Informatics, Data Mining, Econometrics and Financial Economics: A Connection*

EI2015-34

Chia-Lin Chang
Department of Applied Economics
Department of Finance
National Chung Hsing University
Taichung, Taiwan

Michael McAleer
Department of Quantitative Finance
National Tsing Hua University
Hsinchu, Taiwan
and
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
and
Tinbergen Institute, The Netherlands
and
Department of Quantitative Economics
Complutense University of Madrid
Spain

Wing-Keung Wong
Department of Economics
Hong Kong Baptist University
Hong Kong, China

November 2015

* For financial and research support, the first author is grateful to the National Science Council of Taiwan, the second author is most thankful to the Australian Research Council and the National Science Council of Taiwan, and the third author wishes to acknowledge the Research Grants Council of Hong Kong and Hong Kong Baptist University. The third author would also like to thank Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement.

Abstract

This short communication reviews some of the literature in econometrics and financial economics that is related to informatics and data mining. We then discuss some of the research on econometrics and financial economics that could be extended to informatics and data mining beyond the existing areas in econometrics and financial economics.

Keywords: Econometrics, Financial economics, Informatics, Data Mining, Theory, Statistics.
JEL: C01, C55, C81, C82, G14, L86, P34.

1. Introduction

There are many studies that link econometrics and financial economics to informatics and data mining. In this short communication, we discuss some of the available research that links financial economics to informatics and data mining. We then discuss some of the work related to previous contributions in financial economics that could be linked to informatics and data mining. Thereafter, we will discuss extensions of work that can be undertaken to link financial economics and econometrics to informatics and data mining.

The contributions of financial economics and econometrics to informatics and data mining consists of several areas: (1) theoretical mathematical modelling; (2) statistics and econometric modelling; (3) problems; (4) circumventing problems and refining theory; (5) simulation; (6) applications of theoretical mathematical, statistical and econometric models to some important issues in financial economics; and (7) extending the theory.

The remainder of the paper follows along the lines mentioned above.

2. Theoretical Models

In this short communication, we illustrate how to connect financial economics to informatics and data mining by using the Markowitz portfolio optimization theory developed by Markowitz (1952, 1959). The theory is a milestone in modern finance theory for optimal portfolio construction, asset allocation, utility maximization, and investment diversification.

According to the theory, portfolio optimizers respond to the uncertainty of an investment by selecting portfolios that maximize profit, subject to achieving a specified level of calculated risk or, equivalently, minimize variance, subject to obtaining a predetermined level of expected gain.

Thereafter, there have been many extensions of the theory. For example, Kroll, Levy and Markowitz (1984) extend the work to show, for various utility functions and empirical returns distributions, that the expected utility maximizer could typically do very well if they acted knowing only the mean and variance of each distribution for a case with an infinite number of alternate distributions, namely those available from the standard portfolio constraint set. Elton, Gruber, and Padberg (1976) develop a single model to solve the problem.

3. Statistical and Econometric Models

As the theoretical model of the classical MV portfolio optimization have been well developed by Markowitz (1952, 1959) and others, there have been many studies to provide the estimation of the classical MV portfolio optimization. For example, Sharpe (1967) suggests reformulating the classical MV portfolio optimization as parametric linear-programming problems, using a linear approximation to the true quadratic formula for a portfolio's risk. Elton, Gruber, and Padberg (1976) develop a single model to solve the problem.

In order to solve the portfolio optimization problem, Markowitz and Perold (1981) propose dividing the portfolio variance into "within scenario" and "between scenario" variances, and extend the results to models in which scenarios and factors both appear where factor distributions and effects may or may not be scenario sensitive. Perold (1984) develops an algorithm to estimate the parameters for large-scale portfolio optimization.

4. Problems

There are many studies that have questioned the performance of the MV optimization estimates. Some studies simply suggest disregarding the results or abandoning the entire MV approach. On the other hand, some suggest that the MV estimates do more harm than good. For example, Michaud (1989) argues that MV optimization is one of the outstanding puzzles in modern finance, and that it has yet to meet with widespread acceptance by the investment community, particularly as a practical tool for active equity investment management. He calls this puzzle the "Markowitz optimization enigma", and calls the MV optimizers "estimation-error maximizers."

Canner, Mankiw, and Weil (1994) examine popular advice on portfolio allocation among cash, bonds, and stocks. They find that the advice is inconsistent with the mutual-fund separation theorem, which states that all investors should hold the same composition of risky assets.

