Al Kut Journal of Economics and Administrative Sciences /ISSN: 1999 -558X /ISSN Online 2707-4560/ Vol (16) Issue: 52-2024 (August)

A special issue on the proceedings of the Conference on Technology Transfer to Iraq (Capabilities, Mechanisms, and Visions)

AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

مقارنة بين الطرائق المقلصة الخطية واللاخطية لتقدير مصفوفة التباين ذات الابعاد العالية (دراسة محاكاة)

Comparison Between Linear and Nonlinear Shrinkage Methods to Estimate High Dimensional Variance Matrix

http://dx.doi.org/10.29124/kjeas.1652.7

م . م . سارة قيس عيسى[1]

كلية الامام الكاظم (ع)

للعلوم الإسلامية

أ. م . د. أحمد مهدي صالح[2]

كلية الادارة والاقتصاد

جامعة واسط

المستخلص

هناك العديد من التطبيقات الاحصائية التي تتطلب تقدير التباين او فعندما تكون ابعاد هذه المصفوفة كبيرة بالنسبة الى حجم العينة اي ان المصفوفة ذات ابعاد تكون قريبة الى حجم العينة او اكبر منها. ستكون هناك صعوبات في ايجاد تقدير جيد لها اذ ان اغلب المصفوفات بتلك الابعاد ستعاني من صعوبة ايجاد المعكوس لهذه المصفوفات . لذلك فان طرق التقدير الكلاسيكية مثل طريقة الامكان الاعظم او طريقة المربعات الصغرى ستعطي تقديرات متحيزة ويكون التقدير بعيدا عن قيمته الحقيقية في المجتمع. يهدف البحث الى التوسع في استخدام لمقدرات المقلصة لتقدير مصفوفة التباين والتباين المشترك في حالة استخدام عينات ذات ابعاد كبيرة .وهنا سيتم تقدير تلك المصفوفة باستخدام طريقة التقليص اللاخطي والخطي والمقارنة فيما بينها بالاعتماد على اصغر مربعات خطاء. حيت تم استخدام مقدر الاوراكل (Oracle Estimator) كتقدير مقلص غير خطي لمصفوفة التباين والتباين المشترك بالاضافة الى استخدام مقدرين مقلصين خطيين لتقدير المصفوفة هما (Fisher & Sun FS Estimator) و (Rao-Blackwell Ledoit-Wolf (RBLW) Estimator) حيث تم اجراء محاكاة لاحجام عينات مختلفة وبابعاد كبيرة وحساب اصغر مربعات حطاء عند ازدياد حجم العينة بالنسبة الى ابعاد مصفوفة التباين والتباين المشترك.

Abstract

There are many statistical applications that require estimation for var-covariance matrix or its inverse and when the dimensions of this matrix are relatively large to the sample size or when the dimensions of the matrix are close to the sample size or larger. There will be difficulties in finding a good estimation for it. Most Matrices with high dimension suffer from the difficulty of finding the inverse of theme. Therefore the classical methods of estimation such least squares or maximum likelihood will give biased estimators and far from its true value. This search aims at expanding usage of shrinkage estimation to estimate the var- covariance matrix in the case of using samples with large dimensions.We will estimate the var-covariance matrix by using Nonlinear Shrinkage Estimator (***Oracle Estimator***) and two Linear Shrinkage Estimators (***Fisher & Sun (FS) Estimator***) and (***Rao-Blackwell Ledoit-Wolf (RBLW) Estimator***) and make comparison among them based on (MMSE) minimum mean square errors. Here we make a simulated experiment with high dimensions samples with multiple sizes and calculate MMSE as the increasing in sample size to the large dimension of var-covariance Matrix

1- مقدمة

في حالة غياب المعلومات حول القيمة الحقيقية لمصفوفة التباين والتباين المشترك (Σ) فمن الطبيعي ان ينصب الاهتمام حول ايجاد تقدير جيد وكفوء لهذه المصفوفة كونها تدخل في العديد من التطبيقات الاحصائية المهمة وعنما تزداد ابعاد تلك المصفوفة تزداد صعوبة ايجاد التقدير لها بالطرق الكلاسيكية لذلك كان من الضروري استعمال طرق اخرى غير تقليدية كالمقدرات المقلصة .

تعود مسالة تقدير مصفوفة التباين والتباين المشترك الى عام 1960 قدم الباحث Stein [15] افضل تمثيل حصل عليه من مقدر مقلص لتباين العينة . وبعد ذلك اقترحت العديد من المقدرات المقلصة ولها مقاييس اداء مختلفة فعلى سبيل المثال قدم الباحث Haff [6] مقدر يدخل ضمن اسلوب بيز الجزئي (Empirical Bayes) وكذلك اشتق الباحثان [3] Dey & Srinivasan مقدر

(Minimax) تحت دالة خسارة (Entropy) والتي قدمت بالاصل من Stein .

وتمكن الباحثان Ledoit & Wolf [12] باقتراح مقدر جديد عندما يكون حجم العينة اصغر من عدد المتغيرات في العينة تحت الدراسة حيث يعمل هذا المقدر على تصغير متوسط مربعات الخطاء (MSE) حيث يعمل هذا المقدر جيدا في ظل الحجوم الصفيرة للعينات وبازدياد ابعاد مصفوفة التباين والتباين المشترك ..تكمن أهمية البحث في إيجاد تقدير هجين يعمل بصورة جيدة في ضل الابعاد العالية.

