \hat{J}^{(0)}, \\ \text{with} \\ \hat{J}^{(0)} &= \sum_{i=0}^{n_0} Y_i^{(0)} (U_{i+1}^{(0)} - U_i^{(0)})^T \gamma, \\ \hat{J}^{(1)} &= \sum_{i=0}^{n_1} Y_i^{(1)} (U_{i+1}^{(1)} - U_i^{(1)})^T \gamma,\end{aligned}\tag{23}$$

where the superscripts (0) and (1) denote the observations from the subsamples according to $T=0$ and $T=1$. The estimator in (23), in the simple case that Y is a binary, is $\sqrt{n}$ -consistent and can be improved for efficiency by a one-step estimator. (See Korostelev and Muller (1995); and Hardle et al. (2004)).

Horowitz and Hardle (1996) extended the above approach to the multivariate-categorical T and an arbitrary distribution of Y .

3.3 Bandwidth Selection

Selecting the bandwidth h is needed since the practical performance of the estimation method depends significantly on it. Delecroix et al. (2003) proposed an empirical rule for selecting h , they introduced the following criterion to be maximized:

$$\hat{S}(\beta, h) = \frac{1}{n} \sum_{i=1}^n \Psi(Y_i, \hat{g}_{-i}(X_i^T \beta))\tag{24}$$

where $\Psi: \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function verifying some mild regularity conditions. (See Delecroix et al. (2003)). The criterion in (24) is maximized with respect to h , i.e.

$$(\hat{\beta}, \hat{h}) = \arg \max_{\beta, h} \hat{S}(\beta, h)\tag{25}$$

It is shown that this method produces a root- n consistent estimator of β and an asymptotically optimal estimator of h , i.e.

$$\hat{h} / h_{opt} \rightarrow 1,$$

where h_{opt} is the theoretical optimal value of the bandwidth.

4. Testing the single-index model (SIM)

Although the unknown nature of g provides a great flexibility for the model, the hypothesis in Eq.(2) remains somewhat restrictive: geometrically, it amounts to say that the regression function is constant when the regressors are varying along any direction orthogonal to β . Horowitz and Hardle (1994) designed a test that considers the following hypotheses:

argument of the function. Eq. (17) shows that δ equals γ up to scale. A one way to estimate δ can be used also to estimate γ up to scale.

Powell et al. (1989) used the density $f(\cdot)$ of U as the weight function, i.e.

$$w(u) = f(u) \quad (18)$$

This estimator is sometimes referred to a density weighted average derivative estimator (DWADE).

For deriving this estimator, the concentration will be on using the weight function in (18). Powell et al. (1989) showed that:

$$\hat{\delta} = -\frac{2}{n} \sum_{i=1}^n \hat{\nabla}_{f_n}(U_i) Y_i, \quad (19)$$

Where ∇_f is estimated by:

$$\hat{\nabla}_f = \left(-\partial_1 \hat{f}_h(u), \dots, -\partial_q \hat{f}_h(u) \right)_r. \quad (20)$$

Here, the $\partial_j \hat{f}_h(u)$ are the partial derivatives of the multivariate kernel density estimator:

$$\partial_j \hat{f}_h(u) = \frac{1}{nh_1 \dots h_q} \sum_{i=1}^n \partial_j K \left(\frac{u_1 - U_{i1}}{h_1}, \dots, \frac{u_q - U_{iq}}{h_q} \right) \quad (21)$$

It is shown that the estimator $\hat{\delta}$ has asymptotically normal distribution under some regularity conditions and it achieves $\sqrt{n}$ -convergence. (See Powell et al. (1989)).

Second: A model with both continuous and discrete variables

Derivatives can only be calculated if the variable under study is continuous. Thus the WADE method fails for discrete regressors. A case with only one dichotomous variable will be introduced to the model. The model in (16) is extended by an additional term as follows:

$$E[Y \setminus U, T] = g(U^T \gamma) + T^T \beta \quad (22)$$

With U the continuous and T the discrete part of the regressors. Simply, that discrete part is supposed as a univariate binary variable T and that Y is binary as well. In this case, the model "split" into two submodels

$$\begin{aligned} E[Y \setminus U, T] &= P[Y = 1 \setminus U, T] = g(U^T \gamma) & \text{if } T=0, \\ E[Y \setminus U, T] &= P[Y = 1 \setminus U, T] = g(U^T \gamma + \beta) & \text{if } T=1. \end{aligned}$$

There are in fact two models to be estimated, one for $T=0$ and one for $T=1$. It is noted that γ will be estimated from the first equation only, and the difference between the two indices is exactly β .

Korostelev and Muller (1995) proposed a very simple estimator. They showed that the binary regressor variable can be estimated by

estimator is asymptotically efficient. Klein and Spady (1993) illustrated this method in the case of binary response models. They showed that if Y being binary, then:

$$g(X\beta) = E[Y \mid X = x] = p[Y = 1 \mid X = x]$$

which lead to the following conditional distribution of Y given X :

$$\ell_{g,\beta}(Y \mid X) = g(X\beta)^Y (1 - g(X\beta))^{1-Y}$$

The estimator of β is thus given by

$$\hat{\beta} = \arg \max_{\beta} \sum_{i=1}^n \{Y_i \log[\hat{g}_{(-i)}(X\beta)] + (1 - Y_i) \log[1 - \hat{g}_{(-i)}(X\beta)]\}.$$

3.2.3 Weighted average derivative estimation

Although their many advantages, M-estimators admit an important drawback: they require solving an intricate optimization problem in a high dimensional space. A different type of estimator will be considered with two advantages:

- (1) Any distributional assumption on Y will not be needed.
- (2) The resulting estimator is direct. i.e. non-iterative.

In spite of slightly worst theoretical properties, direct estimators are highly attractive, as they provide the estimator on an analytic form. Furthermore, the average derivative method is only directly applicable to models with continuous regressors.

The basic idea is to identify β as the average derivative and thus the studied estimator is called average derivative estimator (ADE) or weighted average derivative estimator (WADE). The following notation will be used:

$$X = (U, T)$$

for the regressors, where U is a q -dimensional refers to continuous variables and T is a p -dimensional of discrete or categorical variables. Two models will be considered:

First: A model with a d -dimensional vector of continuous variables only, i.e.,

$$E[Y \mid X] = m(U) = g(U^T \gamma). \quad (16)$$

The vector of weighted average derivatives δ is given by

$$\delta = E\{\nabla_m(U)w(U)\} = E\{\dot{g}(U^T \gamma)w(U)\}, \quad (17)$$

Where $\nabla_m(U) = (\partial_1 m(U), \dots, \partial_q m(U))^T$ is the vector of partial derivatives of $m(\cdot)$, $\dot{g}$ is the derivative of $g(\cdot)$ and $w(\cdot)$ is a weight function, and ∂_j is the partial derivative w.r.t. the j th

density of Y given X depends upon β and g . It is supposed that the conditional density depends on X only through $X\beta$. The likelihood is defined as follows:

$$L_g(\beta) = \prod_{i=1}^n \ell_{g,\beta}(X_i, Y_i) = \prod_{i=1}^n \ell_{g,\beta}(Y_i \setminus X\beta = X_i^T \beta) f(X_i), \quad (11)$$

where $\ell_{g,\beta}(\cdot)$ and $\ell_{g,\beta}(\cdot \setminus \cdot)$ denote to the joint and conditional distributions of X and Y , respectively and f is the marginal density of X . Hence the log-likelihood is:

$$LL_g(\beta) = \sum_{i=1}^n \log \ell_{g,\beta}(Y_i \setminus X\beta = X_i^T \beta) + \sum_{i=1}^n \log f(X_i). \quad (12)$$

The term $\sum_{i=1}^n \log f(X_i)$ does not depend on g and β , so that maximum $LL_g(\beta)$ amounts to maximization $\sum_{i=1}^n \log \ell_{g,\beta}(Y_i \setminus X = X_i)$. If g was known, the maximum likelihood estimator of β should be given by the following maximization problem:

$$\beta^* = \arg \max_{\beta} \sum_{i=1}^n \log \ell_{g,\beta}(Y_i \setminus X\beta = X_i^T \beta). \quad (13)$$

Standard theory of maximum likelihood estimation implies that β^* is $\sqrt{n}$ -consistent, efficient and asymptotically normal under regularity conditions.

Nevertheless, since g is unknown, β^* is not feasible. This problem can be overcome by replacing g in (13) with its Nadaraya-Watson leave-one-out estimator, thus forming a pseudo-likelihood. (See Klein & Spady (1993); Weisberg & Welsh (1994); and Delecroix et al. (2003)).

The pseudo-likelihood estimator (PMLE) is given by:

$$\hat{\beta} = \arg \max_{\beta} \sum_{i=1}^n \log \ell_{\hat{g}(-i),\beta}(Y_i \setminus X\beta = X_i^T \beta) \quad (14)$$

If the conditional distribution of Y given X is not known, it can be estimated nonparametrically as follows:

$$\hat{\ell}_{g,\beta}(Y_i \setminus X\beta = X_i^T \beta) = \frac{\sum_{j \neq i} K\left(\frac{Y_i - Y_j}{h}\right) K\left(\frac{X_i^T \beta - X_j^T \beta}{h}\right)}{\sum_{j \neq i} K\left(\frac{X_i^T \beta - X_j^T \beta}{h}\right)} \quad (15)$$

which is a kernel estimator of the joint distribution of $(X^T \beta, Y)$ divided by the classical kernel estimator of the marginal density of $X^T \beta$. Delecroix et al. (2003) showed that the resulting

$$\frac{1}{n} \sum_{i=1}^n w(X_i) [Y_i - \hat{g}_{(-i)}(X_i^T \beta)]^2 \quad (7)$$

where $\hat{g}_{(-i)}$ is the leave-one-out Nadaraya-Watson estimator of g assuming the parameter β would be known. (See Ichimura (1993)). The estimator $\hat{g}_{(-i)}(X_i^T \beta)$ appearing in criterion (7), is a weighted Nadaraya-watson estimator, becomes

$$\hat{g}_{(-i)}(X_i^T \beta) = \frac{\sum_{j \neq i} w(X_j) K\left(\frac{X_i^T \beta - X_j^T \beta}{h}\right)}{\sum_{j \neq i} w(X_j) K\left(\frac{X_i^T \beta - X_j^T \beta}{h}\right)} Y_j, \quad (8)$$

With h denotes a bandwidth and K_h a scaled (compact support) kernel.

The choice of the weight function w affects the efficiency of the estimator. Newey and Stocker (1993) found the efficiency bound for semiparametric models. In a single-index context, SLS estimator achieves this bound if

$$w(x) = \frac{1}{\sigma^2(x)} \quad (9)$$

If $\sigma^2(x)$ is unknown, a consistent estimator $\hat{\sigma}^2(x)$ has to be used in (9). Such an estimator can be obtained using the following two-steps procedure:

- First: estimate β by $\hat{\beta}_1$, the minimizer of the unweighted version of (7), which is a root-n but inefficient estimator.
- Second: Let e_i be the i th residual from the estimated model, i.e.

$$e_i = Y_i - \hat{g}_{(-i)}(X_i^T \hat{\beta}_1) \quad (10)$$

And set $\hat{\sigma}^2(x)$ equal to a nonparametric estimator (e.g. the Nadarya-Watson estimator) of the mean regression of e_i^2 on x . It is noted that if a function V is known such that

$$\sigma^2(x) = V[g(X^T \beta)],$$

$\hat{\beta}_1$ will be computed first by taking the weight function w to be identically equal to 1 as above. Then in Eq.(8) $w(X_j)$ is replaced by $\{V[\hat{g}(X^T \hat{\beta})]\}^{-1}$ in order to compute $\hat{\beta}$. (See Horowitz (1996)).

3.2.2 Semiparametric maximum likelihood (SML)

Another optimization method is based on the parametric maximum likelihood methods. In the single-index context, the joint distribution of X and Y and the conditional

$$K_h \frac{1}{nh} \sum_{i=1}^n \left(\frac{u - \hat{\eta}_i}{h} \right). \quad (4)$$

It will be shown that β can be estimated with an $n^{-\frac{1}{2}}$ rate of convergence in probability, i.e. there exists estimator $\hat{\beta}$ such that

$$(\hat{\beta} - \beta) = O_p \left(n^{-\frac{1}{2}} \right), \quad (5)$$

which is the typical rate of convergence for parametric estimators. Besides, it is well known that no nonparametric estimator of the regression function can achieve this rate. The convergence of $\hat{\beta}$ is thus faster than the fastest possible rate of any nonparametric estimator of g . Hence, the estimation of g is direct via standard methods once an estimator of β is known. It is noted, however, that an important attraction of g , is that the model remains flexible enough, but as this nonparametric estimation is done on an univariate index η , the curse of dimensionality is avoided.

3.2 Estimation of β

Many methods of estimating β have been proposed in the literature. Estimators of β can be classified according to whether they require solving nonlinear optimization problem (M-estimator), i.e. indirect estimation or direct estimation. There are two essentially different approaches for this purpose:

- I. An iterative approximation of β using semiparametric least squares (SLS) or pseudo maximum likelihood estimation (PMLE).
- II. A direct (non-iterative) estimation of β using the average derivative of the regression function.

Two important particular M-estimators will be examined, which are generalizations of the parametric least squares and maximum likelihood estimators.

3.2.1 Semiparametric least squares (SLS)

As in the parametric least squares problem, the idea is to minimize the mean square distance between the observed values Y_i and the values given by the model $g(x_i^T \beta)$. If g was known, the classical least squares estimator is given by

$$\beta^* = \arg \min \frac{1}{n} \sum_{i=1}^n w(X_i) [Y_i - g(X_i^T \beta)]^2, \quad (6)$$

where $w(\cdot)$ is an appropriate weight function, and $\arg \min (\cdot)$ means that the function $(\cdot)$ has a unique minimum value at β^* .

To employ the nonparametric technique to substitute the (inner) unknown conditional expectation, a univariate consistent smoother will be used. For simplicity, a Nadaraya-Watson estimator will be used. In the single-index context, the least squares criterion to becomes:

The SIM in (2) is more restrictive than a fully nonparametric model, so it is important to show if the index $x^T \beta$ and $g(\cdot)$ are identified and estimable nonparametrically. It is shown that the index and the function $g(\cdot)$ is restricted to a suitably small class of functions.

Matzkin (1992, 1994) gave some examples of suitable classes of functions. Each contains some single-index models but none contains all. Thus the single index specification contains models that are not within Matzkin's classes of identifiable, nonparametric, structural models that are not single-index models. Similarly, there are identifiable, nonparametric, structural models that are not single-index models. Therefore, Matzkin's classes of identifiable, nonparametric, structural models are neither more nor less general than the class of single-index models.

3. Estimation of the single-index model(SIM)

It is assumed that $g(\cdot)$ and β of the single-index model (SIM) are identified. The strategy of estimating the SIM is based on the following procedure:

- a- Estimation of β by $\hat{\beta}$.
- b- Computing the index values $\hat{\eta} = X\hat{\beta}$.
- c- Estimating the link function $g(\cdot)$.

The estimation of the link function $g(\cdot)$ will be presented first, which is based on $\hat{\eta} = X\hat{\beta}$.

3.1 Estimation of $g(\cdot)$

Suppose that β is known. Then g can be estimated using a univariate nonparametric method for the regression of Y on $\eta = X^T \beta$, various methods are proposed. (See Hardle (1990); and Linton(1994)). This paper uses kernel estimators. Although it is well known that it is not the more efficient one, the Nadaraya-Watson kernel estimator is used in many situations because of its easiness of implementation and interpretation and its mathematical tractability. (See Nadaraya (1964); and Watson (1964)).

Let $\{(X_i, Y_i), i=1,2,\dots,n\}$ be the sample and define $\hat{\eta}_i = X_i^T \hat{\beta}$. Let K be the kernel function usually taken to be a bounded symmetric probability density and h is a bandwidth, i.e. a smoothing parameter. Then the Nadaraya-Watson estimator of the regression function is

$$\hat{g}_h(u) = \frac{\sum_{i=1}^n K_h(u - \hat{\eta}_i) Y_i}{\sum_{i=1}^n K_h(u - \hat{\eta}_i)}, \quad (3)$$

where

with increasing dimension of X (the so-called curse of dimensionality), difficulty in representing relationships graphically, and the inability to make predictions at points X outside the support of the observed X values. (see Carroll, et al. (1997); and Horowitz (1998)).

Semiparametric models propose a mix of both approaches, while permits to compensate for their respective drawbacks. They are characterized by a twofold parameterization, say θ and γ , where θ lies in a finite-dimensional space Θ and γ lies in an infinite-dimensional space Γ . One of the most popular semiparametric model is the SIM. SIMs relax some of the restrictive assumptions of parametric models, in addition to achieve dimension reduction and, thereby, large estimation precision of $E[Y|X=x]$ when X is multidimensional. Finally, SIMs are often easy to compute, and their results are easy to interpret.

In section(2) a formal definition of SIMs and identification conditions are given. Section (3) presents the main estimation procedures in the SIM context, while section (4) considers testing the single-index hypothesis. A simulation study that permits to compare the practical performances of these procedures is presented in section (5). Finally, section (6) introduces the conclusions and discussion on this work.

2. Definition and identification conditions of a single-index model(SIM)

In a single-index model, the conditional mean function $E[Y|X=x]$ has the form:

$$E(Y | X = x) = g(x^T \beta), \quad (2)$$

where Y is a scalar random variable, X is a $p \times 1$ random vector, and β is an unknown $p \times 1$ constant vector and g is an unknown function. The inferential problem in (2) is to estimate β and g from observations of (Y, X) .

It is noted that the model in (2) is indeed a semiparametric since β lies in a finite-dimensional space, and the function g belongs to a functional space so that it can be seen as an infinite dimensional set of parameters. (See Ichimura (1993)).

Restriction must be imposed in order to make β and g uniquely determined by the population distribution of (X, Y) . Ichimura (1993) showed that β and g are identified if:

- g is differential and not constant on the support of $x^T \beta$.
- X admits at least one continuously distributed component.
- The support of X is not contained in any proper linear subspace of R^p .
- $\beta \in \beta$, with $\beta = \{\beta \in R^p : \beta^{(1)} = 1\}$.

If some covariates are discrete, then two other conditions are needed:

- Varying the values of the discrete components must not divide the support of $x^T \beta$ into disjoint subsets.
- g must be a periodic function.

On the Single-Index Model Theory

Dr. Magda Mohamed M. Haggag

*Department of Statistics, Mathematics, and Insurance
Faculty of Commerce, Alexandria University (Damanhour Branch)*

Email: magmhag@yahoo.com

Abstract

One of the most popular semiparametric regression models in literature is certainly the single-index model (SIM). Single-index models generalize linear regression, where the link function is kept unknown and has to be estimated via nonparametric techniques. In this paper a complete summary of the theory of the SIM is introduced. Identification conditions, estimation of the link function, estimation of the index and goodness of fit tests are presented. Also, a simulation study permitting to compare the practical performances of different estimators of the index is done.

Key Words:, average derivative estimator, generalized linear model, kernel smoothing, M-estimators, semiparametric least squares, single-index model, pseudo-likelihood.

1. Introduction

A single index model (SIM) is a class of semiparametric models that can be considered as a nonparametric approach to generalized linear model (GLM), i.e. to generalize the known link function in generalized linear regression to an unknown, nonparametric link function. (See, Ichimura (1993); Ruppart et al. (2003); Yatchew (2003) ; and Hardle et al. (2004)).

Consider the model,

$$Y = m(X) + \varepsilon, \quad (1)$$

where Y is a scalar output, X is a p -dimensional vector of regressors, $m(\cdot)$ is a function from a subset of R^p to the real line, and ε is an univariate random disturbance. The interest will be in estimating the conditional mean function $E(Y|X=x)$ from a sample $\{(X_i, Y_i): i=1, \dots, n\}$. This function is also known as the regression function, is equal to $m(x)$ provided that $E(\varepsilon|X=x)$ is zero, as it is assumed in general.

