Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology: Connections

Chia-Lin Chang¹
Michael McAleer²
Wing-Keung Wong³

¹ National Chung Hsing University
² Department of Finance, Asia University, Taiwan; Discipline of Business Analytics, University of Sydney Business School, Australia; Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; Department of Economic Analysis and ICAE, Complutense University of Madrid, Spain; Institute of Advanced Sciences, Yokohama National University, Japan
³ Department of Finance, Fintech Center, and Big Data Research Center, Asia University, Taiwan; Department of Medical Research, China Medical University Hospital, Taiwan; Department of Economics and Finance, Hang Seng Management College, Hong Kong, China; Department of Economics, Lingnan University, Hong Kong, China
Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology: Connections*

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

Michael McAleer
Department of Finance, Asia University, Taiwan
and
Discipline of Business Analytics, University of Sydney Business School, Australia
and
Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
and
Department of Quantitative Economics, Complutense University of Madrid, Spain
and
Institute of Advanced Sciences, Yokohama National University, Japan

Wing-Keung Wong
Department of Finance, Fintech Center, and Big Data Research Center, Asia University, Taiwan
and
Department of Medical Research, China Medical University Hospital, Taiwan
and
Department of Economics and Finance, Hang Seng Management College, Hong Kong, China
and
Department of Economics, Lingnan University, Hong Kong, China

January 2018

* For financial and research support, the first author is most grateful to the National Science Council, Ministry of Science and Technology (MOST), Taiwan, the second author wishes to thank the Australian Research Council and the National Science Council, Ministry of Science and Technology (MOST), Taiwan, and the third author acknowledges the National Science Council, Ministry of Science and Technology (MOST), Taiwan, Research Grants Council of Hong Kong, Asia University, China Medical University Hospital, Hang Seng Management College, Lingnan University, and Research Grants Council of Hong Kong. The third author would also like to thank Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement.
Abstract

The paper provides a review of the literature that connects Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology, and discusses some research that is related to the seven disciplines. Academics could develop theoretical models and subsequent econometric and statistical models to estimate the parameters in the associated models, as well as conduct simulation to examine whether the estimators in their theories on estimation and hypothesis testing have good size and high power. Thereafter, academics and practitioners could apply theory to analyse some interesting issues in the seven disciplines and cognate areas.

Keywords: Big Data, Computational science, Economics, Finance, Management, Theoretical models, Econometric and statistical models, Applications.

JEL: A10, G00, G31, O32.
1. Introduction

There are many studies that link Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology. Computational Science could be used to work on big data as well as small data. Thus, analysis in big data is part of Computational Science.

As many, if not all, theorems for small data do not hold for big data, and thus, analysis of big data becomes a separate topic, different from that of small data. In addition, Computational Science for both big and small data can be applied to many cognate areas, including Science, Engineering, Medical Science, Experimental Science, Psychology, Social Science, Economics, Finance, Management, and Business.

In this paper, we will discuss different types of utility functions, stochastic dominance (SD), mean-risk (MR) models, portfolio optimization (PO), and other financial, economic, marketing and management models as these topics are popular in Big Data, Computational Science, Economics, Finance, and Management. Academics could develop theory, and thereafter develop econometric and statistical models to estimate the associated parameters to analyze some interesting issues in Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology.

Academics could then conduct simulations to examine whether the estimators calculated for estimation and hypothesis testing have good size and high power. Thereafter, academics and practitioners could apply their theories to analyze some interesting issues in the seven disciplines and other cognate areas.

The plan of the remainder of the paper is as follows. In Section 2, a number of comprehensive theoretical models of risk and portfolio optimization are discussed. Alternative statistical and econometric models of risk and portfolio optimization are analyzed in Section 3. Alternative procedures for conducting simulations are examined in Section 4. A brief discussion of empirical models in several cognate disciplines is presented in Section 5. Some concluding remarks are given in Section 6.
2. Theoretical Models

It is important to commence any rigorous research in computational sciences for big data as well as small data in the areas of Economics, Finance, and Management by developing appropriate theoretical models. The authors have been developing some theories to extend those that have been discussed in a number of existing literature reviews. We discuss some of our research in the following subsections.

2.1 Portfolio optimization

The mean-variance (MV) portfolio optimization procedure is the milestone of modern finance theory for asset allocation, investment diversification, and optimal portfolio construction (Markowitz, 1952b). In the procedure, investors select 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. However, the estimates has been demonstrated to depart seriously from their theoretic optimal returns. Michaud (1989) and others have found the MV-optimized portfolios do more harm than good. Bai, et al. (2009a) have proved that this phenomenon is natural.