2- مصفوفة التباين والتباين المشترك وأهمية تقديرها

أن تقدير مصفوفة التباين والتباين المشترك أو معكوسها يعتبر من أساسيات الكثير من التطبيقات الإحصائية لأنها تعكس العلاقة بين المتغيرات تحت الدراسة حيث أن مصفوفة التباين والتباين المشترك لمتجه المتغيرات العشوائية $\underline{X}$ هو [5]

$$\underline{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_p \end{bmatrix}$$

فان مصفوفة التباين والتباين المشترك هي

$$cov(\underline{X}) = E\left[\left(\underline{X} - \underline{\mu}\right)\left(\underline{X} - \underline{\mu}\right)'\right]$$

$$= \begin{bmatrix} var(x_1) & cov(x_1, x_2) & cov(x_1, x_3) & \cdots\cdots & cov(x_1, x_p) \\ cov(x_2, x_1) & var(x_2) & cov(x_2, x_3) & \cdots\cdots & cov(x_2, x_p) \\ cov(x_3, x_1) & cov(x_3, x_2) & var(x_3) & \cdots\cdots & cov(x_3, x_p) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ cov(x_p, x_1) & cov(x_p, x_2) & cov(x_p, x_3) & \cdots\cdots & var(x_p) \end{bmatrix}$$

ويرمز لها بالرمز Σ حيث تعد عملية تقديرها مهمة ويدخل مقدر $(\hat{\Sigma})$ في الكثير من المواضيع الإحصائية ومنها .

2- 1 الانحدار

أن التداخل بين تقدير مصفوفة التباين والتباين المشترك وقيمة معلمات الانحدار قد لوحظ من أمد بعيد وهناك العديد من الأمثلة التطبيقية حول استعمال هذه المصفوفات في عمليات التنقية للتنبؤ والتقدير [2] . و يدخل تقدير مصفوفة التباين والتباين المشترك عن طريق المربعات الصغرى والإمكان الأعظم وكذلك بواسطة انحدار الحرف Ridge Regression في التنبؤ والاختبار لعدة معالم في الانحدار .

2-2 الأساليب القياسية لمتعدد المتغيرات

عند استعمال أساليب الانحدار المتعدد بعد تقدير مصفوفة التباين والتباين المشترك أساسيا لأنة يمثل المرحلة الابتدائية (Initial Stage) في العديد من هذه التطبيقات مثل تحليل المركبات الأساسية Principal Component Analysis

والتحليل المميز الخطي Linear Discernment Analysis فلابد من وجود تقدير كفؤ لمصفوفة التباين والتباين المشترك لتقليل اخطاء التصنيف Classification Errors [5].

3-2 اختبار متعدد متغيرات

يدخل تقدير مصفوفة التباين والتباين المشترك $(\hat{\Sigma})$ في عدة اختبارات خاصة بمتعدد المتغيرات فعلى سبيل المثال أذا أردنا اختبار تساوي متجه المتوسط لاي متجه تكون أحصاءة الاختبار T^2 أذ أن [5].

$$T^2 = n\left(\underline{\overline{X}} - \underline{\mu}\right)' \left(\hat{\Sigma}\right)^{-1} \left(\underline{\overline{X}} - \underline{\mu}\right)$$

وكذلك يدخل تقدير مصفوفة التباين والتباين المشترك $(\hat{\Sigma})$ في العديد من الاختبار الأخرى .

3- بعض مجالات تطبيق البيانات ذات الابعاد الكبيرة

تظهر البيانات عالية الابعاد في الكثير من مجالات العلوم الحديثة حيث ازداد ظهورها بازدياد التطور العلمي والتكنولوجي في السنوات الاخيرة ومنها

1-3 بيانات التقنية الاحيائية

وهي من المجالات العلمية الحديثة وتظهر فية البيانات عالية الابعاد لتنوع بياناتها كالبيانات الخاصة بالعرق او التحليل العرقي DNA او بيانات الجينات والموروثات كتسلسل الاحماض الامينية في المورثات وتمتد الى سلاسل طويلة ومعقدة كذلك تظهر تلك البيانات في المجال الزراعي خصوصاً في مجال التعديل الوراثي للمحاصيل [14]

2-3 البيانات المالية

تعتبر عملية تحليل البيانات المالية كبيانات التجارة العالمية واسعار الاسهم وغيرها مهمة للغاية وتظهر بشكل ابعاد كبيرة حيث زاد الاهتمام مؤخرا بذلك النوع من البيانات في ضل حرب الهيمنة الاقتصادية بين اقطاب التجارة في العالم حيث يهتم البحث المالي بادق التفاصيل ولايهمل اي متغير ويحلل اثار وانعكاسات التغيرات المالية والاقتصادية والسياسية على عالم المال ولحاجة الباحثين الدائمة للتنبؤ بالمستقبل الحالي والاقتصادي للاسواق والمناطق التجارية [14]

3-3 بيانات صور الاقمار الاصطناعية

تتطلب عملية تحليل الصور الواردة من الاقمار الصناعية التعامل مع قاعدة بيانات واسعة من الصور كصور التنبؤ بالمناخ او الصور الفلكية وغيرها حيث تكون هذة الصور عالية الدقة وعالية الابعاد باحجام هائلة تتطلب كثيراً من الدقة والاحترافية للتعامل معها [14]