Several main approaches can be considered to deal with this problem. First, one can adopt the parametric modeling, assuming that $m(\cdot)$ and the distribution of ε are known up to a finite set of parameters. These parameters can be estimated using a least squares or by a maximum likelihood method. While the parametric modeling is easy in computation and interpretation of the results, it admits an important drawback, which is the lack of flexibility. Also, the nonparametric approach has some drawbacks which include decreasing precision

References

- 1) Amemiya, T.(1985). "Advanced Econometrics", first edition (Cambridge: Harvard University Press).
- 2) Andrews, G., E., Askey, R. and Roy, R.(1999)."Special Functions", first edition, Cambridge University Press, Cambridge.
- 3) Bell, W.W., (1968), "Special Functions for Scientists and Engineers." D.Var Nostrand Company Ltd Princeton, New Jersey, Toronto, Melbourne.
- 4) Cramer, J.S. (1987), Mean and Variance of R in Small and Moderate Samples, Journal of Econometrics, 35,253-266.
- 5) Greene, W.H. (2000)." Econometric Analysis ",(Upper Saddle River. NJ: Prentice Hall).
- 6) Judge,G., Hill, C., Griffiths, W., Lee, T. and Lulkepol, H. (1982)." Introduction to the Theory and Practices of Econometrics", (New York: John Wiley and Sons).
- 7) Nagar, A.L. (1959), The bias and moment matrix of the general K-class estimators of the parameters in structural equations, Econometrica, 58, 145-163.
- 8) Philips, P.C.B (1977), A general theorem in the theory of asymptotic expansion as approximations to the finite sample distributions of econometric estimators, Econometrica, 45b, 1517-1534.
- 9) Sawa, T. (1972). "Finite Sample Properties of k- Class Estimators", Econometrica, 40:653-80.
- 10) Theil, H. (1971),"Principles of Econometrics", (New York: John Wiley & Sons).

To sum up, R^2 has an upward bias which can be substantial in small samples, but it is anyhow very unreliable, even at moderate sample sizes, because of its dispersion. With less than fifty observations or so there is little point in quoting R^2 at all and once we are beyond such numbers the bias has virtually disappeared

Fig. (3) and (4) shows how the standard error varies with ϕ , m and k , the effect of ϕ is particularly strong. For a standard error of R^2 of 0.03 or less we must have at least twenty observation if $\phi = 0.9$. Such high values of the true correlation coefficient are probably exceptional, but with ϕ at 0.667, nearly two hundred observations are needed to reduce the standard error to 0.03.

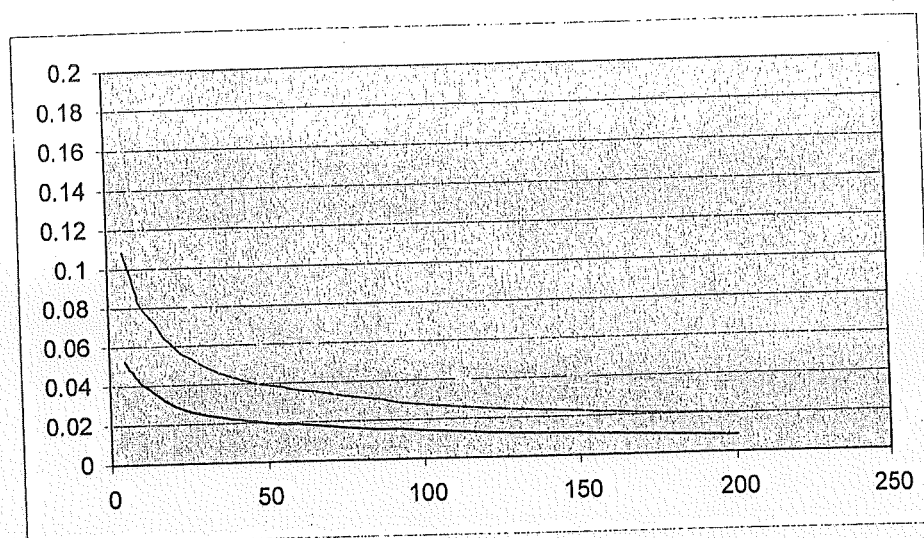


Fig. (3) Standard deviation of R^2 as a function of T for $k = 2$ and selected values of ϕ

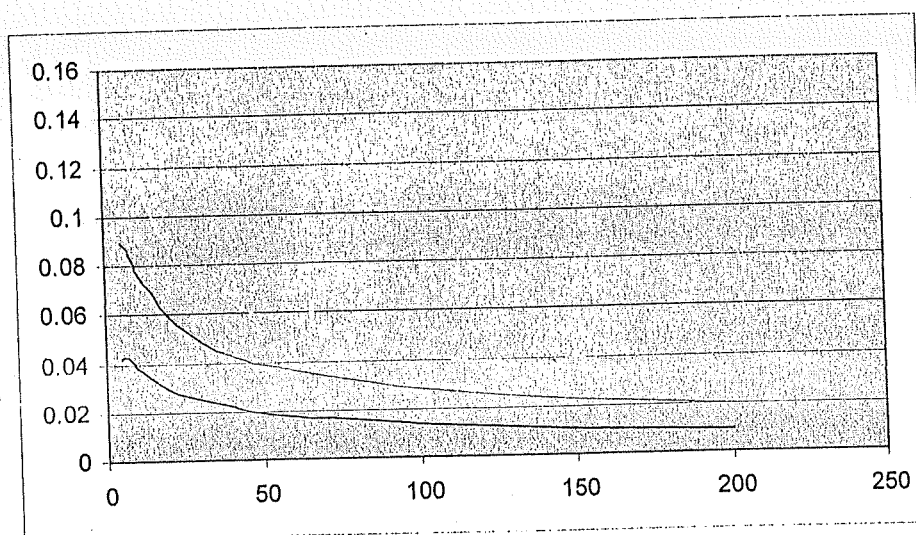


Fig. (3) Standard deviation of R^2 as a function of T for $k = 3$ and selected values of ϕ

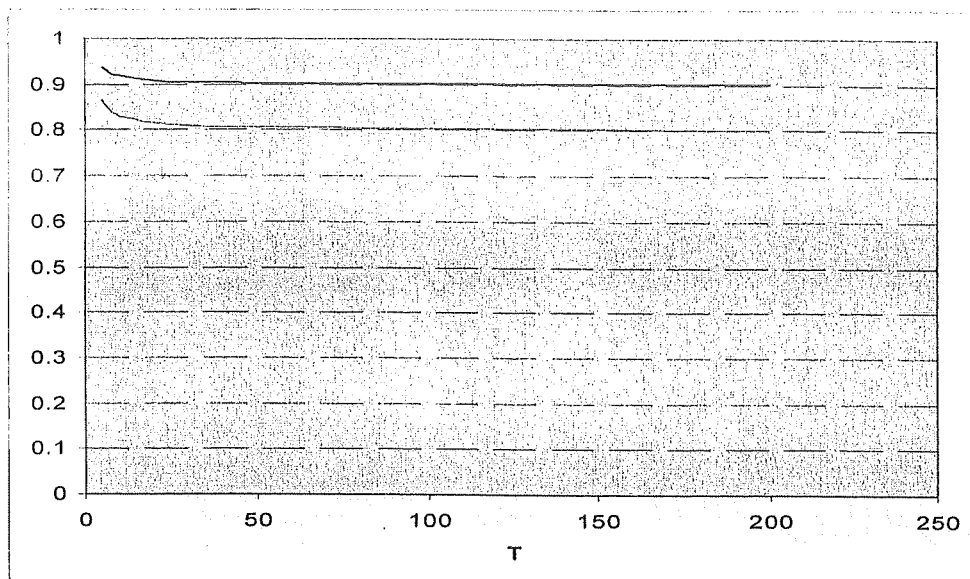


Fig. (1) Expected value of R^2 as a function of T for $k=2$ and selected values of ϕ

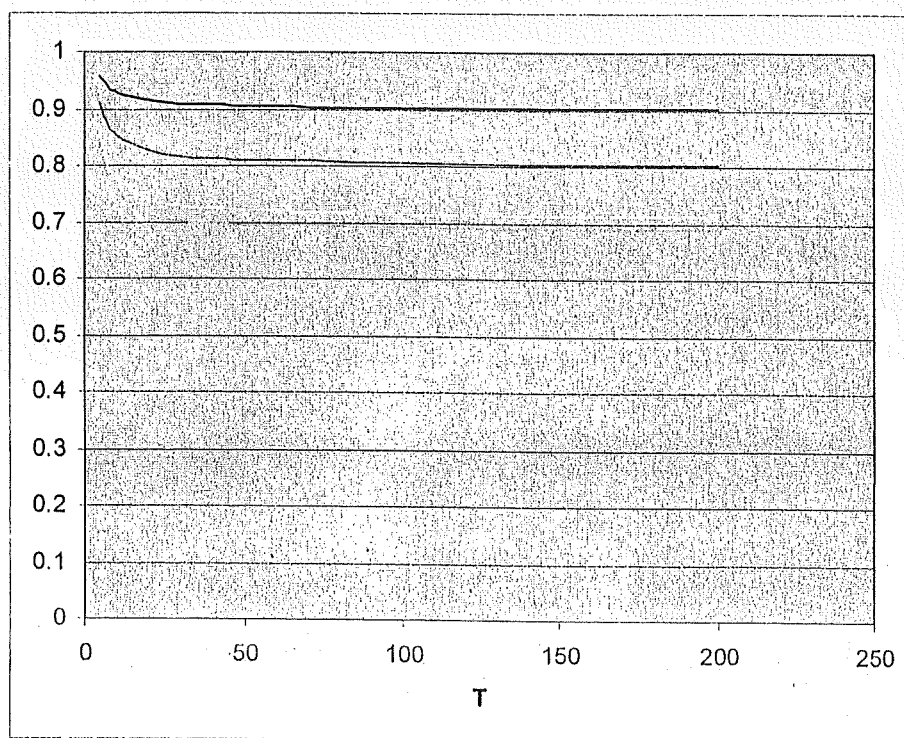


Fig. (2) Expected value of R^2 as a function of T for $k=3$ and selected values of ϕ

Table (2): Standard deviation of R^2 for selected values of k , of ϕ , and of T

$\phi =$		$k=2$				$k=3$			
		0.9	0.8	0.7	0.667	0.9	0.8	0.7	0.667
T=	5	0.052	0.109	0.164	0.18	0.042	0.089	0.136	0.15
	6	0.049	0.102	0.151	0.163	0.043	0.088	0.132	0.144
	7	0.047	0.096	0.14	0.151	0.042	0.086	0.127	0.138
	8	0.045	0.091	0.132	0.142	0.041	0.083	0.122	0.133
	9	0.043	0.086	0.126	0.136	0.04	0.08	0.118	0.128
	10	0.041	0.082	0.12	0.13	0.038	0.077	0.114	0.124
	20	0.03	0.06	0.089	0.099	0.029	0.059	0.087	0.097
	30	0.025	0.05	0.075	0.084	0.025	0.049	0.074	0.083
	40	0.022	0.043	0.065	0.074	0.022	0.043	0.065	0.073
	50	0.02	0.039	0.059	0.067	0.019	0.039	0.059	0.067
	100	0.014	0.028	0.043	0.049	0.014	0.028	0.042	0.048
	150	0.011	0.023	0.035	0.04	0.011	0.023	0.035	0.04
	200	0.009	0.02	0.03	0.035	0.009	0.02	0.03	0.035

Fig. (1) and (2) shows that $E(R^2)$ converges rather quickly to ϕ from above. R^2 thus has a definite upwards bias which is however rapidly reduced as the sample size increases. Very roughly the bias is about 0.03 or less with twenty observations when we have one regressor ($k=2$), or with thirty to forty observations with two regressors. But this is a rough indication only, as the exact values vary with ϕ .

3. Numerical Computation of the Bias and the Standard Deviation of R^2

In this section we shall exhibit the result of numerical computation of the bias and the standard deviation of R^2 . Observing these numerical results, we shall be able to get a clear picture of the behavior of the first two moments for various values of ϕ , of m , and of k

Table (1): Mean of R^2 for selected values of k , of ϕ , and of T

$\phi =$		$k=2$				$k=3$			
		0.9	0.8	0.7	0.667	0.9	0.8	0.7	0.667
T=	5	0.936	0.865	0.787	0.76	0.958	0.91	0.858	0.84
	6	0.93	0.853	0.771	0.743	0.947	0.89	0.828	0.807
	7	0.926	0.845	0.76	0.731	0.94	0.876	0.808	0.784
	8	0.922	0.839	0.752	0.722	0.935	0.866	0.793	0.768
	9	0.92	0.835	0.745	0.715	0.931	0.858	0.782	0.756
	10	0.918	0.831	0.741	0.711	0.928	0.852	0.773	0.747
	20	0.909	0.815	0.722	0.693	0.914	0.826	0.738	0.71
	30	0.906	0.81	0.717	0.69	0.909	0.817	0.727	0.701
	40	0.904	0.808	0.715	0.688	0.907	0.813	0.723	0.697
	50	0.903	0.806	0.714	0.688	0.905	0.81	0.72	0.694
	100	0.902	0.804	0.712	0.687	0.903	0.806	0.715	0.691
	150	0.901	0.803	0.711	0.687	0.902	0.804	0.713	0.69
	200	0.901	0.802	0.711	0.687	0.901	0.803	0.712	0.689

Lemma 1:

Hence, the asymptotic variance of R^2 to the order δ_1^{-4} is given by

$$\begin{aligned}\bar{V}(R^2) = & (n-m)\delta_1^{-2} - 2(n-1)(n-m)\delta_1^{-3} \\ & + [3(n-2) - (n-m)](n-1)(n-m)\delta_1^{-4} - o(\delta_1^{-5})\end{aligned}\quad (11)$$

Lemma 2:

From the connection between R^2 and $\bar{R}^2$ we have

$$\bar{R}^2 = 1 - \frac{T-1}{T-K}(1-R^2)$$

Then,

$$\bar{E}(\bar{R}^2) = 1 - \frac{T-1}{T-K}(1-\bar{E}(R^2))$$

from equation (9) we get, the large δ asymptotic mean to the order δ_1^{-4}

$$\bar{E}(\bar{R}^2) = 1 - \frac{T-1}{T-K} \left[\begin{aligned} & (n-m)\delta_1^{-1} + (n-m)(n-1)\delta_1^{-2} \\ & + (n-m)(n-2)(n-1)\delta_1^{-3} \\ & - (n-m)(n-3)(n-2)(n-1)\delta_1^{-4} + o(\delta_1^{-5}) \end{aligned} \right] \quad (12)$$

And the large δ asymptotic variance to the order δ_1^{-4} is

$$\begin{aligned}\bar{V}(\bar{R}^2) = & 1 - \frac{T-1}{T-K}(1-\bar{V}(R^2)) \\ \bar{V}(\bar{R}^2) = & 1 - \frac{T-1}{T-K} \left[\begin{aligned} & 1 - (n-m)\delta_1^{-2} + 2(n-1)(n-m)\delta_1^{-3} \\ & - [3(n-2) - (n-m)](n-1)(n-m)\delta_1^{-4} + o(\delta_1^{-5}) \end{aligned} \right] \end{aligned} \quad (13)$$

• The proofs of these theorems and lemmas can be distained from the author upon request

and the residual vector is

$$\begin{aligned}\hat{\varepsilon} &= Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y \\ &= [I - X(X'X)^{-1}X']Y = MY\end{aligned}\quad (5)$$

$$\text{where } M = I - X(X'X)^{-1}X'$$

then

$$\varepsilon'\varepsilon = Y'MMY = Y'MY \quad (6)$$

from (4) and (6) we can rewrite R^2 as follow

$$R^2 = 1 - \frac{Y'MY}{Y'NY} = \frac{Y'NY - Y'MY}{Y'NY} = \frac{Y'(N-M)Y}{Y'NY} \quad (7)$$

$$R^2 = \frac{Y'N_2Y}{Y'NY} \quad (8)$$

$$\text{where } N_2 = N - M, \quad N = I - \frac{1}{T}LL', \quad M = I - X(X'X)^{-1}X'$$

Theorem 1:

The approximate first order moment (mean) of R^2 is given by

$$\begin{aligned}\bar{E}(R^2) &= 1 - (n-m)\delta_1^{-1} + (n-m)(n-1)\delta_1^{-2} - (n-m)(n-2)(n-1)\delta_1^{-3} \\ &\quad + (n-m)(n-3)(n-2)(n-1)\delta_1^{-4} - o(\delta_1^{-5})\end{aligned}\quad (9)$$

where

$$\delta_1 = \frac{1}{2}\mu'N\mu, \quad m = \frac{K-1}{2}, \quad n = \frac{T-1}{2}$$

Theorem 2:

The approximate second order moment of R^2 is given by

$$\begin{aligned}\bar{E}(R^2)^2 &= 1 - 2(n-m)\delta_1^{-1} + ((n-m) + (2n-1))(n-m)\delta_1^{-2} \\ &\quad - (2(n-m) + (2n-2))(n-1)(n-m)\delta_1^{-3} \\ &\quad + ((3n-6)(n-m) + (2n-3)(n-2))(n-1)(n-m)\delta_1^{-4} + o(\delta_1^{-5})\end{aligned}\quad (10)$$

where

$$\delta_1 = \frac{1}{2}\mu'N\mu, \quad m = \frac{K-1}{2}, \quad n = \frac{T-1}{2}$$

In view of the above problems with the exact results, obtaining the approximate moments have been popular in econometrics since they are simpler to derive, provide simpler expression, easier to calculate, and often provide useful inferences. Further, the approximations that will be presented and used in the present thesis are useful for the moderately large samples and they lie between the often-unknown exact sample results, and infinite sample limiting results, which are routinely discussed in textbooks and used in applied work even though they give poor results for small and moderate large samples.

The objective of the thesis is to derive easily computable expressions for the mean and variance of R^2 in the standard linear regression model. To put this sample statistic in proper perspective we shall derive its mean and variance for various sample sizes under the standard assumptions of econometric theory. This means that the regressor variables are regarded as given, non random constant. In this respect the model differs from the classical treatment of correlation in the setting of a multivariate Normal distribution, and the result differ too.

2. The Large- δ Asymptotic Mean and Variance for R^2 and $\bar{R}^2$

We consider the standard linear regression model with Normal disturbances as given in any textbook of econometrics.

$$Y = X\hat{\beta} + \hat{\varepsilon}, \quad \varepsilon \sim MN(0, \sigma^2 I) \quad (1)$$

$$\text{and} \quad Y \sim MN(X\beta, \sigma^2 I) \quad (2)$$

$$\text{let} \quad Y^* = \frac{Y}{\sigma} \quad (3)$$

$$E(Y) = E\left(\frac{Y}{\sigma}\right) = \frac{1}{\sigma} X\beta = \mu \quad \text{and} \quad \text{var}(Y^*) = \text{var}\left(\frac{Y}{\sigma}\right) = \frac{1}{\sigma^2} \cdot \sigma^2 I = I$$

so, we can say that,

$$Y^* \sim MN(\mu, I) \quad \text{where} \quad Y^* = \frac{Y}{\sigma}, \quad \mu = \frac{1}{\sigma} X\beta \quad (4)$$

the familiar measure of goodness of fit,

$$R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'NY} \quad , \text{ where } N = I - \frac{1}{T} LL'$$

The Small Sample Properties of the Coefficient of Determination and Its Adjusted Version in Linear Regression Models

Ghazal Abdel-Aziz Amer, Alaa Abd-Elaziz, Soha Othman

ABSTRACT

In this chapter we consider the standard linear regression model with normal disturbances as given in any textbook of econometrics, but we employ a slightly non standard presentation, and so, we get a non standard presentation of R^2 . The objective of the thesis is to derive easily computable expressions for the mean and variance of R^2 in the standard linear regression model. To put this sample statistic in proper perspective we shall derive its mean and variance for various sample sizes under the standard assumptions of econometric theory. This means that the regressor variables are regarded as given, non random constant. In this respect the model differs from the classical treatment of correlation in the setting of a multivariate Normal distribution, and the result differ too.

Key word: The coefficient of determination,

1. Introduction:

There are some techniques for obtaining the exact moments of Econometric estimators and test statistics. These techniques require the specification of the density of the data vector Y , for example, normal, and they provide the results, which hold for any size of the data. The exact results are, however, difficult to derive under a general non-normal framework. Even under a specified density the exact expressions for the moments in many situations are sufficiently complicated and do not lend themselves to further algebraic manipulation for deducing any clear inference. In some situation, especially in the commonly used empirical econometric models such as dynamic models, probit models, the evaluations of exact expressions may be extremely difficult or cannot be derived by the presently known mathematical and statistical tools.