In contrast to the theorem, popular advisors recommend that aggressive investors hold a lower ratio of bonds to stocks than conservative investors. They explore various possible explanations of this puzzle, and conclude that the portfolio recommendations can be explained if popular advisors base their advice on the unconditional distribution of nominal returns.

The authors also find that the cost of this money illusion is small, as measured by the distance of the recommended portfolios from the mean-variance efficient frontier. Simaan (1997) finds that the estimation error is more severe in small samples (small observations relative to the number of assets), and for investors with high risk tolerance.

The MV model's lower estimation risk is most striking in small samples and for investors with a low risk tolerance. Frankfurter, Phillips and Seagle (1971) find that the portfolio selected according to the Markowitz MV criterion is likely not as effective as an equally weighted portfolio, while Zellner and Chetty (1965) and others show that the Bayesian decision rule under a diffuse prior outperforms the MV optimization.

5. Circumventing Problems and Refining Theory

There are many studies that investigate the reasons why the MV optimization estimate is far from satisfactory. For example, Best and Grauer (1991) find that an MV-efficient portfolio's weights, mean, and variance can be extremely sensitive to changes in asset means when only a budget constraint is imposed on the investment problem. They also find that a positively weighted MV-efficient portfolio's weights are extremely sensitive to changes in asset means, but the portfolio's returns are not when nonnegativity constraints are imposed. Their empirical study shows that a small increase in the mean of just a single asset drives half the securities from the portfolio, yet the portfolio's expected return and standard deviation can remain virtually unchanged.

Using an exact finite-sample statistical procedure for testing hypotheses about the weights of mean-variance efficient portfolios, Britten-Jones (1999) finds that the sampling error in estimates of the weights of a global efficient portfolio is large. On the other hand, Laloux, Cizeau, Bouchaud, and Potters (1999) find that the Markowitz portfolio optimization scheme, based on a purely historical determination of the correlation matrix, is not adequate because the lowest eigenvalues dominating the smallest risk portfolio are dominated by noise.

The traditional estimated return for the Markowitz mean-variance optimization has been demonstrated to depart seriously from its theoretic optimal return. Bai, Liu, and Wong (2009) extend the work by first proving that the phenomenon of "Markowitz optimization enigma" is natural and the estimated optimal return is always larger than its theoretic counterpart. Their findings support the idea from Laloux, Cizeau, Bouchaud, and Potters (1999) and others on this issue that the empirical correlation matrix plays an important role in the problem. They also find that estimation of the optimal return is poor due to the inadequate estimation of the asset allocations.

When the dimension of the data is large, which links the issue to Big Data, by the theory of the large dimensional random matrix, it is well known that the sample covariance matrix is not an efficient estimator of the population covariance matrix. Therefore, substituting the sample mean and covariance matrix into the MV optimization procedure will result in a serious departure of the optimal return estimate, and the corresponding portfolio allocation estimate from their theoretic counterparts when the number of the assets is large. They call this return estimate the "plug-in" return and its corresponding estimate for the asset allocation the "plug-in allocation." This phenomenon can also be called "over-prediction."

In order to circumvent this over-prediction problem, Bai, Liu, and Wong (2009) use a new method by incorporating the idea of the bootstrap into the theory of large dimensional random matrix. They develop new bootstrap-corrected estimates for the optimal return and its asset allocation, and prove that these bootstrap-corrected estimates can analytically correct the over-prediction and drastically reduce the error. They also show that the bootstrap-corrected estimate of return and its corresponding allocation estimate are proportionally consistent with their counterpart parameters.

Bai, Liu, and Wong (2009) propose a bootstrap-corrected estimator to correct the overestimation, but there is no closed form for their estimator. Thus, it has to be obtained by employing a bootstrap approach. As a result, it is difficult for practitioners to adopt the estimate in reality. In order to circumvent this limitation, Leung, Ng, and Wong (2012) develop a new estimator for the optimal portfolio return based on an unbiased estimator of the inverse of the covariance matrix and its related terms, and derive explicit formulae for the estimator of the optimal portfolio return.

As the authors derive the explicit formulae for their optimal return estimator, academics and practitioners can easily apply it in any real data analysis. In addition, they prove that their proposed estimated return is consistent when the sample size and the dimension increases to infinity proportionally.

Another advantage of their proposed improved approach to return estimation over the bootstrap-corrected estimation is that they provide not only the closed form for the optimal return estimate and its allocation estimate, but also the closed form of its standard deviation, whereas the bootstrap-corrected estimation does not. In addition, the standard deviation of their proposed improved return estimator is smaller than that of the plug-in estimate.