4-3 تحليل طيف الصور

ظهرت حاجة لتحليل طيف الصور لدواعي علمية وامنية كتحليل الطيف الحراري للصور بالنسبة للباحثين في مجال الفلك كما تتطورت عمليات تحليل الصور للتعرف على الاشخاص Face Detector وكذلك صور مناطق الاجهاد للمهتمين بدراسة المعادن وكذلك صور تحليل الطيف الحراري للاجسام وانتقال الحرارة Thermodynamic لذا يتوجب ان تكون هذة الصور عالية الدقة مما يجعل بيانتها كبيرة باحجام هائلة Massive Size لتحتوي على معلومات كافية لاجراء التحليل [14]

4- اصغر متوسط مربعات خطاء

قبل الدخول الى انواع مقدرات مصفوفة التباين والتباين المشترك لابد من عرض بسيط لمفهوم (MMSE) لكونة يمثل اسلوب مقارنة بين مختلف طرق تقدير مصفوفة التباين والتباين المشترك فالمقدر الافضل هو المقدر الذي يحقق اقل (MMSE) . حيث اقترح الباحثان Frost & Savarino استعمال القياس التربيعي للقياس التربيعي للمسافة من المصفوفة المقدرة والمصفوفة الحقيقية بالاستناد الى Frobenious Norm اذ انه في حالة كون المصفوفة مربعة متماثلة فان [4]

$$\|\boldsymbol{Z}\|_F^2 = \boldsymbol{Trace}\ (\boldsymbol{Z})^2$$

أي أن متوسط مربعات الخطاء سيكون

$$E\left[\left\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\right\|_F^2\right] = \boldsymbol{Trace}\left(\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\right)^2 \qquad \textbf{... (1)}$$

يكون MMSE اسلوباً جيداً للمقارنة بين مختلف مقدرات مصفوفة التباين والتباين المشترك

5-طريقة التقدير المقلصه لمصفوفة التباين والتباين المشترك

فلتكن $\{X_i\}_{i=1}^n$تمثل p من المتجهات العشوائية بمتوسط (0) وتباين $\sum$ متماثلة التوزيع أي أنها متجهات كأوسية Gaussian Vectors المطلوب هنا أيجاد المقدر $\widehat{\sum}$ الذي يعمل على جعل MMSE المعرف بالمعادلة (1) اقل مايمكن من الصعوبة حساب $\widehat{\sum}$ بدون قيود اضافية وعندئذ سوف يجد الباحث نفسه مقيد بنوع من المقدرات الذي يجعله يستخدم طريقة التقليص Shrinkage كما أشار الباحثان Ledoit & Wolf [12]

أن المقدر الاعتيادي لمصفوفة التباين والتباين المشترك هو $\hat{S}$

$$\widehat{\boldsymbol{S}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i' \qquad \textbf{... (2)}$$

آذ أن هذا المقدر هو مقدر غير متحيز

$$E(\hat{S}) = \Sigma$$

ويمثل مقدر الإمكان الأعظم أيضا عندما (n>p) ولكنة ليس بالضرورة يحقق اصغر MSE بسبب التباين العادي وبسبب شروط ازدياد P لذلك من الناحية المنطقية أن يؤخذ مقدار أخر لتقدير $\sum$ وهو المقدر المعروف بالمعادلة (3)

$$\widehat{\boldsymbol{F}} = \frac{\boldsymbol{Tr}\,(\widehat{\boldsymbol{S}})}{\boldsymbol{p}} * \boldsymbol{I} \qquad \textbf{.... (3)}$$

حيث أن I هي مصفوفة الوحدة بدرجة P وهذا المقدر يقلل التباين على حساب التحيز أذا يزداد تحيز هذا المقدر بزيادة P [13]

أن الحل المعقول للربط بين التحيز الصغير والتباين الصغير يتحقق من خلال التقليص بين $(\widehat{S}\ , \widehat{F})$ مما ينتج عنه مقدر جديد يسمى مقدر مختلط (Mix Estimator) وغالباً ما يفضل تسميته Shrinkage كما هو معروف وهو[13] .

$$\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{1} - \widehat{\boldsymbol{\rho}})\, \widehat{\boldsymbol{S}} + \widehat{\boldsymbol{\rho}}\, \widehat{\boldsymbol{F}} \qquad \textbf{... (4)}$$

وهذا المقدر يعتمد على معلمة الربط $(\widehat{P})$ وهي معلمة واقعه بين الصفر والواحد ويمكن الإشارة إلى المقدر $\hat{F}$ بأنة صف التقليص Shrinkage Target سيتم تقدير معلمة الربط او معلمة الخلط اي تجعل MSE المعرفة بالمعادلة (1) اقل مايمكن .

6- مقدر الاوراكل

يعتبر مقدر الاوراكل من المقدرات المثلى اللاخطية لمصفوفة التباين والتباين المشترك عند ازدياد ابعاد تلك المصفوفة [13] اذ يعتمد هذا المقدر على استعمال المعامل الامثل غير العشوائي الذي يعمل على تصغير متوسط مربعات الخطاء اي ان $\widehat{\Sigma}_0$ Oracle Estimator هو الحل للمعادلة التالية

$$\boldsymbol{Min}\ \left[\boldsymbol{E}\left\{\left\|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{0}} - \widehat{\boldsymbol{\Sigma}}\right\|_{\boldsymbol{F}}^{\boldsymbol{2}}\right\}\right]$$

s.t

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{0}} = (\boldsymbol{1} - \widehat{\boldsymbol{\rho}})\, \widehat{\boldsymbol{S}} + \widehat{\boldsymbol{\rho}}\, \widehat{\boldsymbol{F}} \qquad \textbf{....(5)}$$

اذ ان $\hat{S}$, $\widehat{F}$ كما في المعادلتين (2) (3) وان قيمة P المثلى يمكن الحصول عليها بتطبيق النظرية التالية [12]