APPENDIX

Table (1): Definition of all candidate series

Variable (Short name)	Variable Long name and units	Sector	Construction details of the series	Information Source
con_gdp	GDP of Construction activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
elec_gdp	GDP of Electricity activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
man_gdp	GDP of Manufacturing activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
tour_gdp	GDP of Tourism activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
pet_gdp	GDP of Petroleum activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
trans_gdp	GDP of Transportation activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
sues_gdp	GDP of Seuz Canal activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
agri_gdp	GDP of Agriculture activity	Real	Quarterly series is constructed by using Proportional Denton (D-4) method	C.B.E & Ministry of Planning
reserr	Real reserves	Financial	Real reserves are obtained by taking the net international reserves at the end of period in million of US dollars, as published by C.B.E, and deflating that number by the consumer price indices that are constructed based on the inflation rates published by CAPMAS	C.B.E
M1real	Real M1	Financial	C.B.E series named money supply and deflated by the consumer price index	C.B.E
quasimr	Real Quasi money	Financial	C.B.E series named quasi money and deflated by the consumer price index	C.B.E
totdepr	Real total deposits	Financial	Total deposits are computed adding up the demand deposits and Time, Savings and Foreign currency deposits series that C.B.E publishes in its publications	C.B.E
rcredps	Real credit to the private sector	Financial	The real credit to the private sector series is constructed using data published by the Central Bank of Egypt and then deflated using the consumer price index	C.B.E
M2real	Real M2	Financial	Sum of real M1 and real quasi money	C.B.E
Govt_reven	Real total revenue of budget sector	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance
gs_taxr	Real taxes on goods and services	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance
income_taxr	Real income tax revenues	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance
cur_expr	Real total current expenditures of budget sector	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance
cons_expr	Real total consumption expenditures including wages, salaries and purchases of goods and services	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance
cap_expr	Real total capital expenditures	Government	Published nominal series are deflated using the consumer price index	C.B.E & Ministry of Finance

APPENDIX

Table (1): Definition of all candidate series (cont.)

Variable (Short name)	Variable Long name and units	Sector	Construction details of the series	Information Source
imports	Volume of imports index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
cons_imports	Volume of imports of consumption goods index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
inter_imports	Volume of imports of intermediate goods index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
cap_imports	Volume of imports of capital goods index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
fuel_imports	Volume of fuel imports index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
pp_imports	Volume of imports of Primary Products index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
exports	Volume of Exports index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
pp_exports	Volume of Exports of primary products index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
sm_exports	Volume of Exports of semi- manufactured products index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
mp_exports	Volume of Exports of manufactured products index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF
fuel_exports	Volume of fuel exports index	External	The index is constructed using data published by C.B.E,commodity price data pinkshets published by the World Bank and commodity prices published in IFS by IMF	C.B.E,World Bank and IMF

- [11] Harm B. and Michael F., 2003, "In Search of Leading Indicators of Economic Activity in Germany," Department of Economics, Hamburg University, Germany.
- [12] Joannes M. and Tahsin S. Sedik, 2003, "In Estimating Indexes of Coincident and Leading Indicators: An Application to Jordan," IMF Working Paper, WP/03/170.
- [13] Koopmans, J. T., 1947, "Measuring Without Theory," Review of Economics and Statistics, Vol. 29, pp. 997-1028.
- [14] Layton, AP, 1987, "A Spectral Analysis of Australia's Leading and Coincident Indices of Cyclical Economic Growth," The Australian Economic Review, 77, PP 39-45.
- [15] Mall, O.P., 1999, "Composite Index of Leading Indicators for Business Cycles in India," Reserve Bank of India Occasional Papers, Vol. 20, No. 3, pp. 373-414.
- [16] Moore, Geoffrey H. and Julius Shiskin, 1967, "Indicators of Business Expansions and Contractions," NBER Occasional Paper, No. 103, Cambridge, Massachusetts: National Bureau of Economic Research.
- [17] Organization for Economic Cooperation and Development, 1998, "Composite Leading Indicators: a Tool for Short-Term Analysis," Available via the Internet: <http://www.oecd.org/pdf/M00003000/M00003226.pdf>.
- [18] Rand, John, and Finn Tarp, 2001, "Business Cycles in Developing Countries: Are They Different?" CREDIT Research Paper, No. 01/21.
- [19] Simone, Alejandro, 2001, "In Search of Coincident and Leading Indicators of Economic Activity in Argentina," IMF Working Paper, WP/01/30.
- [20] Stock, James H., Mark W. Watson, 1989, "New Indices of Coincident and Leading Economic Indicators," Macroeconomics Annual, National Bureau of Economic Research (U.S.).
- [21] Stock, J. and Watson, M., 1988a, "A new Approach to Leading Economic Indicators," Harvard University, Kennedy School of Government.

REFERENCES

- [1] Agenor, Pierre-Richard, C. John McDermott, and Eswar S. Prasad, 2000, "Macroeconomic Fluctuations in Developing Countries: Some Stylized Facts," *The World Bank Economic Review*, Vol.14, No.2, pp. 251-85.
- [2] Andrea B. and Graham V., 2000, "Forecasting Australian Economic Activity Using Leading Indicators," *Economic Research Department, Reserve Bank of Australia*.
- [3] Burns, Arthur. F., and Wesley C. Mitchell, 1946, "Measuring Business Cycles," *NBER Working Paper* No. 2, Cambridge, Massachusetts: National Bureau of Economic Research.
- [4] De Leeuw, Frank, May 1998, "Toward a Theory of Leading Indicators," United States Bureau of Economic Analysis, *Discussion Paper* No. 24:1-(54) Rev. May 1988 (U.S.).
- [5] Dias, Francisco Craveiro, January 1994, "A composite Coincident Indicator for the Portuguese Economy," *Banco de Portugal, Estudos e Documentos de Trabalho Working Papers*, No. 18-93:1-20(Portugal).
- [6] Dua, Pami and Anirvan Banerji, 1999, "An Index of Coincident Economic Indicators for the Indian Economy," *Journal of Quantitative Economics*, Vol.15, PP.177-201.
- [7] Dua, Pami and Anirvan Banerji, 2001, "A Leading Index for the Indian Economy," Available via the Internet: <http://www.cdedse.org/papers/work90.pdf>.
- [8] Emerson, Rebecca A., David F. Hendry, July 1996, "An Evaluation of Forecasting Using Leading Indicators," *Journal of Forecasting*, No.15, PP.271-91 (UK).
- [9] Gorton, Gary, 1984, "Measuring Economic Activity with the Indexes of Leading, Lagging, and Coincident Indicators," *Federal Reserve Bank of Philadelphia Working Papers*, No. 84-6:1-22(U.S.).
- [10] Hamilton, James D., Gabriel Perez-Quiros, January 1996, "What Do the Leading Indicators Lead? ", *Journal of Business*, No. 69:27-49 (U.S.).

leading indicator could be found. The four indicators were selected among other possibilities according to their fit, availability of the series required to construct them and their consistency with economic theory.

From an econometric standpoint, the four indicators produce satisfactory one step ahead forecasts in a root mean squared error sense or mean absolute deviation, given the limited information on which they are based. All models proved a high capacity to predict accurately the sign of 1 and 4 quarter growth rates. From an economic standpoint, the variables in the models' cointegrating vectors and their signs are consistent with what many economists believe about the Egyptian economy. This enhances the confidence in the econometric results and shows another aspect of the search of coincident and leading indicators exercise. Looking for coincident and leading indicators of economic activity can be a useful exercise to confirm the basic hypothesis regarding the relevant source of economic cycles in an economy.

IV.4 Conclusion Regarding the Econometric Properties of the Selected Models

The main conclusions are the following:

- All the presented models have a good fit and do not seem to suffer from co-linearity problems.
- All models pass successfully the misspecification tests.
- All models pass successfully the ex-post forecasting accuracy tests.
- Model 1 produces more precise ex-ante forecasts than the other 3 models. This highlights the importance of a good fit in forecasting given that the rank in ex-ante forecasting precision matched exactly the rank in fit.
- All models produce forecasts that are not biased. This is consistent with the remark that the four models pass successfully the autocorrelation tests. This may indicate that there is no need to correct the prediction either upwards or downwards to improve the forecasting performance of these models.

V. CONCLUSIONS

The two most commonly used methodologies to construct coincident and leading indicators are the NBER and Stock – Watson (1989) (SW) approach. The former is based on the work by Moor and Shiskin (1967), while the later adds an econometric foundation to the NBER-DIC approach. Both of them rely on an abstract concept of economic activity, which is referred as "the state of the economy". The state of the economy is an unobservable variable that must be estimated out of several available series. The estimate of the unobservable state of economy is called a composite index of coincident indicators, and is a measure of economic activity used by these approaches. Leading indicators aim to forecast the behavior of this composite index of coincident indicators.

NBER and SW methodologies are subject to important and well known criticisms. The NBER-DIC methodology has been mainly criticized because the weights of the series in the indicators are calculated in an arbitrary fashion. The absence of econometric techniques in the estimation of the weights makes it likely that the information in a given dataset is not efficiently used. The new version, SW, has overcome this criticism but the data requirements are typically too high for a developing economy such as Egypt.

In order to address these issues, a general methodology to construct indicators that deals with the low number of observations available in this case was proposed. Applying the methodology to the limited quarterly data available in Egypt, four coincident composite indicators could be constructed but no reliable

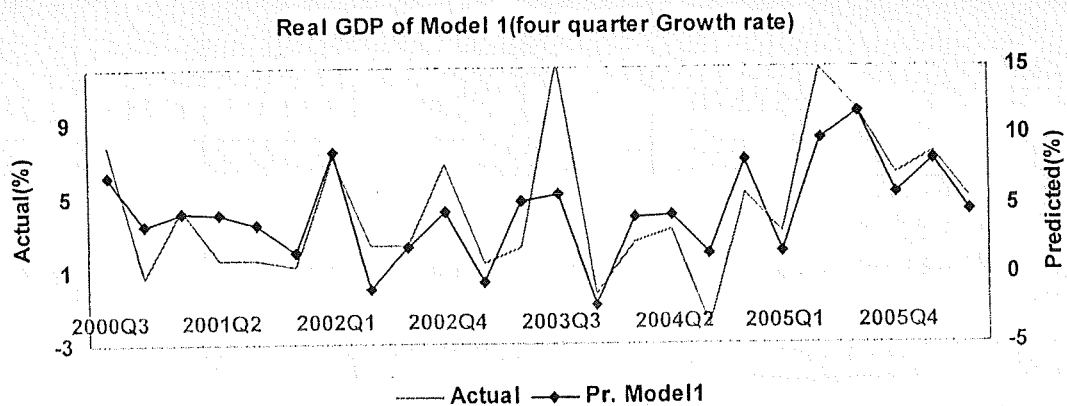
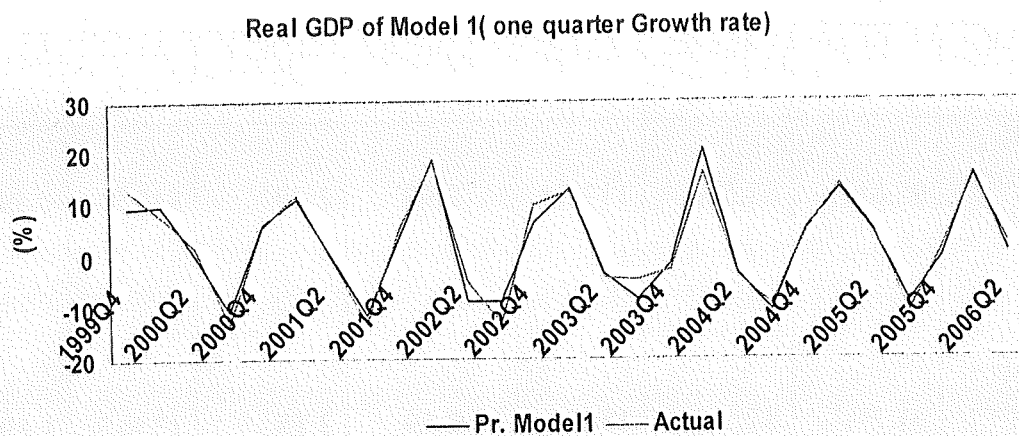
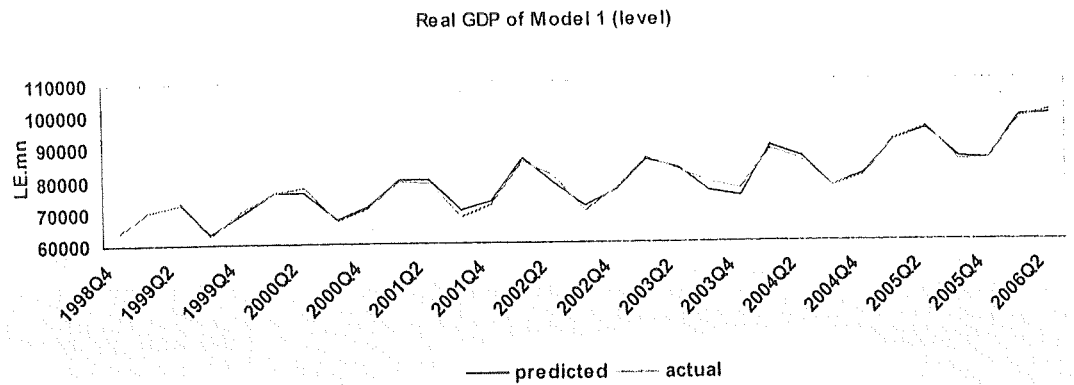


Figure 1: Actual & Predicted values of RGDP for model 1

**Table (4): Sign Accuracy, Root Mean Square Forecast Error,
and Mean Absolute Deviation for the final models**

MODEL	Sign Accuracy (number of correct sign forecasts over total number of forecasts)		Root Mean Square Forecast Error (for level, the unit in LE.mn. For 1 quarter and 4 quarter growth rate, the unit in percentage points.)			Mean Absolute Deviation (For the levels, the unit is in LE.mn. For 1 quarter and 4 quarter growth rate, the unit is in percentage points.)		
	1QGR	4QGR	Level	1QGR	4QGR	Level	1QGR	4QGR
<i>Man_gdp – PP_imports – Agri_gdp</i>	95.8%	87.5%	1172.4	2.09	2.40	913.3	1.59	1.94
<i>Man_gdp – Govt_reven – Agri_gdp</i>	91.7%	87.5%	1545.1	2.39	3.18	1236.3	1.89	2.48
<i>Man_gdp – Trans_gdp</i>	87.5%	91.67%	1619.5	2.81	3.32	1178.4	2.09	2.37
<i>Man_gdp – Govt_reven – MP_exports</i>	87.5%	87.5%	1642.6	3.46	3.13	1389.3	2.93	2.56

growth rate for model 1 (as an example). In this figure, the predicted values and the actual values are shown in the y axis while time is shown on the x axis. The dashed lines are the actual values and the solid lines are the predicted values.

It can readily be seen from this graph, that there is no apparent bias in the case of the first model i.e., there is no a tendency either to under predict or overpredict actual value of real GDP.

Normality tests of the same type as used for the residuals of the regression cannot reject the null of normality of the forecast errors of the level, one quarter growth rate and four quarter growth rate at a 1% significance level in all models. Based on these tests, the assumption that the errors came from a normal distribution seems reasonable. Using this assumption, it can be tested if the mean forecast error for the levels of real GDP, one quarter growth rate and four quarter growth rate are significantly different from 0 using standard statistical theory. This confirms the conclusion reached graphically: There is no significant bias can be detected, the reason for this is that the autocorrelation tests for the four models could not reject the null of no autocorrelation.

Looking at the sign accuracy and measures of forecast point accuracy of the four models, it can be seen that the ranking of models based on fit (R^2) is preserved. Model 1 produces the most precise sign and point forecasts, while model 4 the least precise. The capacity of providing accurate point estimates diminishes substantially in model 4. The root mean square forecast errors and means absolute deviations are greater in models 2, 3 and 4 than its levels in model 1.

IV.2 Misspecification Tests and ex-post Forecasting Performance

Table (3) below shows the p-values for the misspecification tests calculated for the models. The table also shows the p-values for the parameter constancy Chow test.

Table (3): Misspecification tests

Model						Ex-post forecasting test
	LM	ARCH	WH 1	Normality	RESET	Chow
<i>Man_gdp – PP_imports – Agri_gdp</i>	0.98034	0.752399	0.70185	0.325358	0.378380	0.864897
<i>Man_gdp – Govt_reven – Agri_gdp</i>	0.34230	0.993177	0.80201	0.107266	0.65416	0.642157
<i>Man_gdp – Trans_gdp</i>	0.85656	0.502506	0.820461	0.41208	0.520144	0.703509
<i>Man_gdp – Govt_reven – MP_exports</i>	0.17252	0.326596	0.61897	0.66087	0.62244	0.874634

P- Values for all the misspecification tests are all above 0.05 in the table. This means that the null hypothesis of the tests cannot be rejected at a 1% or 5% significance levels. The conclusion from all these tests is that there is no conclusive evidence suggesting that the models are misspecified.

P- Values for the parameter constancy tests are also above 0.05 implying that there is no conclusive evidence to reject the null of parameter constancy at 5% or 1% significance levels. Individual forecast errors t tests and recursive estimation results confirm this conclusion.

IV.3 Ex-ante Forecasting Performance

The main ex-ante forecasting performance statistics of the models are exhibited in Table (4). Figure (1) below illustrates the 1 step ahead performance in predicting the level of real GDP, the one quarter growth rate, and the four quarter

Model 2

$$\begin{aligned}
 D2RGDP = & -1.001 \times DRGDP_{-1} + 1.064 \times D2Man_gdp + 0.522 \times DMan_gdp_{-1} \\
 [T-test\ value] & [-6.093] \quad [9.735] \quad [2.721] \\
 & + 0.086 \times D2Govt_reven + 0.108 \times DGovt_reven_{-1} + 0.377 \times DAgri_gdp \\
 & [2.614] \quad [5.713] \quad [4.492] \\
 & - 2.190 \times ECM_{-2} \\
 & [-11.921]
 \end{aligned}$$

Standard Error of the regression = 0.020909

Model 3

$$\begin{aligned}
 D3RGDP = & -0.51 \times D2RGDP_{-1} - 0.685 \times DRGDP_{-2} + 0.451 \times D3Man_gdp + 0.4553 \times D2Man_gdp_{-1} \\
 [T-test\ value] & [-3.360] \quad [-7.459] \quad [3.207] \quad [2.542] \\
 & + 0.461 \times D3Trans_gdp + 0.5098 \times D2Trans_gdp_{-1} - 2.4885 \times ECM_{-3} \\
 & [6.275] \quad [3.323] \quad [-8.2455]
 \end{aligned}$$

Standard Error of the regression = 0.023142

Model 4

$$\begin{aligned}
 D2RGDP = & -0.403 \times DRGDP_{-1} + 0.771 \times D2Man_gdp + 0.146 \times D2Govt_reven \\
 [T-test\ value] & [-4.802] \quad [7.1764] \quad [6.222] \\
 & + 0.1991 \times D2MP_exports - 1.637 \times ECM_{-2} \\
 & [2.719] \quad [-14.101]
 \end{aligned}$$

Standard Error of the regression = 0.023018

The four models are presented in the error correction form. The coefficients of the error correction variables are known as the feedback coefficients. These coefficients measure the speed of adjustment of deviations from the long run relationships. The error correction variables are simply the deviations of real GDP from the long run relationship.

In general, traditional t-tests show that variables are statistically significant at the 5% and 1% levels. The standard errors of the regression reflect the good fit of all models. For model 1 it is around 1% of the mean of the dependent variable. For models 2,3 and 4 the same statistic are 2%.

IV. Final Models and Their Econometric Properties

All results in this section are obtained using Ordinary Least Squares. Full sample estimates and goodness of fit, misspecification tests. Ex-post forecasting results and ex-ante forecasting results are presented next. A summary of the main conclusions closes the section.

IV.1 Estimation Results and Goodness of Fit

The following Table (2) presents the estimation results of the final models. A few notation clarifications are in order. X_{-j} denotes the j th lag of variable X . $j = 0$ denotes contemporaneous variable. For the $man_gdp - trans_gdp$ model $DiX_i = X_{i-1} - X_{i-3}$ for $j = 0, 1, 2, 3$ and $i = 3 - j$. For example $Dman_gdp_{-2} = man_gdp_{i-2} - man_gdp_{i-3}$. For the other three models $DiX_i = X_{i-1} - X_{i-2}$. For example, $D2man_gdp_{-2} = man_gdp_i - man_gdp_{i-2}$.