6. Simulation

Whenever there is any problem, researchers like to conduct simulations to show the nature of the problem before they develop the theory to explain the problem. For example, before proposing a bootstrap-corrected estimator for portfolio optimization, Bai, Liu, and Wong (2009) conduct simulations to demonstrate the over-prediction problem that the traditional plug-in estimate of the optimal portfolio return over-predicts its theoretical counterpart and the quantity of the over-prediction increases as the dimension increases. The traditional plug-in estimate of the optimal portfolio return is obtained by plugging-in the sample mean and sample covariance into their corresponding theoretical mean and covariance matrix in the optimal portfolio return formula.

When Bai, Liu, and Wong (2009) confirm that there is a serious over-prediction problem and the plug-in estimate of the optimal portfolio return over-predict its theoretical counterpart and the quantity of the over-prediction increases as the dimension increases. They develop new bootstrap-corrected estimates for the optimal return and its asset allocation, and prove that these bootstrap-corrected estimates are proportionally consistent with their counterpart parameters.

Thereafter, they conduct simulations to confirm the consistency of their proposed estimates, implying that the essence of the portfolio analysis problem could be adequately captured by their proposed estimates. Their simulations also show that their proposed method improves the estimation accuracy so substantially that its relative efficiency could be as high as 139 times when compared with the traditional "plug-in" estimate for 300 assets with a sample size of 500. The relative efficiency will be much higher for larger sample sizes and larger numbers of assets. Similar results are also obtained for its corresponding allocation estimate.

Leung, Ng, and Wong (2012) also conduct simulations to show that their proposed estimators dramatically outperform traditional estimators for both the optimal return and its corresponding allocation under different values of p/n (where p is the number of assets) and different inter-asset correlations, especially when p/n is close to 1. In addition, their simulation also shows that their proposed estimators perform better than the bootstrap-corrected estimators for both the optimal return and its corresponding allocation.

7. Real-life Applications

After developing the theory of the bootstrap-corrected estimation for portfolio optimization and conducting simulation to confirm the bootstrap-corrected estimation is superior, Bai, Liu, and Wong (2009) illustrate the superiority of their approach by comparing the estimates of the bootstrap-corrected return and the plug-in return for daily S&P500 data. They choose 500 daily data backward from December 30, 2005, for all companies listed in the S&P500 for their estimation. The authors then choose the number of assets *(p)* from 5 to 400, and for each *p*, they select *p* stocks from the S&P500 database randomly without replacement and compute the plug-in return and the corresponding bootstrap-corrected return.

They also repeat the procedure (*m* =) 10 and 100 times for checking. For each *m* and *p*, they first compute the bootstrap-corrected returns and the plug-in returns. Thereafter, they compute their averages for both the bootstrap-corrected returns and the plug-in returns.

The authors find that as the number of assets, *p*, increases, (1) the values of the estimates from both the bootstrap-corrected returns and the plug-in returns for the S&P500 increase, and (2) the values of the estimates of the plug-in returns increase much faster than those of the bootstrap-corrected returns, so that their differences become wider. These empirical findings are consistent with the theoretical discovery of the "Markowitz optimization enigma" that the estimated plug-in return is always larger than its theoretical value, and their difference becomes larger when the number of assets is large. From their empirical findings, they confirm that their proposed bootstrap-corrected return performs better.

Leung, Ng, and Wong (2012) also illustrate the applicability of their proposed estimate on the investment of the stocks from S&P500 index. They use weekly prices of the component stocks listed in the S&P500 index from 1 January 1990 to 31 December 2009. The data used in their study are adjusted by dividends and stock splits.

8. Concluding Remarks

Bai, Liu, and Wong (2009) and Leung, Ng, and Wong (2012) obtain estimates that improve the Markowitz optimization significantly, and show that their estimates are consistent with the theoretical counterparts. Researchers may well believe that these are the best available, and cannot be improved further. To state the obvious, innovative research continues to improve, so that it is always possible to improve the existing procedures.

As an example, Bai, Li, and Wong (2013) obtain another estimate that performs better than the estimates obtained by Bai, Liu, and Wong (2009) and Leung, Ng, and Wong (2012). Bai, Li, and Wong (2013) develop the best available estimates for the problem of the high-dimensional Markowitz portfolio optimization.