نظرية (1)

فليكن P من المتجهات والمعروفة $\{x_i\}_{i=1}^{n}$ وان $\hat{S}$ هو تقدير لمصفوفة التباين والتباين المشترك لهذه المتغيرات فإذا كانت $(\ (x_i)^n\ )$ متغيرات مستقلة ومتماثلة التوزيع توزيعاً طبيعياً فان حل المعادلة (5) هو

$$\boldsymbol{\rho_0} = \frac{E\{Tr\,(\Sigma-\widehat{S})(\widehat{F}-\widehat{S})\}}{E\{\|\widehat{S}-\widehat{F}\|_F^2\}} \qquad \text{... (6)}$$

أذا أن:

$$E\{Tr\,(\Sigma-\widehat{S})(\widehat{F}-\widehat{S})\} =$$

$$\frac{Tr(\Sigma)}{p}\,E\{Tr(\widehat{S})\} - \frac{E\{Tr^2(\widehat{S})\}}{p} - E\{Tr(\Sigma\,\widehat{S})\} + E\{Tr(\widehat{S}^2)\} \qquad \text{....(7)}$$

وأيضا:

$$E\{\|\widehat{S}-\widehat{F}\|_F^2\}$$

$$= E\{Tr(\widehat{S}^2)\} - 2E\{Tr(\widehat{S}\,\widehat{F})\} + E\{Tr(\widehat{F}^2)\}$$

$$= E\{Tr(\widehat{S}^2)\} - \frac{E\{Tr^2(\widehat{S})\}}{p} \qquad \text{...(8)}$$

وباستعمال التعاريف ألتاليه

$$E\{Tr(\widehat{S})\} = Tr(\Sigma) \qquad \text{...(9)}$$

$$E\{Tr(\widehat{S}^2)\} = \frac{n+1}{n}\,Tr(\Sigma^2) + \frac{1}{n}\,Tr^2(\Sigma) \qquad \text{.... (10)}$$

$$E\{Tr^2(\widehat{S})\} = Tr^2(\Sigma) - \frac{2}{n}\,Tr(\Sigma^2) \qquad \text{...(11)}$$

وعلية سيكون تقدير ρ_0 كالتالي

$$\boldsymbol{\rho_0} = \frac{(1-2/p)\,Tr(\Sigma^2)+Tr^2(\Sigma)}{(n+1-2/p)\,Tr(\Sigma^2)+(1-2/p)Tr^2(\Sigma)} \qquad \text{...(12)}$$

وتتحقق المعادلة (12) في ضل شروط التوزيع الطبيعي فقط آذ عندما يقترب توزيع المعاينة من التوزيع الطبيعي فان الصيغة (12) وتقرب من الصيغة (6) .

8- مقدر FS

اقترح الباحثان ***Fisher & Sun*** [5] مقدرا لمعامل التقليص $\hat{\rho}$ بالاعتماد على مفاهيم الجبر الخطي لحساب التوقعات حيث قام الباحثان بتطوير مقدر Lediot & Wolf [ ] الذي كان بالصورة التالية حيث قدم الباحثان Lediot & Wolf التوضيحات التالية [14].

$$E\{\|S-F\|_F^2\} = E\{\|S-\Sigma+\Sigma-F\|_F^2\}$$

$$= E\{\|S-\Sigma\|_F^2\} + E\{\|\Sigma-F\|_F^2\} + 2E\{\langle S-\Sigma, \Sigma-F\rangle\}$$

$$= E\{\|S-\Sigma\|_F^2\} + E\{\|\Sigma-F\|_F^2\} + 2\langle E[S-\Sigma], \Sigma-F\rangle$$

وبزيادة عدد المتغيرات أي بازدياد ابعاد مصفوفة التباين والتباين المشترك Σ افترض الباحثان ان $\Sigma = E(S)$ مما يجعل الحد الأخير مساويا الى الصفر أي ان [5] .

$$E\{\|S-F\|_F^2\} = E\{\|S-\Sigma\|_F^2\} + E\{\|\Sigma-F\|_F^2\} \quad \ldots(13)$$

وقدم الباحثان المقدر عن طريق النظرية التالية

نظرية (1) : اذا كانت لدينا المعادلة التالية

$$min_{\rho}\, E\left\{\left\|\widehat{\Sigma}-\Sigma\right\|_F^2\right\}$$

$$S.t\ \widehat{\Sigma} = (1-\rho)S + \rho F$$

حيث ان $\hat{\Sigma}$ هو تقدير مصفوفة التباين والتباين المشترك وتكون قيمته مساوية الى القيد فأن قيمة ρ التي تحقق الدالة وتجعل قيمتها اقل ما يمكن هي بالصورة التالية

$$\widehat{\rho} = \frac{E\{\|S-\Sigma\|_F^2\}}{E\{\|S-F\|_F^2\}} \quad \ldots(14)$$

و قام الباحثان Fisher & Sun [5] باستخدام دالة المسافة التربيعية كما هو في معادلة رقم (1) و استخدام S كما هي في المعادلة رقم (6) و F كما هي في المعادلة رقم (2-13) وكذلك تم استخدام نتيجة التوقعات التالية لإيجاد مقدر يحسن من المقدر في المعادلة رقم وكالتالي[].