Table (2): Estimation results for the final models

Model 1

$$\begin{aligned}
 D2RGDP = & -0.826 \times DRGDP_{-1} + 0.686 \times D2Man_gdp + 0.591 \times Man_gdp_{-1} \\
 [T-test\ value] \quad & [-8.93] \quad [7.821] \quad [4.323] \\
 & + 0.326 \times D2PP_imports - 0.371 \times DPP_imports_{-1} + 0.433 \times D2Agri_gdp \\
 & [4.263] \quad [-2.945] \quad [10.807] \\
 & - 2.099 \times ECM_{-2} \\
 & [-20.887]
 \end{aligned}$$

Standard Error of the regression = 0.017092

Table (1): Unrestricted models ranked according to their R^2

MODEL	R^2	Type
Man_gdp- pp_imports - agri_gdp	0.983147	C
Man_gdp-Govt_reven - agri_gdp	0.97461	C
Man_gdp - trans_gdp	0.973298	C
man_gdp - pp_imports	0.973149	C
Man_gdp-Govt_reven-suez_gdp	0.97227	C
Man_gdp-Govt_reven-mp_exports	0.971619	C
Man_gdp-Govt_reven-pp_exports	0.97038	C
man_gdp - cur_expr	0.968196	C
man_gdp - Govt_reven	0.967400	C
man_gdp - Govt_reven	0.95668	C
pp_imports - agri_gdp - m2r	0.953955	C
pp_imports - agri_gdp - total_depr	0.953761	C
pp_imports - agri_gdp - quasimr	0.953445	C
man_gdp - cur_expr	0.94875	C
pp_imports - agri_gdp	0.93296	C
cur_expr-quasimr - fuel_exports	0.931425	C
cur_expr - quasimr - fuel_imports	0.930083	C
pp_imports - total_depr	0.929497	L
pp_imports - m2r	0.928144	L
cur_expr - qusimr	0.87249	C
cur_expr - mp_exports	0.860169	C
Suez_gdp - fuel_exports	0.83839	C

The two unrestricted models that could allow the construction of leading indicators, as defined in this thesis are not selected because they have low fit, and they failed-in subsequent analysis-in many of the important dimensions discussed before. Finally, it can be shown that the model man_gdp - pp_imports -agri_gdp (real GDP of manufacturing sector, volume of imports of primary products index and real GDP of agriculture sector) encompasses all the other models including the other three selected models.

there are many candidate models for each possibility, it may not be feasible to subject all of them to steps 4 and 5. Models can be screened picking the first x models with the best fit for example. Other alternatives are picking the one whose series are available faster or the ones that make more sense from an economic theory standpoint. In this thesis, the four selected unrestricted models satisfy one or more of the criteria mentioned in step 1 but there is no clear cut rule.

4. Using the estimated coefficients of the error correction mechanisms obtained in the estimation of the unrestricted models, the models selected in step 3 are written in the error correction form. Next, they are simplified using the general to specific modeling technique. Writing models in this way has the advantage of reducing the co-linearity usually present among the different lags and easing the simplification process.
5. Forecast accuracy is assessed for the valid simplifications of the error correction models obtained in step 4.

III. The Selected Model

Table (1) shows the different unrestricted models ranked according to their R^2 goodness of fit measure. Of the models presented in this table, only four are selected. These are man_gdp - pp_imports - agri_gdp (real GDP of manufacturing sector, volume of imports of primary products index and real GDP of agriculture sector), man_gdp - Govt_reven - agri_gdp (real GDP of manufacturing sector, real total revenues of government sector and real GDP of agriculture sector), man_gdp - trans_gdp (real GDP of manufacturing sector and real GDP of transportation sector) and man_gdp - Govt_reven - mp_exports (real GDP of manufacturing sector, real total revenues of government sector and volume of exports of manufactured exports index).

These are all unrestricted models that allow the construction of coincident composite indicators. As discussed before, there is no simple clear cut rule to select among these models. A combination of fit and economic theory was given priority in this selection. Fit is extremely important for forecasting performance and a clear economic theory allows to gain confidence in the econometric results.

The steps towards the construction of the indicators are as follows:

1. A series that is individually cointegrated with real GDP for a given lag length is used as a starting point. Then other series are added to the regression one at a time. Once a variable is added, the expanded two variables vector is tested to see if it is also a cointegrating one. Next, the last added series is deleted and trying another one. This is repeated until all the other series in the dataset are tested. The combinations of two variables that appear to be cointegrated with real GDP as a result of these trials are recorded. Next the process is repeated for each combination of two variables that was found to be cointegrated. As a result, the combinations of three variables that seem to be cointegrated with real GDP are recorded.

With less restrictive data limitations, this search process are repeated until the maximum number of variables that can be included for a given lag length is reached. In this case, for a given starting variable and lag length, the search process would stop at unrestricted models with three variable cointegrating vectors. It must be reminded that each possible unrestricted model error terms must be homoskedastic and the parameter must be constant during the period under study. An additional condition imposed is that the coefficients of the cointegrating vector relationship must make economic sense. Both conditions are required to be accepted as a valid unrestricted model.

The same process is repeated for every series that was found to be cointegrated with real GDP. Only two and three lags are considered. This step is completed by ranking the different unrestricted models according to their R square measures. A good fit is a necessary condition for a good forecasting performance and cited measure is one of the simplest ones that can be used. Another possibility is the regression standard error.

2. The unrestricted models found in step one are tested to see if the contemporaneous variables can be eliminated without a significant loss in the fit using zero restrictions tests. If the contemporaneous variables can be eliminated, it means that only past information can be used to obtain an estimate of real GDP. Therefore, an unrestricted model that has this property is considered to be a candidate to construct a leading indicator. The lead time will be determined by the smallest time subscript that enters the regression after step 4 is completed. If the contemporaneous terms cannot be eliminated, then the unrestricted model is considered to be a candidate to construct a coincident indicator.
3. After step 2, a list of unrestricted models ranked according to their fit can be produced for each of the two possibilities: leading and coincident. If

II. Some Key Methodological and Statistical Issues

II.1 Data Sample

The choice of the data sample for estimating composite indices of coincident and leading indicators for Egypt is dictated by data availability and change in the structure of the economy. The data sample in this paper comprises 31 variables from all sectors of the economy and cover the period from the fourth quarter of 1998 to the second quarter of 2006 (i.e., 32 observations).

Series are classified as pertaining to one of the following four sectors: real, financial, government, and external. All monetary and government variables are in real terms. The external variables are expressed in terms of indices and are constructed by using the data published by the Central Bank of Egypt, commodity prices published by the International Monetary Fund (IMF) and the World Bank. The information sources of all the variables used in this paper are summarized in the APPENDIX. Most series are I(1). A common transformation was performed to the series before initiating the construction of indicators. The natural logarithm was applied to each series to guard against non-linearities.

II.2 Steps towards the Construction of the Indicators

The starting point in the construction of a composite indicator is the formulation of an unrestricted model of real GDP using Autoregressive Distributed Lag (ADL) type models. In simple terms, these models involve regressing the desired dependent variable (real GDP) on lagged values of the dependent variable and the current and lagged values of the independent variables. This model takes the form:

$$y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{j=0}^q \gamma_j x_{t-j} + u_t, \quad u_t \sim iid(0, \sigma^2)$$

Because there are p lags on y_t and q lags on x_t , this is sometimes called an ADL(p , q) model. The models are going to be estimated using Ordinary Least Squares (OLS) for three basic reasons.

Another issue is the choice of the number of lags to include in the ADL models. The choice made was to include either 2 or 3 lags. Due to the data constraints, a maximum of 12 parameters in any attempted unrestricted model are estimated. This allow to have up to a three years forecast period to check the forecasting capabilities of the different candidate indicators even though the first forecasts are being calculated using models estimated with one or two degrees of freedom.

I. INTRODUCTION

The construction of Coincident and Leading Indicators of Economic Activity (CLIEA) has a long tradition, starting with the work of Moore and Shiskin (1967). The main purpose of the CLIEA approach is to address a problem that is specially serious in developing countries. Economic time series, in particular real GDP, are reported with a considerable time lag. In addition, the series are revised frequently to incorporate new information. If the magnitude of these revisions is important, first estimates may not provide a good approximation of the current situation and direction of economic activity. This leaves policy makers and investors without objective measures of the current situation and direction of economic activity.

The two most commonly used methodologies to construct coincident and leading indicators are the National Bureau of Economic Research- Department of Commerce (NBER-DOC) approach, and that of Stock and Watson (SW) (1989). The former is based on the work by Moore and Shiskin (1967), while the latter adds an econometric foundation to the NBER-DOC approach.

Although both NBER-DOC and SW methodologies rely on an abstract concept of economic activity, referred to as "the state of the economy", which is an unobservable variable that must be estimated out of several available series, they differ in the strategies used to construct the coincident and leading indicators and they are subject to important and well known criticisms. The NBER-DOC methodology has been criticized on the basis of its lack of foundation in economic theory⁽³⁾, and its lack of attention to the properties of the time series used⁽⁴⁾. The Stock and Watson methodology shares the shortcomings of the NBER-DOC method, but it incorporates econometrics in the construction of indicators. While this is a step forward, a problem of this latter methodology is that the econometric techniques employed involve heavy data requirements which are not met in most developing countries like Egypt.

In order to overcome both the methodological drawbacks of NBER-DOC and the large data requirements of Stock and Watson, an alternative general methodology is proposed and adapted in this paper to describe how to construct composite indicators that produce level estimates of real GDP for every period, using the limited quarterly series available for Egypt. A composite indicator is said to be "coincident" when information of the series composing it up to time t is required to calculate a point estimate of real GDP at time t . A composite indicator is said to be "leading" by τ time period, when information of the series composing it up to time $t-\tau$ is required to calculate a point estimate of real GDP at time t . The leading indicator is then said to have a lead time of τ .

(3) See De Leeuw in Lahiri and Moore (1991) for details.

(4) See Emerson and Hendry (1996).

Estimating Indices of Coincident and Leading Indicators: An Application to Egypt

Adel A. El-Sherif, Ghazal A. Amer⁽¹⁾, and

Sayed M. El-Sayed⁽²⁾

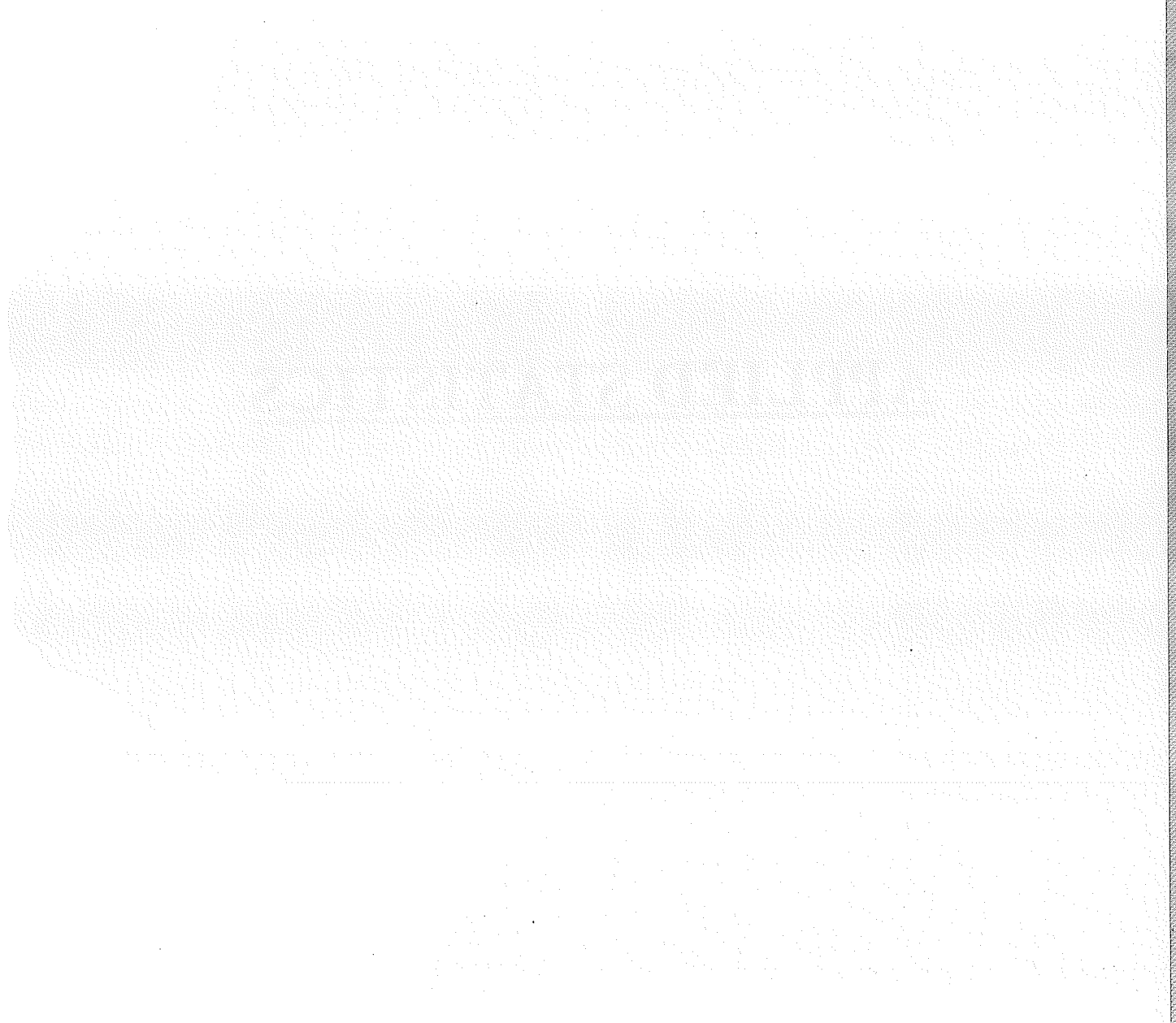
Abstract

Time series of economic activity in developing countries, in particular real GDP, are reported with important lags. Therefore, it is useful to construct indicators that coincide or lead the actual direction and level of economic activity. A general methodology to construct these indicators is proposed and adapted for Egypt. Four coincident indicators could be constructed, but not reliable leading indicator could be found. From an econometric standpoint, the coincident indicators produce satisfactory point estimates of real GDP. From an economic point of view the series that enter the indicator are broadly consistent with what many economists believe is the main source of real GDP fluctuations in Egypt. This enhances the confidence in the econometric results.

Keywords: Egypt, Composite Indicators, Coincident Indicators, Leading Indicators

⁽¹⁾ Department of Applied Statistics & Econometrics, Institute of Statistical Studies & Research, Cairo University.

⁽²⁾ Vice Dean for Graduate Studies & Research, Institute of Statistical Studies & Research, Cairo University.



APPLIED STATISTICS

- Karavanen, J. (2005). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics and Data Analysis*.
- Leonowicz, Z., Karvanen, J. and Shishkin, S.L. (2005). Trimmed estimators for robust averaging of event-related potentials. *Journal of Neuroscience Methods*, **142**, 16-26.
- Pilon, P. and Adamowski K. (1992). The value of regional information to flood frequency analysis using the method of L-moments. *Canadian Journal Civil Engineering*, **19**, 137- 147.
- Ramberg, J.S., Dudewicz, E.J., Tadikamalla, P. and Mykytka E. (1979). A probability distribution and its uses in fitting data. *Technometrics*, **21**, 201-214.
- Rousseeuw, P. and A. Leroy (1987). *Robust Regression and Outlier Detection* (1st ed.). Wiley, New York.
- Sankarasubramanian, A. and K. Srinivasan (1999). Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology*, **218**, 13-34.
- Serfling, G.P. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley& Sons, New York.
- Sillitto, G.P. (1951). Interrelations between certain linear systematic statistics of samples from any continuous population. *Biometrika*, **38**, 377-382.
- Sillitto, G.P. (1969). Derivation of approximations to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, **56**, 641-650.
- Willinger, W., Paxson, V., and Taqqu, M.S. (1998). Self-similarity and heavy tails; structural modeling of network traffic. In *A Practical Guide to Heavy Tails*, eds. Adler, R., Feldman, R., and Taqqu, M.S., pp. 27-53. Boston: Birkhauser.

7 Conclusion

Fractional L-moments are a generalization of the trimmed L-moments. They can be used in practice in a similar way as L-moments and trimmed L-moments. FL-moments provide simple and effective ways of estimate the parameters and inferences for distributions with heavy tails that the mean does not exist. Moreover, they take account of the information at the tail of the distribution by giving less weight for the values at the tails rather than giving them zero weight.

Some questions not addressed in this paper: first one is how to decide on the degree of fraction. It is possible to decide based on the data themselves. This will be discussed in a subsequent paper. The second on the quantile function, how many terms are needed to obtain a usable approximation to population quantile function and what is the optimal choice of s as a function of sample size?. Also, this will be discussed in a subsequent paper.

References

- Abramowitz, M., and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover.
- David, H.A. and Nagaraja, H. (2003). *Order Statistics*, 3rd ed. New York: Wiley.
- Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika*, **53**, 129-141.
- Elamir, E.A.H., and Seheult, A.H. (2001). Control charts based on linear combinations of order statistics. *Journal of Applied Statistics*, **28**, 457-468.
- Elamir, E.A.H., and Seheult, A.H. (2003). Trimmed L-moments. *Computational Statistics and Data Analysis*, **43**, 299-314.
- Elamir, E.A.H. and Seheult, A.H. (2004). Exact variance structure of sample L-moments. *Journal of Statistical Planning and Inference*, **124**, 337-359.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events*. Berlin: Springer-Verlag.
- Harter, H.L. (1984). Another look at plotting positions. *Communication in Statistics-Theory and Methods*, **13**, 1613-1633.
- Hosking, J.R.M. (1986). *The Theory of Probability Weighted Moments*. Research Report RC12210. IBM Research, Yorktown Heights.
- Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistics society, B*, **52**, 105-124.
- Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis: an Approach based on L-moments*, Cambridge, England: Cambridge University Press..

Table4. Simulated biased and mean square error (MSE) of the shape and scale parameters from Pareto distribution using FL-moment estimator with different fractional and number of replication is 5000.

$\eta = -0.75$				
Sample sizes				
	25	50	100	200
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	0.385(0.2164)	0.242(0.0948)	0.168(0.0520)	0.115(0.0287)
(0,1)	-.020(0.1504)	-.053(0.0826)	-.045(0.0436)	-.031(0.0239)
(0,1.05)	-.021(0.1509)	-.053(0.0828)	-.045(0.0436)	-.031(0.0239)
(0,2)	-.052(0.2519)	-.033(0.1189)	-.022(0.0523)	-.006(0.0268)
(0,2.05)	-.052(0.2524)	-.033(0.1191)	-.021(0.0524)	-.006(0.0269)
$\sigma = 1$				
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	0.899(2.9800)	0.537(0.7273)	0.302(0.1679)	0.196(0.1698)
(0,1)	0.117(0.2603)	0.042(0.1022)	-.001(0.0488)	-.011(0.0255)
(0,1.05)	0.116(0.2601)	0.041(0.1022)	-.001(0.0487)	-.011(0.0254)
(0,2)	0.086(0.2730)	0.053(0.1126)	0.012(0.0463)	0.005(0.0232)
(0,2.05)	0.086(0.2731)	0.054(0.1127)	0.012(0.0463)	0.005(0.0232)

Table5. Simulated biased and mean square error (MSE) of the shape and scale parameters from Pareto distribution using FL-moment estimator with different fractional and number of replication is 5000.

$\eta = -1.25$				
Sample sizes				
	25	50	100	200
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	0.572(0.3793)	0.475(0.2508)	0.427(0.1976)	0.401(0.1695)
(0,1)	0.245(0.1926)	0.144(0.0999)	0.084(0.0545)	-.058(0.0553)
(0,1.05)	0.231(0.2349)	0.210(0.1121)	0.001(0.0564)	-.058(0.0555)
(0,2)	0.136(0.2642)	0.081(0.1271)	0.028(0.0585)	-.039(-.0440)
(0,2.05)	0.069(0.2792)	0.052(0.1437)	0.013(0.0613)	-.039(0.0439)
$\sigma = 1$				
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	1.239(5.7847)	1.024(2.3187)	0.949(2.1482)	1.86(33.015)
(0,1)	0.254(0.5118)	0.129(0.1790)	0.078(0.0841)	-.001(0.0525)
(0,1.05)	0.199(0.441)	0.661(0.1957)	0.103(0.0960)	-.002(0.0524)
(0,2)	0.118(0.3871)	0.054(0.1381)	0.016(0.0595)	0.008(0.0342)
(0,2.05)	0.074(0.3710)	0.033(0.1570)	0.007(0.0620)	0.008(0.0341)

moments except for $\eta \leq -(\alpha + \beta + 2)$ and $\alpha + \beta$ is integer numbers. We consider this as advantage of FL-moments over L-moments and TL-moments.