2.2 Cost of capital

Gordon and Shapiro (1956) develop the dividend yield plus growth model for individual firms while Thompson (1985) improve the theory by combining the model with analysis of past dividends to estimate the cost of capital and its ‘reliability’. Thompson and Wong (1991) estimate the cost of capital using discounted cash flow (DCF) methods that require forecasting dividends.

Thompson and Wong (1996) extend the theory by proving the existence and uniqueness of a solution for the cost of equity capital, and the cost of equity function is continuously differentiable. Wong and Chan (2004) have extended their theory by proving the existence and uniqueness of reliability.
2.3 Behavioral models

Barberis, Shleifer and Vishny (1998) and others use Bayesian models to explain investors’ behavioral biases by using the conservatism heuristics and representativeness heuristics in making decisions. Lam, Liu, and Wong (2010) extend the theory by developing a model of weight assignments using a pseudo-Bayesian approach that reflects investors' behavioral biases.

They use the model to explain several financial anomalies, including excess, volatility, short-run underreaction, long-run overreaction, and magnitude effects. Lam, Liu, and Wong (2012) extend their work by developing additional properties for the pseudo-Bayesian approach that reflects investors’ behavioral biases, and explain the linkage between these market anomalies and investors’ behavioral biases.

Fung, Lam, Siu, and Wong (1998) extend their work by incorporating the pseudo-Bayesian model with the impact of a financial crisis. They derive properties of stock returns during the financial crisis and recovery from the crisis.

Guo, McAleer, Wong, and Zhu (2017) extend the model by assuming that the earnings shock of an asset follows a random walk model, with and without drift, to incorporate the impact of financial crises. They assume the earning shock follows an exponential family distribution to accommodate symmetric as well as asymmetric information. By using this model setting, they develop some properties on the expected earnings shock and its volatility, and establish properties of investor behavior on the stock price and its volatility during financial crises and subsequent recovery.

Thereafter, they develop properties to explain excess volatility, short-term underreaction, long-term overreaction, and their magnitude effects during financial crises and subsequent recovery. Egozcue and Wong (2010a) extend prospect theory, mental accounting, and the hedonic editing model by developing an analytical theory to explain the behavior of investors with extended value functions in segregating or integrating multiple outcomes when evaluating mental accounting.
Whether to keep products segregated (that is, unbundled) or integrate some or all of them (that is, bundle) has been a problem of profound interest in areas such as portfolio theory in finance, risk capital allocations in insurance, and marketing of consumer products. Such decisions are inherently complex and depend on factors such as the underlying product values and consumer preferences, the latter being frequently described using value functions, also known as utility functions in economics.

Egozcue, Fuentes García, Wong, and Zitikis (2012a) develop decision rules for multiple products, which we generally call ‘exposure units’ to naturally cover manifold scenarios spanning well beyond ‘products’. The findings show, for example, that the celebrated Thaler’s principles of mental accounting hold as originally postulated when the values of all exposure units are positive (that is, all are gains) or all negative (that is, all are losses).

In the case of exposure unit mixed-sign values, decision rules are much more complex and rely on cataloging the Bell-number of cases that grow very fast, depending on the number of exposure units. Consequently, in this paper we provide detailed rules for the integration and segregation decisions in the case up to three exposure units, and partial rules for the arbitrary number of units.

We note that the theory of decision maker’s behavior developed by Egozcue and Wong (2010a) and Egozcue, Fuentes García, Wong, and Zitikis (2012a) is for marketing, and they develop a theory for consumer behavior.

2.4 Modelling different types of investors

We have been developing some theories, estimation, and testing to examine different utility functions and the preferences of different types of investors. We summarize some of the results here. Readers may refer to Sriboonchitta, Wong, Dhompongsa, and Nguyen (2009) and Bai, Guo, Li, and Wong, (2018) for further information.

2.4.1 Different types of utility functions
Lien (2008) compares the exponential utility function with its second-order approximation under the normality assumption in the optimal production and hedging decision framework. Guo, Lien, and Wong (2016) extend the theory by comparing the exponential utility function with a 2n-order approximation for any integer n. In addition, they propose an approach with an illustration to determine the smallest n that provides a good approximation.

2.4.2 Stochastic Dominance

We have been developing several theories in stochastic dominance, and discuss some here.

2.4.2.1 Behavior of risk averters and risk seekers

Wong and Li (1999) develop some properties for the convex stochastic dominance to compare the preferences of different combinations of several assets for both risk-averse and risk-seeking investors. In addition, they compare the preferences between a convex combinations of several continuous distributions and a single continuous distribution. In addition, Li and Wong (1999) develop some SD theorems for the location-and-scale family and linear combinations of random variables for risk seekers and risk averters.