$$E\{Tr(S)\} = Tr(\Sigma) \qquad \ldots(15)$$

$$E\{Tr(S^2)\} = \frac{n+1}{n}\,Tr(\Sigma^2) + \frac{1}{n}\,Tr^2(\Sigma) \qquad \ldots(16)$$

$$E\{Tr^2(S)\} = Tr^2(\Sigma) + \frac{2}{n}\,Tr(\Sigma^2) \qquad \ldots(17)$$

وعليه سيكون

$$E\{\|S-\Sigma\|_F^2\} = E\{\|S\|^2\} - 2E\{\langle S,\Sigma\rangle\} + \frac{Tr(\Sigma^2)}{p}$$

$$= \frac{n+1}{np}Tr(\Sigma^2) + \frac{1}{np}Tr^2(\Sigma) - \frac{2}{p}Tr(\Sigma^2) + \frac{1}{p}Tr(\Sigma^2)$$

$$\ldots(18)$$

وللسهولة استخدم الباحثان الصيغ التالية.

$$a_1 = \frac{1}{p}Tr(\Sigma) \qquad \ldots(19)$$

$$a_2 = \frac{1}{p}\,Tr(\Sigma^2) \qquad \ldots(20)$$

وعليه سيكون.

$$\boldsymbol{E\{\|S-\Sigma\|_F^2\} = \frac{n+1}{n}a_2 + \frac{p}{n}a_1^2 - 2a_2 + a_1^2}$$

$$\boldsymbol{= \frac{1}{n}a_2 + \frac{p}{n}a_1^2} \qquad \dots (21)$$

اما بالنسبة الى المقام في المعادلة (14) فتم الحصول على نتائج مشابه الى أعلاه حيث ان

$$\boldsymbol{E\{\|S-I\|_F^2\} = E\{\|S\|_F^2\} - 2E\{\langle S, I\rangle\} + \|I\|_F^2}$$

$$\boldsymbol{= \frac{n+1}{np}Tr(\Sigma^2) + \frac{1}{np}Tr^2(\Sigma) - \frac{2}{p}Tr(\Sigma) + 1}$$

$$\boldsymbol{= \frac{n+1}{n}a_2 + \frac{p}{n}a_1^2 - 2a_1 + 1} \qquad \dots (22)$$

وعليه سيكون معامل التقليص بالصورة التالية.

$$\boldsymbol{\hat{\rho}_{FS} = \frac{\frac{1}{n}a_2 + \frac{p}{n}a_1^2}{\frac{n+1}{n}a_2 + \frac{p}{n}a_1^2 - 2a_1 + 1}} \qquad \dots (23)$$

وعليه سيكون مقدر FS لمصفوفة التباين والتباين المشترك بالصورة التالية.

$$\boldsymbol{\hat{\Sigma}_{FS} = (1 - \hat{\rho}_{FS})S + \hat{\rho}_{FS}I} \qquad \dots (24)$$

ومن الملاحظ ان معامل التقليص هنا هو دالة بدلالة Σ مصفوفة التباين والتباين المشترك وبالتالي سيكون مقدر FS لمصفوفة التباين والتباين المشترك $\hat{\Sigma}_{FS}$ يعتمد بصورة مباشرة على مصفوفة التباين والتباين المشترك للمجتمع التي تكون في اغلب الأحيان غير معلومة لذلك لابد من الحصول على مقدر لمصفوفة التباين والتباين المشترك من بيانات العينة اذ من الممكن الحصول على مقدر لمعامل التقليص بدلالة قياسات العينة وكالتالي .

فعندما $n, p \to \infty$, $p/n \to c$ اذ ان $0 < c < \infty$ تم الحصول على المقدرات التالية[14] .

$$\hat{a}_1 = \frac{1}{p} Tr(\hat{S}) \qquad \dots (25)$$

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)p}\left[Tr(\hat{S}^2) - \frac{1}{n} Tr^2(\hat{S})\right] \qquad \dots (26)$$

وعلية سيكون مقدر معامل التقليص FS بدلالة قياسات العينة بالصيغة التالية.

$$\hat{\rho}_{FS2} = \frac{\frac{1}{n}\hat{a}_2 + \frac{p}{n}\hat{a}_1^2}{\frac{n+1}{n}\hat{a}_2 + \frac{p}{n}\hat{a}_1^2 - 2\hat{a}_1 + 1} \qquad \dots (27)$$

اذ ان $\hat{a}_1$, $\hat{a}_2$ تكون حسب المعادلتين (26) , (25) . وعليه سيكون مقدر FS لمصفوفة التباين والتباين المشترك بدلالة قياسات العينة بالصورة التالية.

$$\hat{\Sigma}_{FS2} = (1 - \hat{\rho}_{FS2})S + \hat{\rho}_{FS2} I \qquad \dots (28)$$

9-مقدر RBLW

قدم الباحثان Ledot wolf تحسيناً لمقدر (FS) تحت شروط التوزيع الطبيعي حيث ان فلسفة (RBLW) من حقيقة ان المقدر لمصفوفة التباين والتباين المشترك ($\sum$) في ظل التوزيع الطبيعي هو مقدر كافي Sufficient ومبدئياً فان المقدر FS هو دالة ليس فقط ل $\hat{s}$ ولكن الى كماليات أخرى لذلك تم تحسين المقدر من خلال نظرية Rao – Blackwell theorem حيث تنص النظرية أذا كانت g(x) هو مقدر للمعلمة Q فان التوقع الشرطي لــ g(x) بوجود T(X) حيث أن T هو

مقدر كافي Sufficient لن يكون ابداً اسواء من المقدر الأصلي g(x) دالة خسارة بربيعية وان تطبيق النظرية على المقدر L W سيؤدي إلى النتيجة التالية [1]