Solving above equations give

$$\eta = \frac{2(\alpha + \beta + 3) - (3\beta^2 + 3\alpha\beta + 15\beta + 6\alpha + 18)\tau_3^{(\alpha, \beta)}}{(2\alpha + 2\beta + 6) + (3\beta + 6)\tau_3^{(\alpha, \beta)}}$$

And

$$\sigma = \frac{2\Gamma(\beta + 2)\Gamma(\alpha + \beta + \eta + 3)}{\Gamma(\alpha + \beta + 3)\Gamma(\beta + \eta + 1)}\lambda_2^{(\alpha, \beta)}$$

We can estimate them using sample FL-moments as

$$\hat{\eta} = \frac{2(\alpha + \beta + 3) - (3\beta^2 + 3\alpha\beta + 15\beta + 6\alpha + 18)\hat{\tau}_3^{(\alpha, \beta)}}{(2\alpha + 2\beta + 6) + (3\beta + 6)\hat{\tau}_3^{(\alpha, \beta)}}$$

And

$$\hat{\sigma} = \frac{2\Gamma(\beta + 2)\Gamma(\alpha + \beta + \hat{\eta} + 3)}{\Gamma(\alpha + \beta + 3)\Gamma(\beta + \hat{\eta} + 1)}l_2^{(\alpha, \beta)}$$

For FL-moment estimators with different fractions (zeros L-moments, integers TL-moments) bias and mean square error of estimates of the scale and shape parameters were obtained using Monte Carlo simulation. Tables 3, 4 and 5 shows the results for sample sizes 25, 50, 100 and 200 over a range of negative values of η . As η decreases and the tail weight of the distribution increases, larger amount of fraction become preferable: where they have lowest mean square error. This is also true for $\hat{\sigma}$.

Table3. Simulated biased and mean square error (MSE) of the shape and scale parameters from Pareto distribution using FL-moment estimator with different fractional and number of replication is 5000.

$\eta = -0.25$				
Sample sizes				
	25	50	100	200
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	-.144(0.0871)	0.110(0.0420)	0.058(0.0199)	0.033(0.0100)
(0,1)	-.195(0.1679)	-.017(0.0491)	-.009(0.0264)	0.003(0.0129)
(0,1.05)	-.172(0.1746)	-.017(0.0493)	-.009(0.0264)	0.003(0.0130)
(0,2)	-.218(0.3263)	-.009(0.1014)	-.004(0.0492)	0.008(0.0256)
(0,2.05)	-.229(0.3487)	-.009(0.1018)	-.004(0.0493)	0.008(0.0257)
$\sigma = 1$				
(α, β)	Bias (MSE)	Bias (MSE)	Bias (MSE)	Bias (MSE)
(0,0)	0.120(0.1906)	0.149(0.1091)	0.079(0.0385)	0.033(0.0193)
(0,1)	0.065(0.1942)	0.019(0.0751)	0.013(0.0338)	0.003(0.0168)
(0,1.05)	0.080(0.2058)	0.019(0.0751)	0.013(0.0338)	0.003(0.0168)
(0,2)	0.055(0.2398)	0.026(0.0876)	0.017(0.0409)	0.007(0.0197)
(0,2.05)	0.050(0.2424)	0.026(0.0877)	0.017(0.0409)	0.007(0.0198)

Some values of these weights given in table (2). From this table we see that the trimmed L-moments assign zero weights for the extreme observation but sample FL-moments assigns less weight for the extreme observations which capture more information about the tail of the distribution.

Table 2: The weights of the sample FL-moments $l_1^{(\alpha,\beta)}$, $l_2^{(\alpha,\beta)}$, and $l_3^{(\alpha,\beta)}$ for sample size $n = 10$ and different choices of α and β .

Est.	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
$l_1^{(0,0)}$	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
$l_1^{(1,1)}$	0	0.067	0.117	0.150	0.167	0.167	0.150	0.117	0.067	0
$l_1^{(99,99)}$	0.001	0.067	0.117	0.149	0.166	0.166	0.149	0.117	0.067	0.001
$l_2^{(0,0)}$	-0.100	-0.077	-0.055	-0.033	-0.011	0.011	0.033	0.055	0.078	0.100
$l_2^{(1,1)}$	0	-0.066	-0.083	-0.064	-0.024	0.024	0.064	0.083	0.067	0
$l_2^{(99,99)}$	-0.001	-0.067	-0.083	-0.064	-0.023	0.023	0.064	0.083	0.067	0.001
$l_3^{(0,0)}$	0.100	0.033	-0.017	-0.050	-0.067	-0.067	-0.050	-0.017	0.033	0.100
$l_3^{(1,1)}$	0	0.074	0.037	-0.032	-0.079	-0.079	-0.032	0.037	0.074	0
$l_3^{(99,99)}$	0.002	0.074	-0.036	-0.032	-0.079	-0.079	-0.032	0.036	0.074	0.002

6 Application

The generalized Pareto distribution with scale and shape parameters σ and η has the quantile function

$$x(F) = \frac{\sigma}{\eta} - \frac{\sigma}{\eta} (1 - F)^\eta$$

It is widely used model in fields such as insurance, finance and network traffic distribution where these applications are most likely to be heavy tails. The Fractional L-moments for the distribution via (9) are

$$\lambda_1^{(\alpha,\beta)} = \frac{\sigma}{\eta} \left[1 - \frac{\Gamma(\alpha + \beta + 2)\Gamma(\beta + \eta + 1)}{\Gamma(\beta + 1)\Gamma(\alpha + \beta + \eta + 2)} \right],$$

$$\lambda_2^{(\alpha,\beta)} = \frac{\sigma\Gamma(\alpha + \beta + 3)\Gamma(\beta + \eta + 1)}{2\Gamma(\beta + 2)\Gamma(\alpha + \eta + \beta + 3)}$$

and

$$\lambda_3^{(\alpha,\beta)} = \frac{-\sigma(\eta - 1)\Gamma(\alpha + \beta + 4)\Gamma(\beta + \eta + 1)}{3\Gamma(\beta + 3)\Gamma(\alpha + \eta + \beta + 4)}$$

When α, β are integer numbers the fractional mean of the distribution does not exist for $\eta \leq -(\beta + 1)$ and so do all higher FL-moments. Hence, for the L-moments the mean does not exist for $\eta \leq -1$ and so do all higher L-moments. Also, when α, β are fractional, the fractional mean exists for all values of η and so do all higher fractional

5 Sample FL-moments

We consider estimators of population FL-moments which are functions of order statistics $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ of a random sample $X_1, X_2, \dots, X_n$ of size n from the population via two methods.

5.1 The Direct Biased Estimator

We can write population FL-moments as

$$\lambda_{r+1}^{(\alpha, \beta)} = \frac{1}{r+1} \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} \binom{r+\alpha+\beta+1}{r-k+\beta} (\alpha+k+1) E[x(F) F^{\alpha+k} (1-F)^{\beta+r-k}] \quad (25)$$

This can be estimated based on one of the following plotting positions

$$E(F) = \frac{i}{n+1}, \quad \text{Mod}(F) = \frac{i-1}{n-1} \quad \text{or} \quad \text{Med}(F) \cong \frac{i-0.33}{n+0.38}$$

See; Harter (1984), for example, if we used the $E(F)$

$$\tilde{\lambda}_{r+1}^{(\alpha, \beta)} = \sum_{k=0}^r (-1)^{r-k} \frac{(\alpha+k+1)}{n(r+1)} \binom{r}{k} \binom{r+\alpha+\beta+1}{r-k+\beta} \sum_{i=1}^n \left[\left(\frac{i}{n+1} \right)^{\alpha+k} \left(\frac{n+1-i}{n+1} \right)^{\beta+r-k} \right] x_{i:n}$$

This form is very useful in proving the asymptotic normality of sample FL-moments.

5.2 Covariance Estimator

We obtain this estimator directly from (15) as

$$\tilde{\lambda}_{r+1}^{(\alpha, \beta)} = \begin{cases} \text{cov}(x, T_r^{(\beta, \alpha)}(F_n)) + \hat{E}(X) & r = 0 \\ \text{cov}(x, T_r^{(\beta, \alpha)}(F_n)) & r \geq 1 \end{cases} \quad (26)$$

Where F_n is the empirical cumulative distribution function of x and $\hat{E}(x) = \bar{x}$ the usual sample mean. This estimator takes advantage of availability of covariance function in many statistical packages such as S-plus, Minitab.

5.3 Weight Function

It possible to write the weight function for the equation (25) as following

$$w_i(r, \alpha, \beta) = \sum_{k=0}^r \frac{(-1)^{r-k} \binom{r}{k} \binom{i-1}{k+\alpha} \binom{n-i}{\beta+r-k}}{(r+1) \binom{n}{r+1+\alpha+\beta}} \quad \text{For } r = 0, 1, \dots, i = 1, \dots, n$$

$$\tilde{\tau}_{r+1}^{(\alpha,\beta)} = \frac{(r+1)\Gamma(M+2)\Gamma(r+\alpha+1)\Gamma(r+\beta+1)\Gamma(\alpha+\beta+3)}{2\Gamma(r+M+1)\Gamma(\alpha+2)\Gamma(\beta+2)\Gamma(r+\alpha+\beta+2)} \frac{\lambda_{r+1}^{(\alpha,\beta)}}{\lambda_2^{(\alpha,\beta)}} \quad (24)$$

This is bounded where

$$|\tilde{\tau}_{r+1}^{(\alpha,\beta)}| \leq 1$$

This takes advantage of L-moments ratio.

For example consider the Pareto distribution with $X(F) = \frac{\sigma}{\eta} - \frac{\sigma}{\eta}(1-F)^\eta$. The FL-skewness

$$\text{is } \tau_3^{(\alpha,\beta)} = \frac{-2(\eta-1)(\alpha+\beta+3)}{3(\beta+2)(\alpha+\eta+\beta+3)} \text{ and the scaled FL-skewness is } \tilde{\tau}_3^{(\alpha,\beta)} = \frac{-(\alpha+2)(\eta-1)}{(M+2)(\alpha+\beta+\eta+3)}.$$

For $\alpha = \beta = 1.05$, $\tau_3^{(1.05,1.05)} = \frac{-10.20(\eta-1)}{9.15(\eta+5.10)}$, which exceeds 1 when $\eta \leq -1.9$ and

approaches its limit $10.10/9.15$ as $\eta \rightarrow -2.05$. While $\tilde{\tau}_3^{(1.05,1.05)} = \frac{-(\eta-1)}{(\eta+5.10)}$

approaches its upper bound 1 as $\eta \rightarrow -2.05$ and is equal to 0.906 when $\eta = -1.90$.

As an example Table 1 shows the L-moments ratio and scaled FL-moments ratio for the generalized lambda distribution

$$X(F) = a + \frac{F^c - (1-F)^d}{b}$$

Where a , b , c and d are the parameters; see, Ramberg et al. (1979).

Table 1 Scaled ratio FL-moment from generalized lambda distribution with different choices of parameters.

Parameters (a,b,d,c)	$(\tau_3, \tilde{\tau}_3^{(1.05,1.05)})$	$(\tau_4, \tilde{\tau}_4^{(1.05,1.05)})$	$(\tau_5, \tilde{\tau}_5^{(1.05,1.05)})$	$(\tau_6, \tilde{\tau}_6^{(1.05,1.05)})$
(0,1,1,1)	(0,0)	(0,0)	(0,0)	(0,0)
(0,1,0.75,0.75)	(0,0)	(0.017,0.008)	(0,0)	(0.003,0.001)
(0,1,0.5,0.5)	(0,0)	(0.047,0.020)	(0,0)	(0.011,0.0027)
(0,1,0.1349,0.1349)	(0,0)	(0.124,0.049)	(0,0)	(0.043,0.009)
(0,1,0.001,0.001)	(0,0)	(0.166,0.064)	(0,0)	(0.066,0.013)
(0,-1,-.25,-.25)	(0,0)	(0.272,0.099)	(0,0)	(0.138,0.025)
(0,-1,-.5,-.5)	(0,0)	(0.428,0.145)	(0,0)	(0.273,0.046)
(0,-1,-1,-1)	(0,0)	(∞ ,0.286)	(0,0)	(∞ ,0.132)
(0,-1,-1.5,-1.5)	(0,0)	(∞ ,0.528)	(0,0)	(∞ ,0.354)
(0,1,0.001,0.1349)	(0.270,0.166)	(0.125,0.050)	(0.068,0.019)	(0.044,0.009)
(0,1,0.195,0.921)	(-.065,-.034)	(0.037,0.014)	(-.017,-.004)	(0.012,0.002)
(0,1,0.154,0.002)	(-.259,-.155)	(0.120,0.048)	(-.063,-.018)	(0.041,0.009)
(0,-1,-.001,-.1349)	(0.392,0.226)	(0.218,0.082)	(0.139,0.036)	(0.099,0.019)
(0,-1,-.195,-.921)	(-.907,-.409)	(0.869,0.247)	(-.830,-.150)	(0.808,0.106)
(0,-1,-1,-2)	(∞ ,0.956)	(∞ ,0.940)	(∞ ,0.917)	(∞ ,0.904)

We note from the table that, at some values of the parameters the L-moment ratios do not exist. While, the FL-moment ratios exist for the same values. Thus, we conclude that the FL-moments can expand the range of applicability of the method to include more parameter values.

This is similar to L-moments ratio; see, Hosking (1990). L-moments ratio is bounded by 1 in absolute value.

If the quantile function $x(F)$ is differentiable, we can integrate (9) by parts to obtain

$$\lambda_{r+1}^{(\alpha, \beta)} = \frac{(r-1)! \Gamma(r+\alpha+\beta+2)}{(r+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)} \int_0^1 x'(F) F^{\alpha+1} (1-F)^{\beta+1} P_{r-1}^{(\beta+1, \alpha+1)}(F) dF \quad (23)$$

Where $x'(F) = 1/f(x)$ is the 1st derivative of quantile function, and

$$\int_0^1 F^{\alpha} (1-F)^{\beta} P_r^{(\beta, \alpha)}(F) dF = - \frac{F^{\alpha+1} (1-F)^{\beta+1} P_{r-1}^{(\beta+1, \alpha+1)}(F)}{r}$$

see, Abramowitz and Stegun (1972).

To obtain bound for the FL-moment ratio we find from (23) that

$$\lambda_{r+1}^{(\alpha, \beta)} = \frac{(r-1)! \Gamma(r+\alpha+\beta+2)}{(r+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)} \int_x F^{\alpha+1}(x) [1-F(x)]^{\beta+1} P_{r-1}^{(\beta+1, \alpha+1)}(F) dx$$

Where $x'(F) = 1/f(x)$ and $dF(x) = f(x)dx$

Also

$$\lambda_2^{(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta+3)}{2\Gamma(\alpha+2)\Gamma(\beta+2)} \int_x F^{\alpha+1}(x) [1-F(x)]^{\beta+1} dx$$

Note that $P_0^{(\alpha, \beta)}(F) = 1$.

From Abramowitz and Stegun (1972)

$$\text{Max}_F |P_{r-1}^{(\alpha+1, \beta+1)}(F)| = \binom{r+M}{r-1} \quad \text{Where } M = \max(\alpha, \beta),$$

Therefore,

$$\begin{aligned} |\lambda_{r+1}^{(\alpha, \beta)}| &\leq \frac{(r-1)! \Gamma(r+\alpha+\beta+2)}{(r+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)} \binom{r+M}{r-1} \int_x F^{\alpha+1}(x) [1-F(x)]^{\beta+1} dx \\ &= \frac{(r-1)! \Gamma(r+\alpha+\beta+2)}{(r+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)} \binom{r+M}{r-1} \frac{2\Gamma(\alpha+2)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+3)} \lambda_2^{(\alpha, \beta)} \end{aligned}$$

Hence

$$\left| \tau_{r+1}^{(\alpha, \beta)} \right| = \frac{|\lambda_{r+1}^{(\alpha, \beta)}|}{\lambda_2^{(\alpha, \beta)}} \leq \frac{2\Gamma(r+M+1)\Gamma(\alpha+2)\Gamma(\beta+2)\Gamma(r+\alpha+\beta+2)}{(r+1)\Gamma(M+2)\Gamma(r+\alpha+1)\Gamma(r+\beta+1)\Gamma(\alpha+\beta+3)}$$

This bound could be greater than 1 for values of $\alpha, \beta > 0$ also increase with increasing value of r . To take advantage of limited values of ratio L-moment we defined scaled FL-moments ratio as

Following Sillitto (1969) and Hosking (1990) we find from (19) that

$$0 = \lim_{s \rightarrow \infty} \int_0^1 F^\alpha (1-F)^\beta R^2(F) dF =$$

$$\int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - \sum_{r=0}^{\infty} \frac{(r+1)^2 (2r+\alpha+\beta+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)}{r! (r+\alpha+\beta+1)^2 \Gamma(r+\alpha+\beta+1)} \lambda_{r+1}^{2(\alpha,\beta)}$$

This can be written as

$$0 = \int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - \frac{\Gamma(\alpha+1) \Gamma(\beta+1)}{(\alpha+\beta+1) \Gamma(\alpha+\beta+1)} \lambda_1^{2(\alpha,\beta)}$$

$$- \sum_{r=1}^{\infty} \frac{(r+1)^2 (2r+\alpha+\beta+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)}{r! (r+\alpha+\beta+1)^2 \Gamma(r+\alpha+\beta+1)} \lambda_{r+1}^{2(\alpha,\beta)}$$

Hence

$$\int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - \frac{\Gamma(\alpha+1) \Gamma(\beta+1)}{(\alpha+\beta+1) \Gamma(\alpha+\beta+1)} \lambda_1^{2(\alpha,\beta)}$$

$$= \sum_{r=1}^{\infty} \frac{(r+1)^2 (2r+\alpha+\beta+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)}{r! (r+\alpha+\beta+1)^2 \Gamma(r+\alpha+\beta+1)} \lambda_{r+1}^{2(\alpha,\beta)}$$

From this equation we obtain

$$FV^{(\alpha,\beta)}(X) = \int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - \frac{\Gamma(\alpha+1) \Gamma(\beta+1)}{(\alpha+\beta+1) \Gamma(\alpha+\beta+1)} \lambda_1^{2(\alpha,\beta)}$$

$$= \int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - \frac{\Gamma(\alpha+\beta+2)}{(\alpha+1) \Gamma(\beta+1)} \left[\int_0^1 F^\alpha (1-F)^\beta x(F) dF \right]^2$$

$$= \sum_{r=1}^{\infty} \frac{(r+1)^2 (2r+\alpha+\beta+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)}{r! (r+\alpha+\beta+1)^2 \Gamma(r+\alpha+\beta+1)} \lambda_{r+1}^{2(\alpha,\beta)}$$

When $\alpha = \beta = 0$ we obtain the variance of population in terms of L-moments; see, Sillitto (1969), as

$$FV^{(0,0)}(X) = \sigma^2 = \int_0^1 x^2(F) dF - \lambda_1^2 = \sum_{r=1}^{\infty} (2r+1) \lambda_{r+1}^2$$

It is clear from the representation of fractional variance in terms of FL-moments that $FV^{(\alpha,\beta)}(X) > 0$.

4 FL-moments Ratio

We define the FL-moments ratio as

$$\tau_{r+1}^{(\alpha,\beta)} = \frac{\lambda_{r+1}^{(\alpha,\beta)}}{\lambda_2^{(\alpha,\beta)}}, \quad r = 2, 3, \dots \quad (22)$$

$$\int_0^1 F^\alpha (1-F)^\beta P_m^{(\beta, \alpha)}(F) P_n^{(\beta, \alpha)}(F) dF = \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)} \delta_{mn}$$

Note that, $\delta_{mn} = 1$ for $m = n$ and $\delta_{mn} = 0$ for $m \neq n$.

The weighted mean square error is minimizing by choosing

$$c_r = \frac{(r+1)(2r+\alpha+\beta+1)}{r+\alpha+\beta+1} \lambda_{r+1}^{(\alpha, \beta)}$$

and the second derivative is positive where

$$2 \sum_{r=0}^s \frac{\Gamma(r+\alpha+1)\Gamma(r+\beta+1)}{r!(2r+\alpha+\beta+1)\Gamma(r+\alpha+\beta+1)} > 0$$

That

$$\int_0^1 F^\alpha (1-F)^\beta [R_s(F)]^2 dF \rightarrow 0 \quad \text{as } s \rightarrow \infty$$

For all $r = 0, 1, 2, \dots, \alpha, \beta \geq 0$ and $s \rightarrow \infty$.

3.1 Fractional Mean and Fractional Variance

We define the fractional mean and the fractional variance which is the generalization of the standard mean and variance of population in terms of FL-moments.

Corollary1 The fractional mean and variance can be defined as

$$\mu^{(\alpha, \beta)} = \lambda_1^{(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} E[x(F)F^\alpha (1-F)^\beta] \quad (20)$$

And

$$FV^{(\alpha, \beta)}(X) = \sum_{r=1}^{\infty} \frac{(r+1)^2 (2r+\alpha+\beta+1)\Gamma(r+\alpha+1)\Gamma(r+\beta+1)}{r!(r+\alpha+\beta+1)^2 \Gamma(r+\alpha+\beta+1)} \lambda_{r+1}^{2(\alpha, \beta)} \quad (21)$$

For $\alpha, \beta \geq 0$.