We also note that one could apply several techniques to analyse investing empirical issues. With more tools involved, the findings and conclusions drawn could become more interesting. For example, Hoang, Wong, and Zhu (2015) incorporate the mean-variance portfolio optimization developed by Markowitz (1952), Bai, Liu, and Wong (2009), and Leung, Ng, and Wong (2012), with the mean-risk criteria (mean-variance, mean-MVaR, and mean-Omega), and stochastic dominance analysis to assess the effect of gold quoted on the SGE in Chinese portfolios composed of stocks and government and corporate bonds.

They conclude that the safe-haven characteristic of gold in the Chinese context since it helps to improve the performance of portfolios in periods of turmoil, especially for risk seekers. They also find that risk-averse investors prefer portfolios from the efficient frontier, while risk-seekers prefer the equal-weighted portfolio. These findings can provide important information for banks, fund managers, and individual investors regarding portfolio diversification with gold in China.

Recently, Niu, McAleer, Guo, and Wong (2015) develop a theory to construct confidence interval for the new economic performance measurement (see Homm and Pigorsch, 2012) via the Aumann-Serrano index. Their contribution to the literature will be useful in comparing the performance of different assets.

References

Bai, Z.D., Li, H.X., Wong, W.K. (2013), The Best Estimation for High-Dimensional Markowitz Mean-Variance Optimization, MPRA Paper 43862, University Library of Munich, Germany.

Bai, Z.D., Liu, H.X., Wong, W.K. (2009), Enhancement of the Applicability of Markowitz's Portfolio Optimization by Utilizing Random Matrix Theory, Mathematical Finance, 19(4), 639-667.

Best, M.J., Grauer, R.R. (1991), On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results, Review of Financial Studies, 4(2), 315-342.

Britten-Jones, M. (1999), The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights, Journal of Finance, 54(2), 655-671.

Canner, N., Mankiw, N.G., Weil, D.N. (1997), An Asset Allocation Puzzle, American Economic Review, 87(1), 181-191.

Elton, E.J., Gruber, M.J., Padberg, M.W. (1976), Simple Criteria for Optimal Portfolio Selection, Journal of Finance, 31(5), 1341-1357.

Frankfurter, G.M., Phillips, H.E., Seagle J.P. (1971), Portfolio Selection: The Effects of Uncertain Means, Variances and Covariances, Journal of Financial and Quantitative Analysis, 6, 1251-1262.

Van Hoang, T.H., Wong, W.-K. Zhenzhen Zhu, Z.Z. (2015), Is Gold Different for Risk-Averse and Risk-Seeking Investors? An Empirical Analysis of the Shanghai Gold Exchange, Economic Modelling, 50, 200-211.

Homm, U., Pigorsch, C. (2012), Beyond the Sharpe Ratio: An Application of the Aumann-Serrano Index to Performance Measurement, Journal of Banking and Finance, 36, 2274-2284.

Kroll, Y., Levy, H., Markowitz, H.M. (1984), Mean-Variance Versus Direct Utility Maximization, Journal of Finance, 39, 47-61.

Leung, P.L., Ng, H.Y., Wong, W.K. (2012), An Improved Estimation to Make Markowitz's Portfolio Optimization Theory Users Friendly and Estimation Accurate, with Application on the US Stock Market Investment, European Journal of Operational Research, 222, 85-95.

Markowitz, H.M. (1952), Portfolio Selection, Journal of Finance, 7, 77-91.

Markowitz, H.M. (1959), Portfolio Selection, New York, Wiley.

Markowitz, H.M., Perold, A.F. (1981), Portfolio Analysis with Factors and Scenarios, Journal of Finance, 36, 871-877.

Michaud, R.O. (1989), The Markowitz Optimization Enigma: Is 'Optimized' Optimal?, Financial Analysts Journal, 45, 31-42.

Niu, C.Z., McAleer, M., Guo, X., Wong, W.K. (2015), Statistical analysis of an economic and risk performance measure of stocks, unpublished paper.

Perold, A.F. (1984), Large-Scale Portfolio Optimization, Management Science, 30(10), 1143-1160.

Simaan, Y. (1997), Estimation Risk in Portfolio Selection: The Mean Variance Model Versus the Mean Absolute Deviation Model, Management Science, 43(10), 1437-1446.

Zellner, A., Chetty, V.K. (1965), Prediction and Decision Problems in Regression Models from the Bayesian Point of View, Journal of the American Statistical Association, 60, 608-616.