لتكن $\{(x_i)^n\}$ P من المتغيرات العشوائية المستقلة من التوزيع الطبيعي وكالمصفوفة التباين والتباين المشترك هي ($\sum$) ومقدر من العينة $\hat{S}$ ((العينة الخاصة بالمتغيرات $\{(x_i)^n\}$))

فان التوزيع الشرطي لمقدر المصفوفة LW هو

$$\hat{\Sigma}_{RBLW} = E\left[\frac{\hat{\Sigma}_{LW}}{\hat{S}}\right] \qquad \dots (29)$$

$$\hat{\Sigma}_{RBLW} = (1 - \hat{\rho}_{RBLW})\hat{S} + \hat{\rho}_{RBLW}\hat{F} \qquad \dots (30)$$

علماً أن :

$$\hat{\rho}_{RBLW} = \frac{(n-2)/n \cdot Tr(\hat{S}^2) + Tr^2(\hat{s})}{(n+2)[Tr(\hat{S}^2) - Tr^2(\hat{S})/p]} \qquad \ldots(31)$$

10- المحاكاة

نظرا لصعوبة الحصول على بيانات باابعاد كبيرة في الواقع العملي لذا لجاء الباحث الى اسلوب المحاكاة وهنا تم تقسيم المحاكاة الى قسمين

اولا هو ان تكون مصفوفة التباين والتباين المشترك Σ هي ناتج العملية الجزئية المتزايدة لحركة براون Increment Process of Fractional Brownian Motion (FBM) واستخدمت هذه الطريقة لضمان خصائص الكاوسية Gaussian Properties حيث تكون مصفوفة التباين والتباين المشترك Positive Definite [11] والتي تكون حسب الصيغة

$$\Sigma_{ij} = \frac{1}{2}\left[(|i-j|+1)^{2h} - 2|i-j|^{2h} + (|i-j|-1)^{2h}\right] \qquad \ldots(32)$$

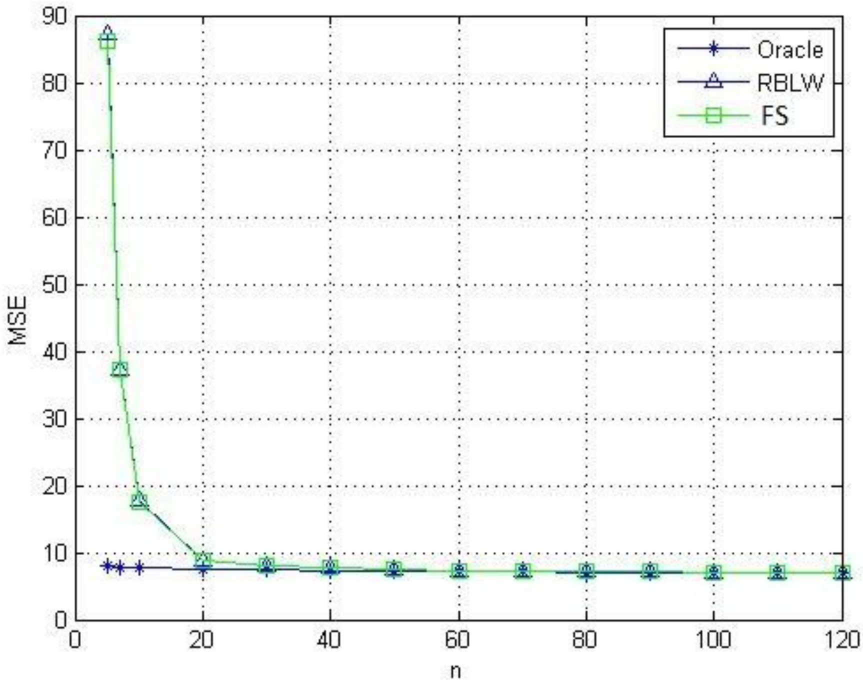
حيث ان $h \in [0.5 , 1]$ وتسمى بمعلمة هورس Hurst Parameter حيث يتم اختيار قيم عشوائية لهذا المؤشر تتراوح قيمها في الفترة [0.5 , 1] وهنا قام الباحث باختيار قيمتين هما [0.8 , 0.6]حيث كررت التجربة 100 مرة وباستعمال برنامج MATLAB كانت النتائج كالتالي

جدول (1)

متوسط مربعات الخطاء للطرائق الثلاث عندما h=0.6

	Minimum Mean Squares Error , h = 0.6			
		Oracle	**FS**	**RBLW**
Sample Size	**5**	**7.9488**	**87.1621**	**86.2258**
	7	**7.8026**	**37.1207**	**37.2164**
	10	**7.7036**	**17.6649**	**17.5952**
	20	**7.4946**	**8.6877**	**8.7310**
	30	**7.4093**	**7.9047**	**7.9226**
	40	**7.3474**	**7.6815**	**7.6961**

	50	**7.2739**	**7.4270**	**7.4280**
	60	**7.2367**	**7.3222**	**7.3221**
	70	**7.1651**	**7.2461**	**7.2511**
	80	**7.1007**	**7.1536**	**7.1559**
	90	**7.0749**	**7.1362**	**7.1384**
	100	**6.9914**	**7.0394**	**7.0380**
	110	**6.9215**	**6.9636**	**6.9629**
	120	**6.8969**	**6.9434**	**6.9437**