Proof:

When $r = 0$, we obtain the fractional mean from (9) as

$$\mu^{(\alpha, \beta)} = \lambda_1^{(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} E[x(F)(1-F)^\alpha F^\beta]$$

This is the usual mean weighted by power of cumulative distribution function multiplying by a constant and is equal to population mean $\mu = E(X)$ when the distribution is symmetric and $\alpha = \beta$.

$$\mu^{(\alpha, \alpha)} = \mu = \frac{\Gamma(2\alpha+2)}{\Gamma^2(\alpha+1)} E[x(F)F^\alpha (1-F)^\alpha]$$

3 Representation of Quantile Function in terms of FL-moments

FL-moments can be defined as the coefficients in the expansion of the quantile function as a weighted sum of shifted Jacobi polynomial as L-moments, see Sillitto (1969), Hosking (1986) and Hosking (1990).

Theorem 1: Let X be a continuous real-valued variable such that its fractional mean is finite, quantile function $x(F)$ and fractional L-moments $\lambda_{r+1}^{(\alpha, \beta)}$, $r = 0, 1, 2, \dots$ then the representation

$$x(F) = \sum_{r=0}^{\infty} \frac{(r+1)(2r+\alpha+\beta+1)}{r+\alpha+\beta+1} \lambda_{r+1}^{(\alpha, \beta)} P_r^{(\beta, \alpha)}(F) \quad (18)$$

This is convergent in the weighted mean square with weight function $F^\alpha (1-F)^\beta$.

Proof:

For the number of terms s

$$R(F) = x(F) - \sum_{r=0}^s c_r P_r^{(\beta, \alpha)}(F) \quad (19)$$

Our aim is to minimize the following

$$\text{Min.} \int F^\alpha (1-F)^\beta R^2(F) dF$$

This can be written as

$$\begin{aligned} \int F^\alpha (1-F)^\beta R^2(F) dF &= \int F^\alpha (1-F)^\beta \left[x(F) - \sum_{r=0}^s c_r P_r^{(\beta, \alpha)}(F) \right]^2 dF = \\ &= \left[\int F^\alpha (1-F)^\beta x^2(F) dF - 2 \sum_{r=0}^s c_r \int F^\alpha (1-F)^\beta x(F) P_r^{(\beta, \alpha)}(F) dF \right. \\ &\quad \left. + \sum_{r=0}^s \sum_{v=0}^s c_r c_v \int F^\alpha (1-F)^\beta P_r^{(\beta, \alpha)}(F) P_v^{(\beta, \alpha)}(F) dF \right] \\ &= \int_0^1 F^\alpha (1-F)^\beta x^2(F) dF - 2 \sum_{r=0}^s c_r \frac{(r+1)\Gamma(r+\alpha+1)\Gamma(r+\beta+1)}{r!\Gamma(r+\alpha+\beta+2)} \lambda_{r+1}^{(\alpha, \beta)} \\ &\quad + \sum_{r=0}^s c_r^2 \frac{\Gamma(r+\alpha+1)\Gamma(r+\beta+1)}{r!(2r+\alpha+\beta+1)\Gamma(r+\alpha+\beta+1)} \end{aligned}$$

By using the following in the second term

$$\lambda_{r+1}^{(\alpha, \beta)} = \frac{r!\Gamma(r+\alpha+\beta+2)}{(r+1)\Gamma(r+\alpha+1)\Gamma(r+\beta+1)} \int [x(F) F^\alpha (1-F)^\beta P_r^{(\beta, \alpha)}(F)] dF$$

And the following orthogonal property of the shifted Jacobi polynomial

2.3 An Expression for FL-moments as a Covariance

Direct rewriting of (9) yield

$$\lambda_{r+1}^{(\alpha, \beta)} = \int_0^1 x(F) T_r^{(\beta, \alpha)}(F) dF \quad (13)$$

Where

$$T_r^{(\beta, \alpha)}(F) = \sum_{k=0}^r (-1)^{r-k} \frac{(\alpha + k + 1) \binom{r}{k} \binom{r + \alpha + \beta + 1}{r - k + \beta}}{(r + 1)} F^{k + \alpha} (1 - F)^{r - k + \beta} \quad (14)$$

Using orthogonality of the function $T_r^{(\beta, \alpha)}(F)$ we obtain

$$\lambda_{r+1}^{(\alpha, \beta)} = \begin{cases} \text{Cov}(x(F), T_0^{(\beta, \alpha)}(F)) + E(X) & r = 0 \\ \text{Cov}(x(F), T_r^{(\beta, \alpha)}(F)) & r \geq 1 \end{cases} \quad (15)$$

Thus FL-moments can be interpreted as the covariance of X and its rank. Also, this representation yields a very useful inequality comparing the second FL-moments with population standard deviation:

$$\lambda_2^{(\alpha, \beta)} = \text{Cov}(x(F), T_1^{(\beta, \alpha)}(F)) = \text{Corr}(x(F), T_1^{(\beta, \alpha)}(F)) \sigma_{T_1^{(\beta, \alpha)}} \sigma_x \quad (16)$$

It could be shown that

$$E[F^b (1 - F)^a] = \frac{\Gamma(b + 1) \Gamma(a + 1)}{\Gamma(b + a + 2)}$$

Using this we find

$$\sigma^2(T_1^{(\beta, \alpha)}) = \frac{(\alpha + 1)(\alpha + \beta + 1)\Gamma^2(\alpha + \beta + 3)\Gamma(2\beta + 1)\Gamma(2\alpha + 1)}{2(\beta + 1)\Gamma(2\alpha + 2\beta + 4)\Gamma^2(\beta + 1)\Gamma^2(\alpha + 2)}$$

Thus we obtain

$$\lambda_2^{(\alpha, \beta)} \leq \sigma_x \sqrt{\frac{(\alpha + 1)(\alpha + \beta + 1)\Gamma^2(3 + \alpha + \beta)\Gamma(2\alpha + 1)\Gamma(2\beta + 1)}{2(\beta + 1)\Gamma(2\alpha + 2\beta + 4)\Gamma^2(\alpha + 1)\Gamma^2(\beta + 2)}} \quad (17)$$

σ_x is standard deviation of the population.

For example, when $\alpha = \beta = 0$, we obtain $\lambda_2^{(0,0)} \leq \sigma / \sqrt{3}$, also for $\alpha = \beta = 1$, $\lambda_2^{(1,1)} \leq \sigma \sqrt{6/35}$.

where

$$P_r^{(\beta, \alpha)}(F) = \sum_{k=0}^r (-1)^k \frac{\Gamma(r+\alpha+1)\Gamma(r+\beta+1)}{k!\Gamma(r-k+1)\Gamma(r-k+\beta+1)\Gamma(\alpha+k+1)} F^k (1-F)^{r-k} \quad (10)$$

is the shifted Jacobi polynomial in the interval $[0, 1]$ and we assume that $\alpha, \beta \geq 0$ although the shifted Jacobi polynomial is defined for $\alpha, \beta > -1$.

When $\alpha = \beta = 0$, $\lambda_{r+1}^{(0,0)}$ is the L-moments, also integer values of α, β , $\lambda_{r+1}^{(\alpha, \beta)}$ gives TL-moments.

2.2 Relations

In this section we use some recurrence relation of Jacobi polynomial to derive corresponding relations between FL-moments with different degrees of fractional.

Relation 1

For $r = 0, 1, 2, \dots$ and $\alpha, \beta \geq 0$

$$(r+1)(r+\alpha+\beta+2)\lambda_{r+1}^{(\alpha, \beta)} = (r+1)(2r+\alpha+\beta+2)\lambda_{r+1}^{(\alpha, \beta+1)} + (r+2)(r+\beta+1)\lambda_{r+2}^{(\alpha, \beta)} \quad (11)$$

Proof:

From Abramowitz and Stegun (1972) we find

$$(1-x)(2n+\alpha+\beta+2)P_n^{(\alpha+1, \beta)}(x) = 2(n+\alpha+1)P_n^{(\alpha, \beta)}(x) - 2(n+1)P_{n+1}^{(\alpha, \beta)}(x)$$

This can be expressed in terms of shifted Jacobi polynomial as

$$(1-u)(2r+\alpha+\beta+2)P_r^{(\alpha+1, \beta)}(u) = (r+\alpha+1)P_r^{(\alpha, \beta)}(u) - (r+1)P_{r+1}^{(\alpha, \beta)}(u)$$

Direct substitution of this expression in (9) gives relation 1.

Relation 2

For $r = 0, 1, 2, \dots$ and $\alpha, \beta \geq 0$

$$(r+1)(r+\alpha+\beta+2)\lambda_{r+1}^{(\alpha, \beta)} = (r+1)(2r+\alpha+\beta+2)\lambda_{r+1}^{(\alpha+1, \beta)} - (r+2)(r+\alpha+1)\lambda_{r+2}^{(\alpha, \beta)} \quad (12)$$

Proof:

From Abramowitz and Stegun (1972) we find

$$(1+x)(2n+\alpha+\beta+2)P_n^{(\alpha, \beta+1)}(x) = 2(n+\beta+1)P_n^{(\alpha, \beta)}(x) + 2(n+1)P_{n+1}^{(\alpha, \beta)}(x)$$

This can be expressed in terms of shifted Jacobi polynomial as

$$(1-u)(2r+\alpha+\beta+2)P_r^{(\alpha+1, \beta)}(u) = (r+\alpha+1)P_r^{(\alpha, \beta)}(u) - (r+1)P_{r+1}^{(\alpha, \beta)}(u)$$

Direct substitution of this expression in (9) gives relation 2.

Substituting equation (2) in (3) we obtain

$$\begin{aligned}\lambda_r^{(t_1, t_2)} &= r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \frac{(r+t_1+t_2)!}{(r+t_1-k-1)!(t_2+k)!} \int_0^1 x(F) F^{r+t_1-k-1} (1-F)^{t_2+k} dF \\ &= r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \frac{(r+t_1+t_2)!}{(r+t_1-k-1)!(t_2+k)!} E[x(F) F^{r+t_1-k-1} (1-F)^{t_2+k}] \quad (5)\end{aligned}$$

Clearly TL-moments is defined to Cauchy distribution where $E(X_{2,3})$ exist for $t_1 = t_2 = 1$ and reduce to L-moments when $t_1 = t_2 = 0$, for example,

$$\lambda_1^{(0,0)} = \lambda_1 = \int_0^1 x(F) dF \text{ is the sample mean.}$$

Note that the definition of TL-moments in (4) and (5) does not involve the expectations of the t_1 smallest or the t_2 largest order statistics of the sample size $t_1 + t_2 + r$. Hence, the sample TL-moment assigns zero weights for these values. This may lead to ignoring information at the tail of the distribution. To avoid this we define the Fractional L-moments (FL-moments) from (5) as

$$\lambda_r^{(\alpha, \beta)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \frac{\Gamma(r+\alpha+\beta+1)}{\Gamma(r+\alpha-k)\Gamma(\beta+k+1)} E[x(F) F^{r+\alpha-k-1} (1-F)^{\beta+k}] \quad (6)$$

where $\alpha, \beta \geq 0$.

Next we look at expression of FL-moments in terms of Jacobi polynomial.

2.1 FL-moments in terms of Jacobi Polynomials

The Jacobi polynomial of degree n is defined by

$$P_n^{(\alpha, \beta)}(X) = \frac{1}{2^n} \sum_{k=0}^n \binom{n+\alpha}{k} \binom{n+\beta}{n-k} (X-1)^{n-k} (X+1)^k, \quad \alpha, \beta > -1 \quad (7)$$

See, Abramowitz and Stegun (1972). They are orthogonal on the interval $[-1, 1]$ with respect to the weight function $(1-x)^\alpha (1+x)^\beta$. The shifted Jacobi polynomial can be obtain from (6) via $X = 2F - 1$ as

$$P_n^{(\alpha, \beta)} = \sum_{k=0}^n (-1)^{n-k} \binom{n+\alpha}{k} \binom{n+\beta}{n-k} F^k (1-F)^{n-k} \quad (8)$$

this is orthogonal on $[0, 1]$ with respect to the weight function $F^\beta (1-F)^\alpha$.

By changing the variable in (6) and multiplying by constant we re-write (6) in terms of shifted Jacobi polynomial as

$$\lambda_{r+1}^{(\alpha, \beta)} = \frac{r! \Gamma(r+\alpha+\beta+2)}{(r+1) \Gamma(r+\alpha+1) \Gamma(r+\beta+1)} E[x(F) F^\alpha (1-F)^\beta P_r^{(\beta, \alpha)}(F)], \quad r = 0, 1, \dots, \alpha, \beta \geq 0 \quad (9)$$

ends of the distribution which is reflected in the sample trimmed L-moments by assigning these extreme values weight zeros. This may lead to loss of information. To avoid this, we define the fractional L-moments which assign very small weight for the extreme values. This could help in capture more information about the tail of the distribution.

In Section 2 we define the fractional L-moments and give a condition under which they defined, re-express the FL-moments in term of Jacobi polynomial, and proof some relations between FL-moments with different fractional, express FL-moments in terms of the covariance between the distribution function and its rank. The representation of quantile function in terms of FL-moments and a proof that the FL-moments uniquely define a distribution are given in Section 3. We give the definition of ratio FL-moments and bound on it in Section 4. The estimation of population FL-moments with different three methods is given in Section 5. The estimation of generalized Pareto distribution based on FL-moments with different fractions is illustrated in Section 6.

2 Fractional L-moments (FL-moments)

Let $X_1, X_2, \dots, X_n$ be a sample from a continuous distribution with quantile function $x(F)$, cumulative distribution function $F(x) = F_x = F$, $f(x)$ the density function and let $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ denote the corresponding order statistics. Hosking (1990) defined the r^{th} L-moment as

$$\lambda_r = r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}) \quad r = 1, 2, \dots \quad (1)$$

using the standard form of expected value

$$E(X_{r:n}) = r \binom{n}{r} \int_0^1 x(F) F^{r-1} (1-F)^{n-r} dF, \quad r \leq n \quad (2)$$

see, David (2003). Substituting this expression in (1) gives

$$\begin{aligned} \lambda_r &= r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} \frac{r!}{(r-k-1)!k!} \int_0^1 x(F) F^{r-k-1} (1-F)^k dF \\ &= r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} \frac{r!}{(r-k-1)!k!} E[x(F) F^{r-k-1} (1-F)^k] \end{aligned} \quad (3)$$

While L-moment enjoys certain robustness properties, it is not defined for distributions which has infinite mean; for example, Cauchy distribution.

Elamir and Seheult (2003) extended the idea of L-moments to trimmed L-moment (TL-moments) as

$$\lambda_r^{(t_1, t_2)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r+t_1-k:r+t_1+t_2}), \quad r = 1, 2, \dots, \quad t_1, t_2 = 0, 1, 2, \dots \quad (4)$$

Fractional L-Moments

Elsayed Habib²

Abd-Elftah Kandile¹

Enayat Abd-Elrazik³

Abstract

Linear moments (L-moments) are used for measuring descriptive features and make inferences of a probability distribution but its use is confined to a distribution whose mean is finite. Trimmed Linear moments (TL-moments) are natural extension of L-moments that do not require the mean of the underlying distribution to exist. They do not involve some values at the extreme ends of the distribution which is reflected in the sample TL-moments by assigning these extreme values weight zeros. Therefore, the data at the tails of the distribution it will be ignored. This could lead to the loss of information and should be avoided. However, Fractional Linear moments (FL-moments) is defined as a natural generalization of TL-moments. FL-moments take account of the data at the tail of the distribution where the sample FL-moments captures all the information in the sample about the population counterparts by assigning less weight for extreme values and provide simple and effective ways of estimate the parameters and inferences for distributions with heavy tails that the mean does not exist. The quantile function in terms of FL-moments as a weighted sum of Jacobi polynomial is obtained. Also, the population FL-moments is estimated using three different methods.

Keywords: Estimation; L-moments; Jacobi polynomial; Order statistics; Quantile function.

1 Introduction

As interest in statistical modeling using heavy-tailed distributions is increasing, modeling the frequency of extreme events is of particular interest. In this connection, an extensive L-moment methodology has been developed in support of regional frequency analysis in environmental science, which treats the quantile of distributions of variable such as annual maximum precipitation, stream flow or wind speed observed at each site in a given network; see, Hosking and Wallis (1997) and for other applications; see, for example, Pilon and Adamowski (1992), Sankarasubramanian and Srinivasan (1999) and Elamir and Seheult (2001). TL-moments used by Elamir and Seheult (2003) for the same purpose as L-moments and used by Karavanen (2005) to estimate the mixture quantile of Cauchy distribution and by Leonowicz et al. (2005) to study robust averaging of event-related potentials.

In heavy tailed distributions especially when the mean is not defined the L-moments can not be used; see Embrechts et al. (1997) and Willinger et al. (1998). In this case the trimmed L-moments would be useful where it is defined when the mean is not existed. The trimmed L-moment does not involve some values at the extreme

¹ Professor in the Department of Mathematics, Statistics, and Insurance, Faculty of commerce, Benha University.

² Associate professor in the department of Management and Marketing, Business School, University of Bahrain.

³ Assistant lecturer in the Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha university.

models are almost coincident. While the coefficient of variation is best when $\mathcal{G}$, $\mathcal{G}_1$ are small for the three models as shown in fig. (2.1) and fig. (2.2) especially for large n, r .

Coming back to the compound distributions of sums of truncated Puri and Rubin distribution section we had present two groups of compound distributions. The first group is the compound deterministic - Puri and Rubin (left truncated and non truncated) distributions. We see that the coefficient of variation is best in the compound deterministic - left truncated Puri and Rubin specially for large k as shown in fig. (3.1). The Second group is the compound Bernoulli - Puri and Rubin (left truncated, right truncated and non truncated) distributions. We see that the best model is the compound Bernoulli-left-truncated Puri and Rubin for different α especially when we substitute k by a large number.

According to the best model in each group we can say that it is very useful for us when using these models in different fields such as industry, engineering, queueing theory, the reliability theory, electronics, medicine as in hospital administration and insurance risk theory.

Reference

- [1] Alaa, R. A. A. (2003). "Compound distributions of sums of single truncated random variables", MSC. Thesis sub mited to Department of statistics, Faculty of science, Tanta University, (Abou-El-Ata, m. o.).
- [2] Johnson, N. L., Kemp, A. W., Kotz, S. (2005). "Univariate Discrete Distributions", (third edition), John Wiley & Sons, Inc., Hoboken, New York.
- [3] León, C. A. and Perron, F. (2003). "Extremal properties of sums of Bernoulli random variables", Statistics and Probability Letters, 27, pp. (345-354).
- [4] Lingappaiah, G. S. (1992). "Distribution of the sum of independent right-truncated negative binomial variates", Revista de Matematica e Estadistica, 10, pp. (155-163).
- [5] Medhi, J. (1975). "On the convolutions of left-truncated generalized negative binomial and Poisson variables", Sankhyā, Series B, 37, pp. (293-299).
- [6] ... (1982). "Stochastic Processes", Eastern Limited, first reprint, March (1983). Printed in India.
- [7] Moger, and Odd O. Aalen, (2008). "Regression models for infant mortality data in Norwegian siblings, using a compound Poisson frailty distribution with random scale", Biostatistics, 9(3), pp. (577-591)
- [8] P.Vellaisamy and N.S. Upadhye,(2009). "On the sums of compound negative binominal and gamma random variables", Journal of Applied Probability, V. 46, pp. (272- 283).
- [9] Qu, Y., Greene, T. And Piedmonte, M.R. (1993). "Symmetric Bernoulli distributions and generalized binomial distributions", Biometrical Journal, 35, pp. (229-242).
- [10] Stathis Chadijconstantinidis and Georgios Pitselis.(2008). "Further improved recursions for a class of compound Poisson distributions", Mathematics and Economics, 44(2), pp. (278-286)
- [11] Utz J. Pape, Sven Rahmann, Fengzhu Sun and Martin Vingron. (2008). "Compound Poisson approximation of the number of occurrences of a position frequency matrix (PFM) on both strands", Journal of Computational Biology, V. 15(6), pp. (547-564).
- [12] Wyshak, G. (1974). "A program for estimating the parameters of the truncated negative binomial distribution", Journal of the Royal Statistical Society Series, 23, pp. (87-91).

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$p(S_N = s) = \begin{cases} \alpha & , s = 0 \\ \frac{(1-\alpha)^2 p}{(1-\alpha q)^2} \left(\frac{\alpha p}{1-\alpha q} \right)^{s-1} & , s \geq 1 \end{cases}$$

With the following moments:

i) Mean: $E(S_N) = p$ (3.22)

ii) Variance: $Var(S_N) = p^2 \left[\frac{2\alpha}{(1-\alpha)} + \frac{q}{p} \right]$ (3.23)

Comparison between model III, IV and V:

From the relations of the mean and variance we can compute the coefficient of variation $C(S_N) = \sqrt{\frac{V(S_N)}{E(S_N)^2}}$ for each model, as it shown we draw the compound Bernoulli –Puri and

Rubin distribution (non-truncated, left-truncated and right-truncated for different k). And it is easy to compare between them. Take $\alpha = 0.1, 0.2, 0.3$ and 0.4 , $p = 0.3, 0.7$ in fig. (3.2) and fig. (3.3) respectively.

