Uma comparação de estimadores de encolhimento linear e não linear para a matriz de covariância

Palavras-Chave: ALTA-DIMENSÃO; SHRINKAGE; REGULARIZAÇÃO; MONTE CARLO.

Autores:

ISABELLA NASCIMENTO PERES DA SILVA, UNICAMP - IMECC

Prof(ª). Dr(ª). CARLOS TRUCÍOS, UNICAMP - IMECC

INTRODUÇÃO:

A matriz de covariância desempenha um papel crucial em estatística, econometria e ciência de dados. Sua estimação é essencial, pois grandes erros de estimação resultam em modelos preditivos imprecisos, desempenho insatisfatório fora da amostra e decisões sub-ótimas. Embora a matriz de covariância amostral tenha um desempenho satisfatório em dimensões pequenas e moderadas, tende a fornecer estimativas imprecisas e matrizes mal condicionadas em grandes dimensões, comprometendo o desempenho de qualquer método que a utilize.

Diversos estimadores de encolhimento (shrinkage) têm surgido como uma alternativa a este problema, mas a escolha do estimador mais apropriado para um problema específico ainda não é clara. Assim, experimentos de Monte Carlo são realizados para identificar se os métodos são eficazes na prática.

METODOLOGIA

Os estimadores de encolhimento podem ser divididos em lineares e não lineares. Os métodos lineares são definidos, na sua forma geral, como:

$$\Sigma_n^* = \rho_1 \delta_n + \rho_2 S_n$$

Em que S_n denota a matriz de covariância amostral (amos), δ_n denota uma matriz *target* e as constantes ρ_1 e $\rho_2 \in [0, 1]$ são obtidas pela otimização de:

$$min_{\rho_1 \rho_2}\left[\left|\left|\Sigma_n^* - \Sigma_n\right|\right|^2\right]$$

Diferentes escolhas para δ_n levarão a diferentes estimadores: Ledoit e Wolf (2004a) propõem utilizar uma matriz diagonal cujos elementos são todos iguais à média das variâncias amostrais (ls1).

Ledoit e Wolf (1995) propõem utilizar uma matriz cujos elementos na diagonal são todos iguais e os elementos fora da diagonal também são todos iguais (ls2). Ledoit e Wolf (2004a) propõem utilizar uma matriz cuja diagonal é igual à variância amostral e a correlação entre os pares é sempre a mesma (ls3). Ledoit e Wolf (1995) propõem utilizar uma matriz cujos elementos na diagonal são iguais à variância amostral (ls4). Ledoit e Wolf (2004b) propõe utilizar a matriz diagonal unitária (ls5).

Propostas mais flexíveis foram posteriormente propostas, sendo estas: estimador de encolhimento não linear com inversão da função QuEst (nls_quest) de Ledoit e Wolf (2017), estimador linear analítico (nls_asym) de Ledoit e Wolf (2018), estimador com combinação convexa assintoticamente ótima da matriz de covariância amostral e a matriz dense target (nld) de Ledoit e Wolf (2018), ou os métodos LIS (*Linear-Inverse Shrinkage*), QIS (*Quadratic-Inverse Shrinkage*) e GIS (*Geometric-Inverse Shrinkage*) propostos em Ledoit e Wolf (2022).

A comparação é realizada utilizando experimentos de Monte Carlo com 1000 replicações e os dados são simulados seguindo uma distribuição Normal com uma determinada matriz de covariância. Os estimadores são aplicados para diferentes tamanhos amostrais. Foram utilizadas mais de 10 funções perda diferentes, apresentando aqui apenas algumas delas.

RESULTADOS E DISCUSSÃO:

As Figuras 1 e 2 apresentam os RMSE para os elementos da diagonal (Var) e fora da diagonal (Cov) da matriz de covariância. Em geral, independente da dimensão, a medida que o tamanho amostral aumenta o RMSE diminui. Os estimadores de encolhimento não sempre diminuem o RMSE referente às variâncias. De fato, apenas o ls3, ls4 e nld melhoram a estimação, exceto, quando a dimensão é pequena, em cujo caso, todos os métodos apresentaram um desempenho semelhante. Por outro lado (Fig. 2), o RMSE sempre diminui quando utilizamos estimadores de encolhimento, exceto quando o método nls é utilizado, sendo os métodos ls3 e ls4 aqueles que apresentaram melhor desempenho.