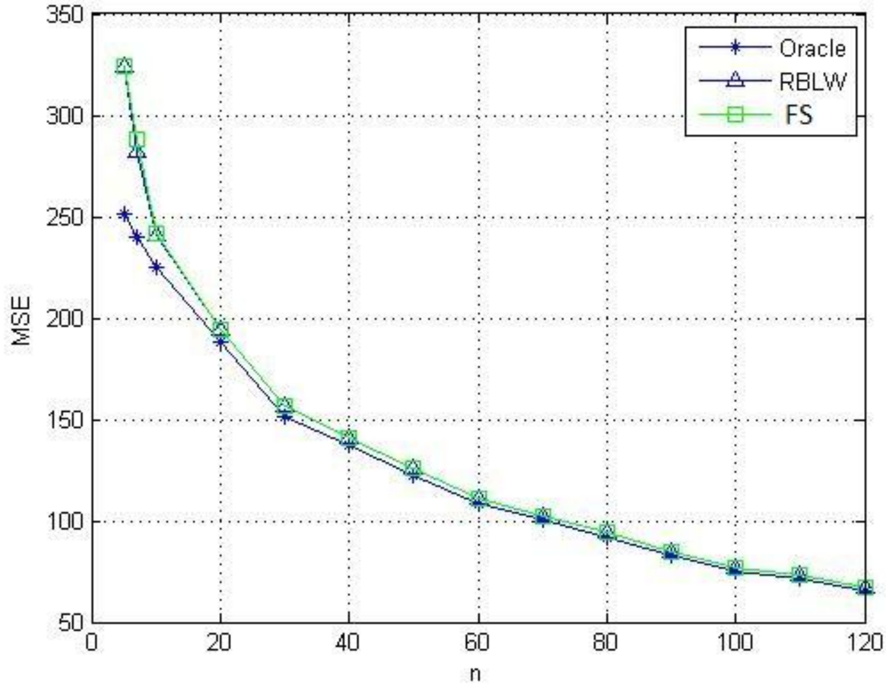
شكل (1)

متوسط مربعات الخطاء للطرق الثلاثة عندما h = 0.6

جدول (2)

متوسط مربعات الخطاء للطرائق الثلاث عندما h=0.8

	Minimum Mean Squares Error , h = 0.8			
Sample Size		**Oracle**	**FS**	**RBLW**
	5	**251.2848**	**324.4587**	**324.4834**
	7	**240.1501**	**281.6623**	**287.8923**
	10	**225.3423**	**240.4072**	**241.7484**
	20	**188.5768**	**194.3619**	**194.4208**
	30	**151.4428**	**156.4365**	**156.6997**
	40	**137.0056**	**140.9980**	**141.0939**
	50	**122.1011**	**125.6100**	**125.6228**
	60	**108.1176**	**110.7878**	**110.8200**
	70	**100.3358**	**102.5587**	**102.5657**
	80	**91.2554**	**94.2333**	**94.2301**
	90	**82.8337**	**84.6840**	**84.7077**
	100	**75.1605**	**77.1729**	**77.1598**
	110	**71.6810**	**73.3021**	**73.3211**
	120	**65.6447**	**66.6918**	**66.7315**



شكل (2)

متوسط مربعات الخطاء للمقدرات الثلاثة عندما h=0.8

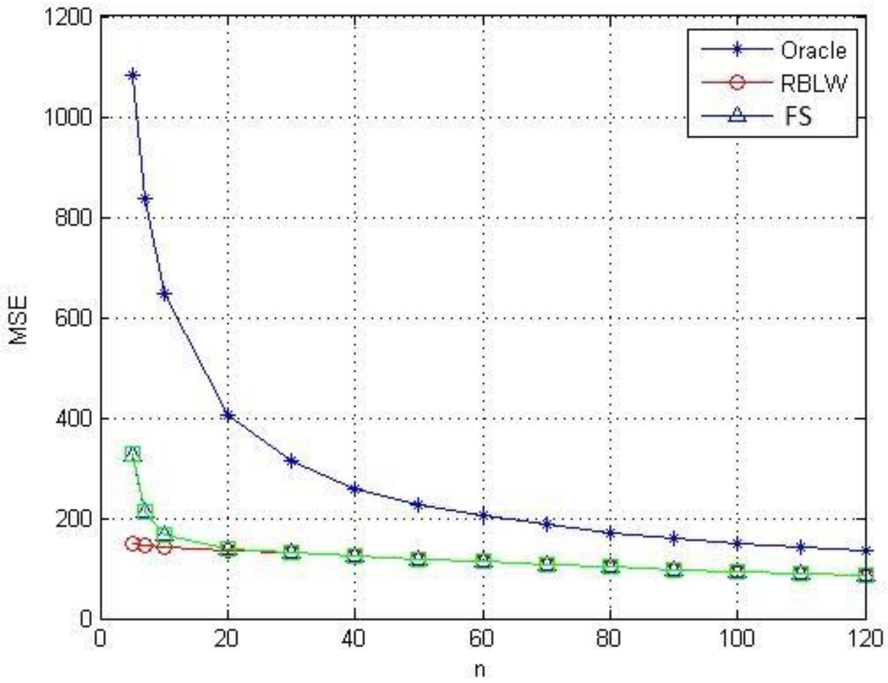
حيث تم اختيار حجوم عينات مختلفة تبداء من 5 الى 120 وحساب قيم MMSE وفق المعادلة (1) للطرق الثلاثة و الى 100 من المتغيرات حيث تكون مصفوفة التباين والتباين المشترك ذات ابعاد 100 x 100 والثاني تم فيه توليد متغيرات تتوزع توزيعا طبيعيا بمتوسط صفر وتباين مساوي الى واحد حسب طريقة بوكس ميللر **Box– Muller** واحتسبت من تلك المتغيرات مصفوفة التباين والتباين المشترك للعينة Σ وكررت التجربة 100 مرة فكانت النتائج كالتالي.