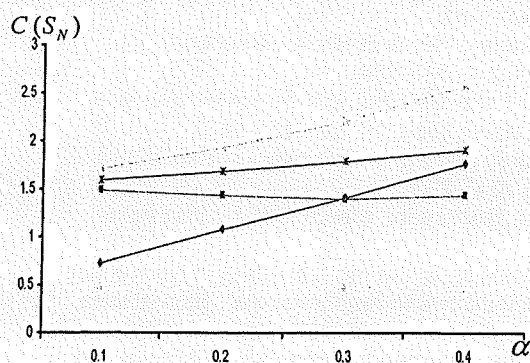


Fig. (3.2)

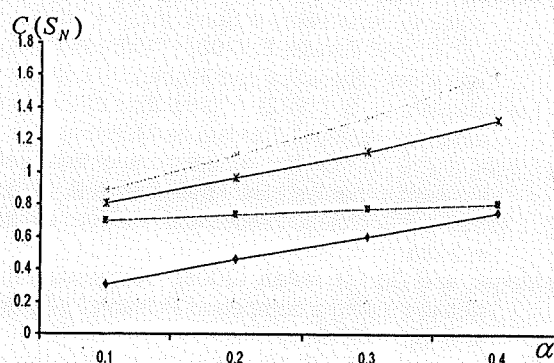


Fig. (3.3)

Now, the best model is the compound Bernoulli-left-truncated Puri and Rubin for different α especially when we substitute k by a large number.

Conclusion:

After we present all the models in the two sections, the compound distributions of sums of negative binomial and (left or right) truncated Puri and Rubin distributions, For each model, the p.g.f., p.m.f. and the corresponding moments such as the mean, and variance are calculated and more generally the descending factorial moments. Then we compare between the groups using the coefficient of variation for each model.

Regarding to the compound distributions of sums of negative binomial distribution section we had shown the compound binomial (left truncated, right truncated and non truncated) – negative binomial distributions using graphs in comparison among them, we find that the three

3.4.1. Moments:

Now we can drive the moments of the compound sum S_N using the previous probability generating function as the following:

i) Mean:

$$E(S_N) = P'_{S_N}(z) \Big|_{z=1} = \frac{p[1-(k+1)\alpha^k + k\alpha^{k+1}]}{1-\alpha^k + \alpha^{k+1}} \quad (3.19)$$

ii) Variance:

$$\begin{aligned} Var(S_N) &= P''(1) + P'(1) - \overline{P'(1)}^2 \\ &= \frac{p^2 \left[3\alpha - (k^2 - k - 2)\alpha^k + (2k^2 - 4k - 6)\alpha^{k+1} - (k^2 - 3k - 2)\alpha^{k+2} \right.}{(1-\alpha)(1-\alpha^k + \alpha^{k+1})^2} \\ &\quad \left. - (k+1)\alpha^{2k} + 3(k+1)\alpha^{2k+1} - (3k+2)\alpha^{2k+2} + k\alpha^{2k+3} - 1 \right] \\ &\quad + \frac{p[1-(k+1)\alpha^k + k\alpha^{k+1}]}{1-\alpha^k + \alpha^{k+1}} \end{aligned} \quad (3.20)$$

$$Var(S_N) = p^2[Var(N) - E(N)] + pE(N)$$

Using the compound theorem, and relations (3.19), (3.20) it is clear that:

$$\begin{aligned} E(S_N) &= E(X_i) \cdot E(N) \\ &= \frac{p[1-(k+1)\alpha^k + k\alpha^{k+1}]}{1-\alpha^k + \alpha^{k+1}} \end{aligned}$$

$$\begin{aligned} Var(S_N) &= Var(X_i) E(N) + Var(N) (E(X_i))^2 \\ &= p^2[Var(N) - E(N)] + pE(N) \\ &= \frac{p^2 \left[3\alpha - (k^2 - k - 2)\alpha^k + (2k^2 - 4k - 6)\alpha^{k+1} - (k^2 - 3k - 2)\alpha^{k+2} \right.}{(1-\alpha)(1-\alpha^k + \alpha^{k+1})^2} \\ &\quad \left. - (k+1)\alpha^{2k} + 3(k+1)\alpha^{2k+1} - (3k+2)\alpha^{2k+2} + k\alpha^{2k+3} - 1 \right] \\ &\quad + \frac{p[1-(k+1)\alpha^k + k\alpha^{k+1}]}{1-\alpha^k + \alpha^{k+1}} \end{aligned}$$

3.5. Model V: Compound Bernoulli-Puri and Rubin distribution (Special case of model III and IV)

It is clear that we can obtain the compound Bernoulli-Puri and Rubin distribution as a special case from model III (the compound Bernoulli-left-truncated Puri and Rubin distribution) when we put $k=0$ or model IV (the compound Bernoulli-right-truncated Puri and Rubin distribution).

So, the p. g. f. of the compound sum $S_N = \sum_{i=1}^N X_i$ when X_i has a deterministic random

variable and N is a simple modified Puri and Rubin random variable is given by:

$$P_{S_N}(z) = \alpha + \frac{(1-\alpha)^2}{1-\alpha q} \left[q \sum_{i=0}^{\infty} \left(\frac{\alpha p z}{1-\alpha q} \right)^i + p \sum_{i=0}^{\infty} z^{i+1} \left(\frac{\alpha p}{1-\alpha q} \right)^i \right] \quad (3.21)$$

Remark 3.1:

Let $k = 1$ in model III, then it reduced to the compound Bernoulli – single left truncated Puri and Rubin with p.m.f.:

$$p(S_N = s) = p(s) = (1 - \alpha) \sum_{i=0}^{\infty} \alpha^i q^{i+1-s} p^s$$

$$\text{and moments: } E(S_N) = \frac{p}{1 - \alpha}, \quad \text{Var}(S_N) = \frac{p(1 - \alpha) - p^2(1 - 2\alpha)}{(1 - \alpha)^2}$$

3.4. Model IV: Compound Bernoulli-Right-truncated Puri and Rubin distribution

Let X_i ($i = 1, 2, \dots, N$) be an independent and identically distributed random variables each has a Bernoulli distribution with p.g.f. and moments as given in model III.

Assume N be a random variable having a simple modified right-truncated Puri and Rubin probability mass function which is in the following form:

$$p_R(n) = \begin{cases} \frac{\alpha}{1 - \alpha^k + \alpha^{k+1}} & n = 0 \\ \frac{(1 - \alpha)^2 \alpha^{n-1}}{1 - \alpha^k + \alpha^{k+1}} & n = 1, 2, \dots, k; \alpha = \text{parameter} \end{cases}$$

clearly, $\sum_{n=0}^k p_R(n) = 1$ which is a probability mass function.

The p. g. f. of the random variable N could be calculated as follows:

$$P_N(z) = \sum_{n=0}^k p_R(n) z^n = \frac{1}{(1 - \alpha^k + \alpha^{k+1})} \left[\alpha + (1 - \alpha)^2 \frac{z(1 - \alpha^k z^k)}{(1 - \alpha z)} \right] \quad (3.16)$$

And its moments are given by:

$$E(N) = \frac{1 - (k+1)\alpha^k + k\alpha^{k+1}}{1 - \alpha^k + \alpha^{k+1}}, \quad (3.17)$$

$$\text{Var}(N) = \frac{2\alpha - k^2 \alpha^k + (2k^2 - 2k - 3)\alpha^{k+1} - (k^2 - 2k - 1)\alpha^{k+2} + \alpha^{2k+1} - \alpha^{2k+2}}{(1 - \alpha)(1 - \alpha^k + \alpha^{k+1})^2}$$

Now applying the compound distribution theorem for the sum $S_N = \sum_{i=1}^N X_i$, when X_i has a Bernoulli random variable and N is a simple modified right-truncated Puri and Rubin random variable, as given in Medhi (1982), the probability generating function of the compound sum S_N is calculated as follows:

$$P_{S_N}(z) = P_N(P_{X_i}(z))$$

$$= \frac{1}{(1 - \alpha^k + \alpha^{k+1})} \left[\alpha + (1 - \alpha)^2 \frac{(q + pz)(1 - \alpha^k (q + pz)^k)}{(1 - \alpha(q + pz))} \right] \quad (3.18)$$

3.3. Model III: Compound Bernoulli-left-truncated Puri and Rubin distribution

Let X_i ($i = 1, 2, \dots, N$) be an independent and identically distributed random variables each has a Bernoulli probability mass function in the form:

$$p(x_i) = p^{x_i} q^{1-x_i}, \quad x_i = 0, 1, \quad p + q = 1$$

The probability generating function $P_{X_i}(z)$ of X_i is calculated as follows:

$$P_{X_i}(z) = \sum_{x_i=0}^1 p(x_i) \cdot z^{x_i} = q + pz \quad (3.11)$$

with moments given by:

$$E(X_i) = P'(1) = p \quad (3.12)$$

$$Var(X_i) = P''(1) + P'(1) - \overline{P'(1)}^2 = pq$$

Let N be a random variable having a simple modified left-truncated Puri and Rubin distribution with p.g.f. and moments as given in model I.

Now applying the compound distribution theorem for the sum $S_N = \sum_{i=1}^N X_i$, when X_i

has a Bernoulli random variable and N is a simple modified left-truncated Puri and Rubin random variable, as given in Medhi (1982), the probability generating function of the compound sum S_N is calculated as the following:

$$P_{S_N}(z) = (1-\alpha) \sum_{i=0}^{\infty} \alpha^i \sum_{j=0}^{\infty} q^{i+k-j} p^j z^j \quad (3.13)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$p(S_N = s) = p(s) = (1-\alpha) \sum_{i=0}^{\infty} \alpha^i q^{i+k-s} p^s$$

3.3.1. Moments:

We can get the moments of the compound sum S_N using the previous probability generating function as the following:

i) Mean:

$$E(S_N) = P'_{S_N}(z) \Big|_{z=1} = \frac{p[k + \alpha(1-k)]}{(1-\alpha)} \quad (3.14)$$

ii) Variance:

$$\begin{aligned} Var(S_N) &= P''(1) + P'(1) - \overline{P'(1)}^2 \\ &= \frac{p(1-\alpha)[k + \alpha(1-k)] - p^2[k - 2\alpha k - \alpha^2(1-k)]}{(1-\alpha)^2} \end{aligned} \quad (3.15)$$

Clearly applying the compound theorem, and equation (3.14), (3.15):

$$E(S_N) = E(X_i) \cdot E(N) = \frac{p[k + \alpha(1-k)]}{(1-\alpha)}$$

$$\begin{aligned} Var(S_N) &= V(X_i) E(N) + V(N) (E(X_i))^2 \\ &= \frac{p(1-\alpha)[k + \alpha(1-k)] - p^2[k - 2\alpha k - \alpha^2(1-k)]}{(1-\alpha)^2} \end{aligned}$$

3.2. Model II: Compound Deterministic- Puri and Rubin distribution (Special case of model I)

In model I (the compound deterministic- left truncated Puri and Rubin distribution) if we put $k = 0$ it reduced to the compound deterministic –Puri and Rubin distribution.

So, the p. g. f. of the compound sum $S_N = \sum_{i=1}^N X_i$ when X_i has a deterministic random variable and N is a simple modified Puri and Rubin random variable is given by:

$$P_{S_N}(z) = \alpha + (1 - \alpha)^2 \sum_{i=0}^{\infty} \alpha^i z^{a+ai} \quad (3.8)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$p(S_N = s) = p(s) = \begin{cases} \alpha & , s=0 \\ 0 & , s=1 \\ (1-\alpha)^2 \alpha^{\frac{s}{a}-1} & , s=a, 2a, 3a, \dots \end{cases}$$

With the following moments:

$$\text{i) Mean: } E(S_N) = a \quad (3.9)$$

$$\text{ii) Variance: } Var(S_N) = \frac{2\alpha a^2}{1-\alpha} \quad (3.10)$$

Comparison between model I and II:

Clearly from the relations of the mean and variance we can compute the coefficient of variation $C(S_N) = \sqrt{\frac{Var(S_N)}{E(S_N)^2}}$ for each model, as it shown we draw the compound deterministic –

Puri and Rubin distribution (non-truncated and left-truncated for different k). Take $\alpha = 0.1, 0.2, 0.3$ and 0.4

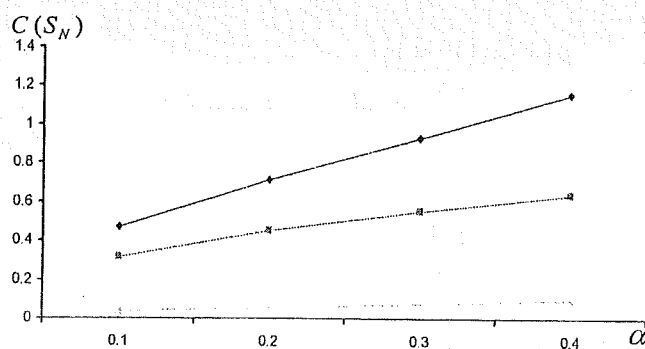


Fig. (3.1)

We see that the coefficient of variation is best for large k as shown in fig. (3.1), the upper line is for compound deterministic –Puri and Rubin distribution then for compound deterministic – left-truncated Puri and Rubin distribution $k = 1, 5$ and 10 respectively.

Let N be a random variable having a simple modified left-truncated Puri and Rubin probability mass function which is in the following form:

$$p_L(n) = (1-\alpha) \alpha^{n-k}, \quad n = k, k+1, \dots$$

clearly, $\sum_{n=k}^{\infty} p_L(n) = 1$ which is a probability mass function.

The p. g. f. of the random variable N could be calculated as follows:

$$P_N(z) = \frac{(1-\alpha) z^k}{(1-\alpha z)} \quad (3.3)$$

And its moments are given by:

$$E(N) = k + \frac{\alpha}{1-\alpha} \quad \text{and} \quad Var(N) = \frac{\alpha}{(1-\alpha)^2} \quad (3.4)$$

Consider the compound sum $S_N = \sum_{i=1}^N X_i$ when X_i has a deterministic random variable and N is a simple modified left-truncated Puri and Rubin random variable. Then as given in Medhi (1982) the probability generating function of the compound sum S_N is calculated as the following:

$$P_{S_N}(z) = (1-\alpha) \sum_{i=0}^{\infty} \alpha^i z^{ai+ak} \quad (3.5)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$p(S_N = s) = p(s) = (1-\alpha) \alpha^{\frac{s}{a}-k}, \quad s = ak, a(k+1), \dots$$

3.1.1. Moments:

We can drive the moments of the compound sum S_N using the previous probability generating function as follows:

i) Mean:

$$E(S_N) = P'_{S_N}(z) \Big|_{z=1} = a \left[k + \frac{\alpha}{1-\alpha} \right] \quad (3.6)$$

ii) Variance:

$$Var(S_N) = P''(1) + P'(1) - \overline{P'(1)}^2 = \frac{a^2 \alpha}{(1-\alpha)^2} \quad (3.7)$$

Using the compound theorem and relations (3.6), (3.7) it is clear that:

$$E(S_N) = E(X_i) \cdot E(N) = a \left[k + \frac{\alpha}{1-\alpha} \right]$$

$$Var(S_N) = Var(X_i) E(N) + Var(N) (E(X_i))^2 = \frac{a^2 \alpha}{(1-\alpha)^2}$$

Comparison between model I, II and III:

Clearly from the relations of the mean and variance we can compute the coefficient of variation $C(S_N) = \sqrt{\frac{V(S_N)}{E(S_N)^2}}$ for each model, it is easily that the three models are nearly coincide for different $\mathcal{G}$ and $\mathcal{G}_1$. Take $n = r = 20, 25$ and 40 , $(\mathcal{G} = 0.3, \mathcal{G}_1 = 0.1)$ in fig. (2.1), and $(\mathcal{G} = 0.8, \mathcal{G}_1 = 0.5)$ in fig. (2.2).

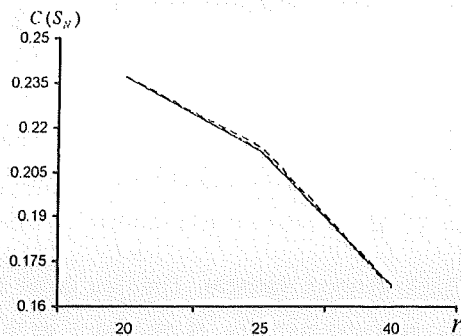


Fig. (2.1)

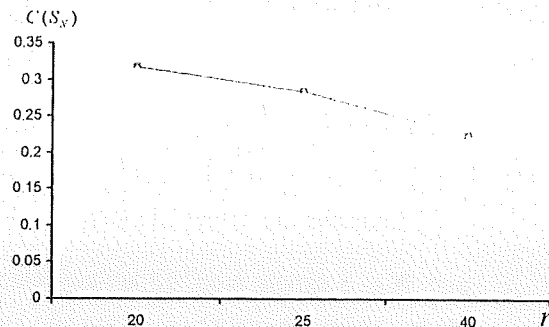


Fig. (2.2)

We see that the coefficient of variation is best when $\mathcal{G}, \mathcal{G}_1$ are small for the three models as shown in fig. (2.1) and fig. (2.2) for different n, r .

3. Compound distribution of sums of Puri and Rubin distributions

Puri and Rubin distribution is one of the interesting distributions in engineering. We introduce five different models such as the compound deterministic-Puri and Rubin (left-truncated or non-truncated) distributions, the compound Bernoulli – Puri and Rubin (left truncated, right truncated or non-truncated) distributions.

3.1. Model I: Compound Deterministic-left-truncated Puri and Rubin distribution

Let X_i ($i = 1, 2, \dots, N$) be an independent and identically distributed random variables each has a deterministic probability mass function in the form:

$$p(x_i) = \begin{cases} 1 & x_i = a \\ 0 & x_i \neq a \end{cases}$$

The probability generating function of X_i is calculated as follows:

$$P_{X_i}(z) = \sum_{x=0}^{\infty} p(x) \cdot z^x = z^a \quad (3.1)$$

And its moments are:

$$\begin{aligned} E(X_i) &= P'(1) = a \\ \text{Var}(X_i) &= P''(1) + P'(1) - \overline{P'(1)}^2 = 0 \end{aligned} \quad (3.2)$$

Lemma 2.1:

$$\sum_{s=m-nj}^{ni} s_{(k)} C_{m-s}^{nj} (1-g)^{ni-s} g^s = g^{m-nj} \sum_{m=0}^k (-1)^m \frac{C_m^k (ni)_{(k)} (nj)_{(m)} (1-g)^m}{(ni)_{(m)}}$$

with $(ni)_{(0)} = (nj)_{(0)} = 1$

2.2.2. Factorial moments:

$$\begin{aligned} \mu_{(k)} = E(s_{(k)}) &= \sum_{s=0}^{\infty} s_{(k)} p(s) = \sum_{s=0}^{\infty} s(s-1)\cdots(s-k+1) p(s) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^r g_1^r (1-g_1)^i (1-g^n)^{-i} (-1)^{i-j} \sum_{s=ni-nj}^{ni} s_{(k)} C_{m-s}^{nj} (1-g)^s g^{ni-s} \end{aligned}$$

using Lemma 2.1

$$\therefore \mu_{(k)} = \sum_{m=0}^k (-1)^m C_m^k (1-g)^m \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^r g_1^r (1-g_1)^i (1-g^n)^{-i} (-g^n)^{i-j} \frac{(ni)_{(k)} (nj)_{(m)}}{(ni)_{(m)}}$$

2.3. Model III: Compound binomial-negative binomial distribution (Special case of model I and II)

It is clear that we can obtain the compound binomial-negative binomial distribution as a special case from model I (the compound single left-truncated binomial-negative binomial distribution) when we include $x=0$ or $x=n$ in model II (the compound single right-truncated binomial-negative binomial distribution).