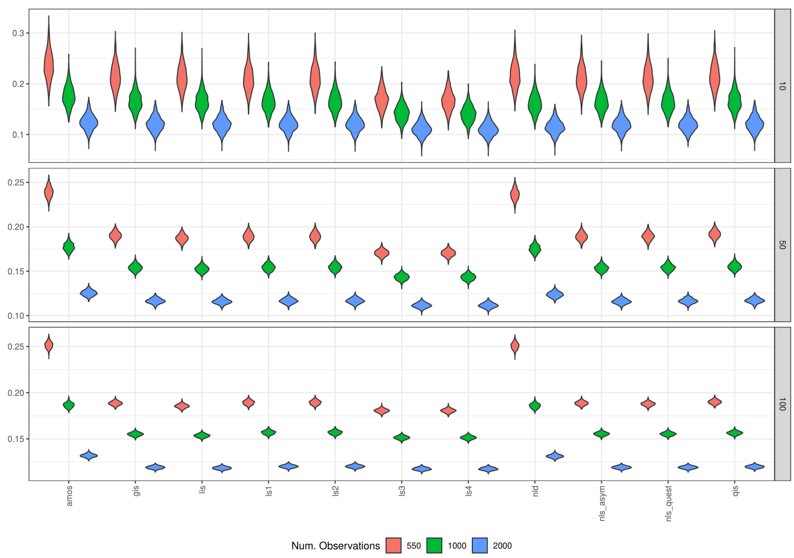

Figura 1 – RSME-Var

Nas outras funções perda avaliadas (mas não apresentadas aqui), os métodos ls3 e ls4 também foram os que apresentaram melhor desempenho.

Dentre todas as dimensões avaliadas, os estimadores de encolhimento não lineares não apresentaram grandes ganhos face aos lineares.

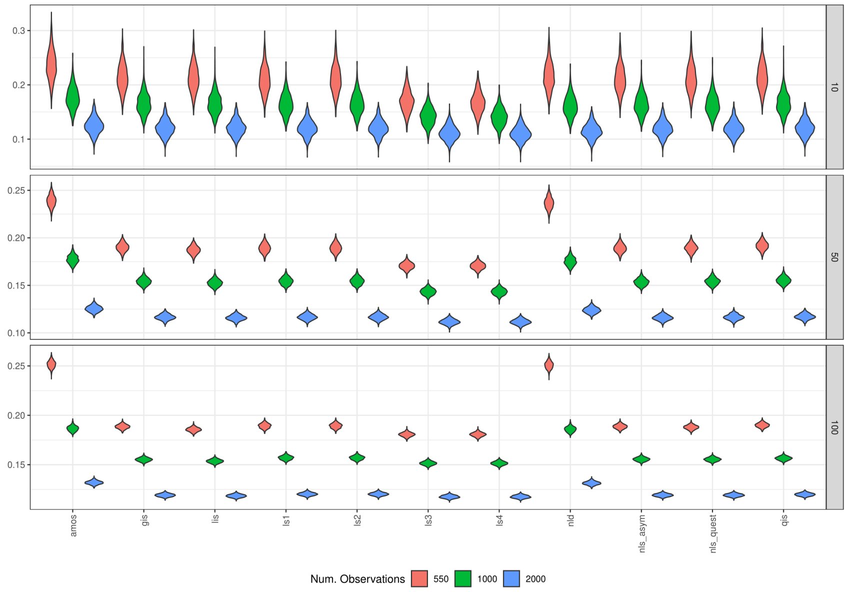

Figura 2 – RSME-Cov

CONCLUSÕES:

Em geral, estimadores de encolhimento lineares e não lineares tiveram uma performance melhor do que a apresentada pela matriz de covariância amostral. Ademais, para as dimensões avaliadas, os métodos ls3 e ls4 são os que apresentaram melhor desempenho.

BIBLIOGRAFIA

Ledoit, O. **Essays on Risk and Return in the Stock Market**. PhD thesis, Massachusetts Institute of Technology, Sloan School of Management, 1995.

Ledoit, O. e Wolf, M. **A well-conditioned estimator for large-dimensional covariance matrices**. Journal of Multivariate Analysis, 2004a, 88(2):365–411.

Ledoit, O., Wolf, M. **Honey, I shrunk the covariance matrix.The Journal of Portfolio Management**, 2004b, 30(4): 110-119.

Ledoit, O. e Wolf, M**. Numerical implementation of the QuEST function**. Computational Statistics and Data Analysis, 2017, 115(1): 199-223.

Ledoit, O. e Wolf, M. **Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli**, 2018, 24 (4B) 3791 - 3832.

Ledoit, O. e Wolf, M. **Analytical nonlinear shrinkage of large-dimensional covariance matrices**. Ann. Statist. 48 (5) 3043 - 3065. 2020. Ledoit, O. e Wolf, M. **Quadratic shrinkage for large covariance matrices**. Bernoulli, 2022, 28(3): 1519-1547.