جدول (3)

متوسط مربعات الخطاء للمقدرات الثلاثة بمتغيرات ولدت حسب طريق بوكس – ميللر Box-Muller بمتوسط صفر وتباين مساوي الى الواحد الصحيح

	Minimum Mean Squares Error			
Sample Size		**Oracle**	**FS**	**RBLW**
	5	**1.0821**	**0.3259**	**0.3266**

	7	**0.8383**	**0.2107**	**0.2120**
	10	**0.6468**	**0.1669**	**0.1667**
	20	**0.4052**	**0.1390**	**0.1390**
	30	**0.3144**	**0.1312**	**0.1312**
	40	**0.2595**	**0.1244**	**0.1244**
	50	**0.2274**	**0.1182**	**0.1182**
	60	**0.2069**	**0.1134**	**0.1134**
	70	**0.1866**	**0.1076**	**0.1076**
	80	**0.1696**	**0.1022**	**0.1022**
	90	**0.1602**	**0.0975**	**0.0975**
	100	**0.1503**	**0.0935**	**0.0935**
	110	**0.1416**	**0.0899**	**0.0899**
	120	**0.1334**	**0.0858**	**0.0858**



شكل (3)

متوسط مربعات الخطاء للمقدرات الثلاثة بمتغيرات ولدت حسب طريق بوكس – ميللر Box-Muller بمتوسط صفر وتباين مساوي الى الواحد الصحيح

9- الاستنتاجات والتوصيات

من خلال الشكل رقم (1) و (2) يتضح ان مقدر الاوراكل Oracle اللاخطي تكون الامثل عندما يكون حجم العينة صغيرا بالنسبة الى عدد المتغيرات حيث يتفوق هذا المقدر على مقدري (RBLW) , (FS) حيث يمتلك اقل متوسط مربعات خطاء وبازدياد حجم العينة يتضح ان المقدرات تقترب من بعضها . ومن ملاحظة الشكل رقم (3) تتضح افضلية مقدر(RBLW) عندما يكون حجم العينة صغير بالنسبة الى عدد المتغيرات تحت افتراضات التوزيع الطبيعي .. ومن هذا تتضح افضلية المقدرات المقلصة اللاخطية على المقدرات المقلصة الخطية في الحالات العامة حيث تتفوق المقدرات المقلصة الخطية على المقدرات المقلصة اللاخطية فقط تحت فرضيات التوزيع الطبيعي..

يوصي الباحثان بالتوسع في استعمال المقدرات المقلصة اللاخطية لتقدير مصفوفة التباين والتباين المشترك واقتراح مقدرات جديدة اخرى للتقدير في ظل ظهور البيانات عالية الابعاد في مختلف مجالات الحياة

كذلك يوصي الباحثان بالتوسع في استعمال اساليب احصائية اخرى للتقدير كالاساليب الحصينة واللامعلمية لتقدير مصفوفة التباين والتباين المشترك ذات الابعاد الكبيرة .

المصادر

1- Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. Ann. Statist. 36 199–227.

2- Buhlmann, P. and Van De Geer.(2011).Statistics for High-dimensional Data Methods, Theory and Applications.Sprenger Press ..

3- Dey, D. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, The Annals of Statistics, vol. 13, no. 4, pp. 1581-1591.

4- Frost, P. A. and Savarino, J. E.(1986). An empirical Bayes Approach to Portfolio selection. Journal of Finincial and Quantitave Analysis, 21: 293- 305

5- Fujikoshi, Y. , Ulyanov, V. , Shimizo, R.(2011). Multivariate Statistics : High-Dimensional and Large-Sample Approximations. Wiley Series in Probability and Statistics

6- Haff. L, (1980). Empirical Bayes estimation of the multivariate normal covariance matrix," The Annals of Statistics, vol. 8, no. 3, pp. 586-597.

7- Horel, A. E. & Kennard, R. W.(1970). Ridge regression; biased estimation for nonorthogonal problems. Technometrics 12, 55-82

8- James .W and Stein .C,(1956). Estimation with quadratic loss," in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, p. 361, University of California press,.

9- Johnstone, I. M. (2007). High dimensional statistical inference and random matreces. Int. Cong. Mathematicians, vol. I. pp. 307- 333. Zurich, Switzerland: European Mathematical Society

10- Johnstone, I. M. and Titterington, D. M. (2009). Statistical Challenges of high-dimensional data. Philosophical Transactions of The Royal Society. 367, 4237-4253

11- Laland, W. , Taqqu, M. , Willinger, W., and Wilson, D.(2002). On the self-similar nature of Ethernet traffic (extended version)," Networking, IEEE/ACM Transactions on, vol. 2, no. 1, pp. 1-15

12- - Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices, Journal of Multivariate Analysis, vol. 88, no.2, pp. 365-411

13- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics. Vol. 40, No. 2, 1024–1060

14- Pourahmadi, M. (2013). High-Dimensional Covariance Estimation: With High-Dimensional Data. Wiley Series in Probability and Statistics.

15- Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39the Annual Meeting IMS. Atlanta, Georgia

16- Stoica, P., Jian, L., Xumin, Z., and Guerci, J. (2008).On using a priori knowledge in space-time adaptive processing," IEEE transactions on signal processing, vol. 56, no. 6, pp. 2598-2602,