The p. g. f. of the compound sum $S_N = \sum_{i=1}^N X_i$, when X_i has binomial random variable with parameters (n, g) and N is a negative binomial random variable with parameters (r, g_1) , is:

$$P_{S_N}(z) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} C_i^{r+i-1} C_j^r g_1^r (1-g_1)^i g^j (1-g)^{ni-j} z^j \quad (2.14)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N given by:

$$p(S_N = s) = p(s) = g_1^r \left[\frac{g}{1-g} \right]^s \sum_{i=0}^{\infty} C_i^{r+i-1} C_s^{ni} [(1-g_1)(1-g)^n]^i, \quad s = 0, 1, 2, \dots$$

With the following moments:

$$\text{i) Mean: } E(S_N) = n g \frac{r(1-g_1)}{g_1} \quad (2.15)$$

$$\text{ii) Variance: } Var(S_N) = n g (1-g) \frac{r(1-g_1)}{g_1} + n^2 g^2 \frac{r(1-g_1)}{g_1^2} \quad (2.16)$$

Also, it is clear that the factorial moments can be found as:

$$\mu_{(k)} = \sum_{i=0}^{\infty} C_i^{r+i-1} (ni)_{(k)} g^k g_1^r (1-g_1)^i, \quad k = 1, 2, \dots$$

With moments:

$$\begin{aligned} E(X_i) &= \frac{n\vartheta(1-\vartheta^{n-1})}{1-\vartheta^n} \\ \text{Var}(X_i) &= \frac{n\vartheta(1-\vartheta)}{(1-\vartheta^n)^2} [1-n\vartheta^{n-1}+(n-1)\vartheta^{n-1}] \end{aligned} \quad (2.10)$$

Let N be a random variable having a negative binomial distribution with p.g.f. and moments as given in model I.

Consider the compound sum $S_N = \sum_{i=1}^N X_i$ when X_i has a single right-truncated binomial

random variable with parameters (n, ϑ) and N is a negative binomial random variable with parameters (r, ϑ_1) . Then as given in Medhi (1982) the probability generating function of the compound sum S_N is calculated as the following:

$$\begin{aligned} P_{S_N}(z) &= \left[\frac{\vartheta_1}{1-(1-\vartheta_1)(1-\vartheta^n)^{-1}[(1-\vartheta)+\vartheta z]^n - (\vartheta z)^n} \right]^r \\ &= \sum_{s=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^i C_{m-s}^{nj} \vartheta_1^r (1-\vartheta_1)^i (1-\vartheta^n)^{-i} (-1)^{i-j} (1-\vartheta)^{m-s} \vartheta^s z^s \end{aligned} \quad (2.11)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$p(S_N = s) = p(s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^i C_{m-s}^{nj} \vartheta_1^r (1-\vartheta_1)^i (1-\vartheta^n)^{-i} (-1)^{i-j} (1-\vartheta)^{m-s} \vartheta^s$$

2.2.1. Moments:

We can drive the moments of the compound sum S_N using the previous probability mass function as follows:

$$\text{i) Mean: } E(S_N) = \frac{nr(1-\vartheta_1)(\vartheta-\vartheta^n)}{\vartheta_1(1-\vartheta^n)} \quad (2.12)$$

ii) Variance:

$$\text{Var}(S_N) = \frac{n\vartheta(1-\vartheta)}{(1-\vartheta^n)^2} [1-\vartheta^n-n\vartheta^{n-1}+n\vartheta^n] \frac{r(1-\vartheta_1)}{\vartheta_1} + \frac{n^2\vartheta^2(1-\vartheta^{n-1})^2}{(1-\vartheta^n)^2} \frac{r(1-\vartheta_1)}{\vartheta_1^2} \quad (2.13)$$

Now by applying the compound theorem, its moments as given in Medhi (1982) and relations (2.12), (2.13) we have:

$$E(S_N) = E(X_i) \cdot E(N) = \frac{nr(1-\vartheta_1)(\vartheta-\vartheta^n)}{\vartheta_1(1-\vartheta^n)}$$

$$\text{Var}(S_N) = V(X_i) E(N) + V(N) (E(X_i))^2$$

$$= \frac{n\vartheta(1-\vartheta)}{(1-\vartheta^n)^2} [1-\vartheta^n-n\vartheta^{n-1}+n\vartheta^n] \frac{r(1-\vartheta_1)}{\vartheta_1} + \frac{n^2\vartheta^2(1-\vartheta^{n-1})^2}{(1-\vartheta^n)^2} \frac{r(1-\vartheta_1)}{\vartheta_1^2}$$

ii) Variance:

The variance of the compound sum S_N could be calculated as the following:

$$\begin{aligned} E(S_N^2) &= \sum_{s=0}^{\infty} s^2 p(s) \\ &= \frac{n\theta(1-\theta)}{1-(1-\theta)^n} \frac{r(1-\theta_1)}{\theta_1} + \frac{n^2\theta^2}{1-(1-\theta)^n} \frac{r(1-\theta_1)}{\theta_1} + \frac{n^2\theta^2}{[1-(1-\theta)^n]^2} \frac{r(1-\theta_1)}{\theta_1^2} \\ &\quad + \frac{n^2\theta^2}{[1-(1-\theta)^n]^2} \frac{r^2(1-\theta_1)^2}{\theta_1^2} - \frac{n^2\theta^2}{[1-(1-\theta)^n]^2} \frac{r(1-\theta_1)}{\theta_1} \end{aligned} \quad (2.7)$$

Thus from (2.6) and (2.7) we get the variance by:

$$\begin{aligned} Var(S_N) &= E(S_N^2) - \overline{E(S_N)}^2 \\ &= \frac{n\theta}{1-(1-\theta)^n} \frac{r(1-\theta_1)}{\theta_1} \left[(1-\theta) + n\theta + \frac{n\theta}{[1-(1-\theta)^n]\theta_1} - \frac{n\theta}{1-(1-\theta)^n} \right] \end{aligned} \quad (2.8)$$

Now by applying the compound theorem, its moments as given in Medhi (1982) and relations (2.6), (2.8) we have the following results:

$$E(S_N) = E(X_i) \cdot E(N) = \frac{n\theta}{1-(1-\theta)^n} \frac{r(1-\theta_1)}{\theta_1}$$

$$\begin{aligned} Var(S_N) &= V(X_i) E(N) + V(N) (E(X_i))^2 \\ &= \frac{n\theta}{1-(1-\theta)^n} \frac{r(1-\theta_1)}{\theta_1} \left[(1-\theta) + n\theta + \frac{n\theta}{[1-(1-\theta)^n]\theta_1} - \frac{n\theta}{1-(1-\theta)^n} \right] \end{aligned}$$

2.1.2. Factorial moments:

In this section we calculate the k^{th} descending factorial moment where:

$$\begin{aligned} \mu_{(k)} &= E(s_{(k)}) = \sum_{s=0}^{\infty} s(s-1)\cdots(s-k+1) p(s) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^i \theta_1^i (1-\theta_1)^i [1-(1-\theta)^n]^{-i} (-1)^{i-j} (1-\theta)^{ni-nj} (nj)_{(k)} \end{aligned}$$

2.2. Model II: Compound single right-truncated binomial-negative binomial distribution

Let X_i ($i = 1, 2, \dots, N$) be an independent and identically distributed random variables each has a single right-truncated binomial probability mass function in the form:

$$p(x_i) = C_{x_i}^n \frac{\theta^{x_i} (1-\theta)^{n-x_i}}{1-\theta^n}, \quad x_i = 1, 2, \dots, n-1, \quad 0 \leq \theta \leq 1$$

The probability generating function of X_i is given as follows:

$$P_{X_i}(z) = \sum_{x=0}^{\infty} p(x) \cdot z^x = \frac{[(1-\theta) + z\theta]^n - [z\theta]^n}{1-\theta^n} \quad (2.9)$$

Let N be a random variable having a negative binomial distribution with probability mass function:

$$p(N = k) = C_{r-1}^{k+r-1} g_1^r (1 - g_1)^k, \quad k = 0, 1, 2, \dots, \quad 0 \leq g_1 \leq 1$$

and probability generating function:

$$P_N(z) = \sum_{k=0}^{\infty} C_{r-1}^{k+r-1} g_1^r [(1 - g_1)z]^k = \left[\frac{g_1}{1 - (1 - g_1)z} \right]^r \quad (2.3)$$

Its moments are:

$$E(N) = \frac{r(1 - g_1)}{g_1} \quad \text{and} \quad \text{Var}(N) = \frac{r(1 - g_1)}{g_1^2} \quad (2.4)$$

Now consider the compound sum $S_N = \sum_{i=1}^N X_i$ when X_i has a single left-truncated

binomial random variable with parameters (n, g) and N is a negative binomial random variable with parameters (r, g_1) . Then as given in Medhi (1982) the probability generating function of the compound sum S_N is calculated as the following:

$$\begin{aligned} P_{S_N}(z) &= P_N(P_{X_i}(z)) \\ &= \left[\frac{g_1}{1 - (1 - g_1)P_{X_i}(z)} \right]^r \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{s=0}^{nj} C_i^{r+i-1} C_j^i C_s^{nj} g_1^r (1 - g_1)^i [1 - (1 - g)^n]^{-i} (-1)^{i-j} (1 - g)^{nj-s} g^s z^s \end{aligned} \quad (2.5)$$

Thus the probability mass function $p(S_N = s)$ of the random variable S_N is:

$$\begin{aligned} p(S_N = s) &= p(s) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i^{r+i-1} C_j^i C_s^{nj} g_1^r (1 - g_1)^i [1 - (1 - g)^n]^{-i} (-1)^{i-j} (1 - g)^{nj-s} g^s, \quad s = 0, 1, 2, \dots \end{aligned}$$

2.1.1. Moments:

We can drive the moments of the compound sum S_N using the previous probability mass function as follows:

i) Mean:

The expected value of S_N could be calculated as:

$$E(S_N) = \sum_{s=0}^{\infty} s p(s) = \frac{n g}{1 - (1 - g)^n} \frac{r(1 - g_1)}{g_1} \quad (2.6)$$

binomial, Poisson and geometric random variables. Leon and Perron (2003) introduced extremal properties of sums of Bernoulli random variables. Utz, et al (2008), introduced compound Poisson approximation of the number of occurrences of a position frequency matrix (PFM) on both strands. Moger and Aalen,(2008),used the compound Poisson frailty distribution with random scale to introduce regression models for infant mortality data in Norwegian siblings. Vellaisamy and Upadhye,(2009), studied the sums of compound negative binomial and gamma random variables. Stathis and Georgio (2008), introduced further improved recursions for a class of compound Poisson distributions.

The objective of this paper is to introduce some interesting models of compound distributions. For each model, we calculate the probability mass function of the compound sums and its moments such as the mean, the variance and the more general relation of the descending factorial moments. The paper really consists of two sections.

Section 2, includes three different models such as the compound single left-truncated binomial – negative binomial distribution, the compound single right-truncated binomial – negative binomial distribution and the compound binomial – negative binomial distribution as special case. And the p.g.f., p.m.f. and the corresponding moments mean and variance, and the descending factorial moments are calculated.

In section 3, we present five models. These models are: the compound deterministic –Puri and Rubin (left-truncated or non-truncated as special case) distributions, the compound Bernoulli – Puri and Rubin (left-truncated, right-truncated or non-truncated as special case) distributions. Also, their p.g.f., p.m.f. and moments such as the mean and the variance are given.

2. Compound distribution of sums of negative binomial distributions

It has shown by different investigators that the negative binomial distribution provides an excellent model because it has a variance larger than the mean. Some of the common examples of such behaviors are the frequency distributions of plant density. We present three different models such as the compound binomial (non truncated, single left-truncated or single right-truncated) – negative binomial distributions.

2.1. Model I: Compound single left-truncated binomial-negative binomial distribution

Let X_i ($i = 1, 2, \dots, N$) be an independent and identically distributed random variables each has a single left-truncated binomial probability mass function in the form:

$$p(x_i) = C_{x_i}^n \frac{\theta^{x_i} (1-\theta)^{n-x_i}}{1-(1-\theta)^n}, \quad x_i = 1, 2, \dots, n, \quad 0 \leq \theta \leq 1$$

The probability generating function of X_i is given as follows:

$$P_{X_i}(z) = \sum_{x=0}^{\infty} p(x) \cdot z^x = \frac{[(1-\theta) + z\theta]^n - (1-\theta)^n}{1-(1-\theta)^n} \quad (2.1)$$

With moments:

$$E(X_i) = \frac{n\theta}{1-(1-\theta)^n} \quad (2.2)$$

$$Var(X_i) = \frac{n\theta[(1-\theta) - [n\theta + (1-\theta)](1-\theta)^n]}{[1-(1-\theta)^n]^2}$$

Compound distributions of sums of negative binomial and truncated Puri-Rubin distributions

El-Arshy, S. M^{*}, Imam, T.A^{*} and Salim, A. M.^{**}

Abstract: Based on the compound distributions, we introduce some new models of the compound sum $S_N = \sum_{i=1}^N X_i$, when N is random variable. These models are, the compound binomial (single left-truncated, single right-truncated or non truncated as special case) – negative binomial distributions, the compound deterministic-Puri and Rubin (left-truncated or non truncated as special case) distributions, the compound Bernoulli – Puri and Rubin (left truncated, right truncated or non truncated as special case) distributions, which are of interest in many fields such as industry, queueing theory, the reliability theory, electronics, medicine as in hospital administration and insurance risk theory. For each model, the p.g.f., p.m.f. and the corresponding moments such as the mean, and variance are calculated and more generally the descending factorial moments.

Keyword: Compound distribution, truncated distribution, discrete distribution, negative binomial distribution, binomial distribution, deterministic distribution, Bernoulli distribution, Puri and Rubin distribution, probability generating function, moments and factorial moments.

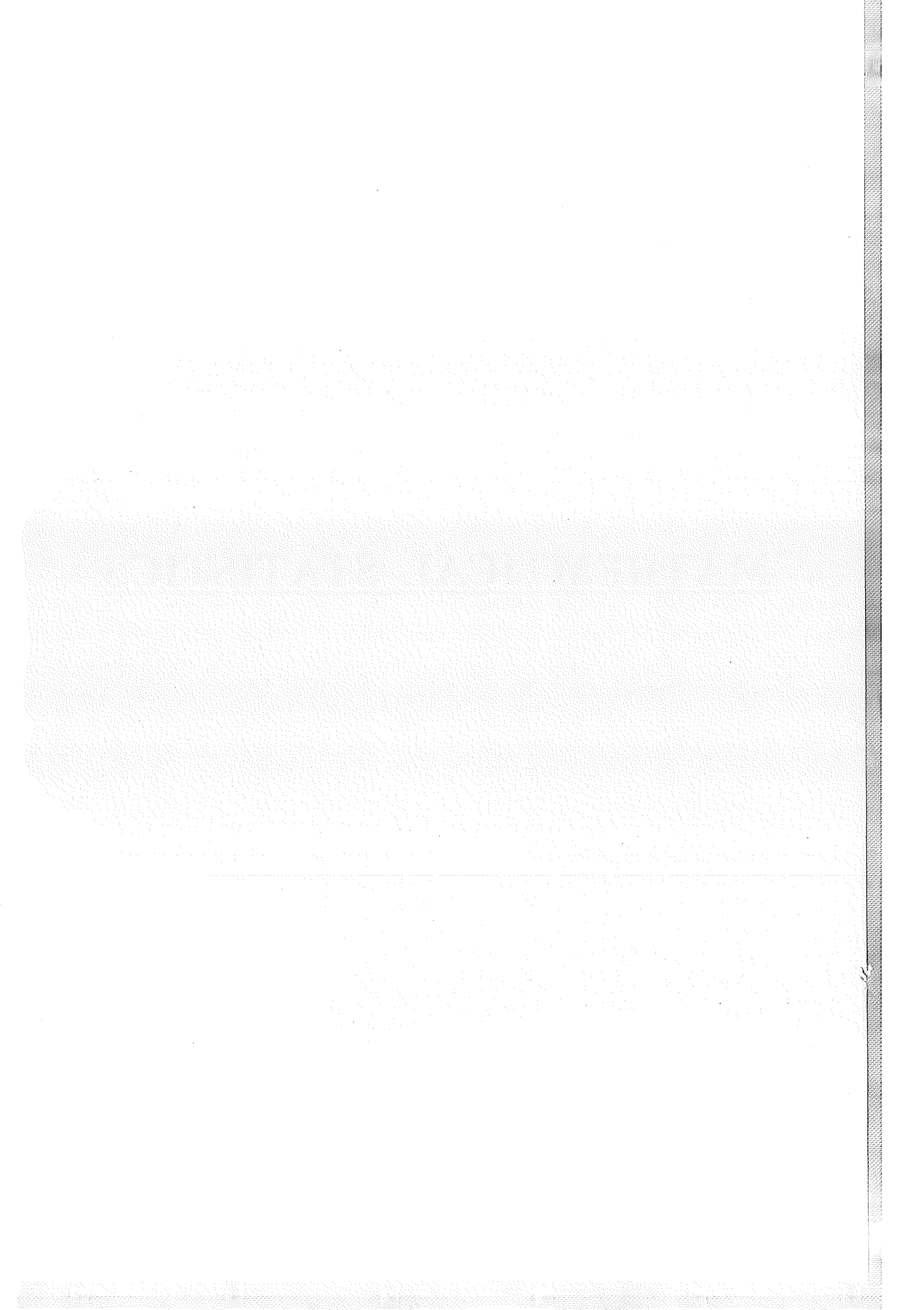
1. Introduction

The study of compound distribution of sum $S_N = \sum_{i=1}^N X_i$ of random variables is of valuable interest especially in the truncated distribution case. It takes the attention of many researchers in many fields such as industry, queueing theory, reliability theory of electronics, medicine as in hospital administration and insurance risk theory and many other fields.

Many researchers were interested in studying the distributions of both the complete (non truncated) and the truncated distributions of the compound sums and its applications, see for example, Alaa, (2003), who introduced compound distributions of sums of single truncated

^{*} Faculty of science Al-Azhar University, Math. Dept.

^{**} Modern Academy For Engineering and Technology, Basic Science Dept.



MATHEMATICAL STATISTICS

1. The first part of the paper is devoted to a general discussion of the problem of the existence of solutions of the system of equations

which is the system of equations of the theory of the motion of a rigid body.

2. In the second part of the paper the problem of the existence of solutions of the system of equations

is considered. It is shown that the system of equations has solutions if and only if the conditions

are satisfied. The conditions are satisfied if and only if the conditions

are satisfied. The conditions are satisfied if and only if the conditions

are satisfied. The conditions are satisfied if and only if the conditions

are satisfied. The conditions are satisfied if and only if the conditions

are satisfied. The conditions are satisfied if and only if the conditions

Index
MATHEMATICAL STATISTICS

1	Compound Distributions of Sums of Negative Binomial and Truncated Puri-Rubin Distributions El-Arishy,S,M, Imam , T.A , Salim, A.M	1-14
2	Fractional L-Moments Elsayed Habib , Abd-Elftah Kandile , Enayat Abd-Elrazik	15-30

Index
APPLIED STATISTICS

1	Estimating Indices of Coincident and Leading Indicators: An Application to Egypt Adel A.El-Sherif, Ghazal A. Amer , Sayed M. El-Sayed	31-48
2	The Small Sample Properties of the Coefficient of Determination and Its Adjusted Version in Linear Regression Models Ghazal A. Amer , Alaa Abd-Elaziz , Soha Othman	49-58
3	On The Single-Index Model Theory Magda M. Haggag	59-72
4	A Mixed Two-Stage Randomized Response Model: An Application to The Phenomenon of Informal Marriage Sahar Adel , Farid N.Guirguis , Mahmoud Mansour	73-87
5	Face Recognition Using Integration Between Multivariate Analysis Techniques And Artificial Neural Networks Abd Alfattah .M. Kandel , Sahar Adel , Fared N.Guirguis	88-101
6	On The Estimation Problem of Parameters of The Polynomial Regression Model When The Explanatory Variable Is Measured with Error Farid N.Guirguis	102-127

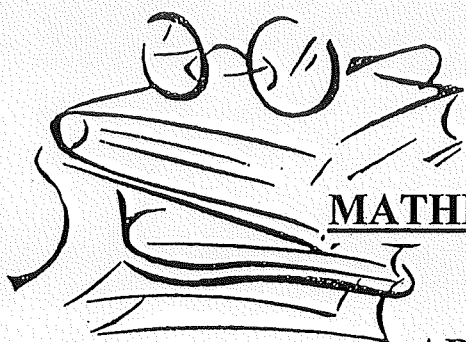


Cairo University

Institute of Statistical Studies and Research



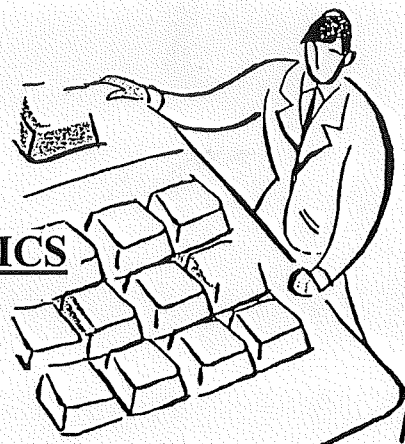
THE ANNUAL CONFERENCE
ON STATISTICS, COMPUTER SCIENCE
AND OPERATIONS RESEARCH



MATHEMATICAL STATISTICS

and

APPLIED STATISTICS



7-10, Dec ,2009