Wong (2007) extends their work by introducing the first three orders of both ascending SD (ASD) and descending SD (DSD) to decisions in business planning and investment for risk-averse and risk-seeking decision makers so that they can compare both return and loss. The author provides tools to identify the first-order SD prospects and discern arbitrage opportunities that could increase their expected utility and expected wealth. Wong (2007) also introduces the mean-variance (MV) rule to decisions in business planning or investment on both return and loss for both risk-averse and risk-seeking decision makers, and show that the rule is equivalent to the SD rule under some conditions.

Chan, Clark, and Wong (2016) analyse properties of SD for both risk-averse and risk-seeking SD (RSD) for risk-seeking investors which, in turn, enables an examination of their behavior. They
first discuss the basic properties of SD and RSD that link SD and RSD to expected-utility maximization. Thereafter, they prove that a hierarchy exists in both SD and RSD relationships and that the higher orders of SD and RSD can be inferred by the lower orders of SD and RSD, but not vice-versa. Furthermore, they study the conditions in which third-order SD preferences are ‘the opposite of’ or ‘the same as’ their counterpart third-order RSD preferences.

In addition, they establish the relationship between the orders of the variances and that of the integrals for two assets, which enables us to establish certain relationships between the dominance of the variances and the second- and third-order SD and RSD for two assets under the condition of equal means. The theory developed in the paper provides a set of tools that enables investors to identify prospects for first-, second-, and third-order SD and RSD, and so enables investors to improve their investment decisions.

Another contribution in the paper is that the authors recommend checking the dominance of the means of the distributions to draw inferences for the preferences for two different assets for third-order risk averters and risk seekers. They illustrate this idea by comparing the investment behavior of both third-order risk averters and risk seekers in bonds and stocks.

Guo and Wong (2016) extend some univariate SD results to multivariate SD (MSD) for both risk averters and risk seekers, respectively, to n order for any n > 0 when the attributes are assumed to be independent and the utility is assumed to be additively separable. Under these assumptions, they develop some properties for MSD for both risk averters and risk seekers. For example, they prove that MSD are equivalent to the expected-utility maximization for both risk averters and risk seekers, respectively.

They show that the hierarchical relationship exists for MSD, and establish some dual relationships between the MSD for risk averters and risk seekers. They develop some properties for non-negative combinations and convex combinations random variables of MSD, and develop the theory of MSD for the preferences of both risk averters and risk seekers on diversification. At last, they discuss some MSD relationships when attributes are dependent, and discuss the importance and the use of the results developed in their paper.
2.4.2.2 Behavior of investors with S-shaped and reverse S-shaped utility functions

Wong and Chan, (2008) extend the work on Prospect SD (PSD) and Markowitz SD (MSD) to the first three orders, and link the corresponding S-shaped and reverse S-shaped utility functions to the first three orders. They provide experiments to illustrate each case of the MSD and PSD to the first three orders, and demonstrate that the higher order MSD and PSD cannot be replaced by the lower order MSD and PSD. Furthermore, they show that an hierarchy exists in both PSD and MSD relationships, arbitrage opportunities exist in the first orders of both PSD and MSD, and for any two prospects under certain conditions, their third order MSD preference will be ‘the opposite of’ or ‘the same as’ their counterpart third order PSD preferences.

2.4.3 Almost Stochastic Dominance

Guo, Zhu, Wong, and Zhu (2013) provide further information on both the expected-utility maximization and the hierarchy property. For almost SD (ASD), Leshno and Levy (2002) propose a definition, and Tzeng et al. (2013) modify it to provide another definition. Guo, Zhu, Wong, and Zhu (2013) show that the former has the hierarchy property but not expected-utility maximization, whereas the latter has the expected-utility maximization but not the hierarchy property.

Guo, Post, Wong, and Zhu, (2014) establish necessary conditions for ASD criteria of various orders. These conditions take the form of restrictions on algebraic combinations of moments of the probability distributions in question. The relevant set of conditions depends on the relevant order of ASD but not on the critical value for the admissible violation area. These conditions can help to reduce the information requirement and computational burden in practical applications. A numerical example and an empirical application for historical stock market data illustrate the moment conditions. The first four moment conditions, in particular, seem appealing for many applications.

Guo, Wong, and Zhu (2016) extend ASD theory for risk averters to include ASD for risk-seeking investors. Thereafter, they study the relationship between ASD for risk seekers and ASD for risk

### 2.5 Indifference curves

Meyer (1987) extends MV theory to include comparisons among distributions that differ only by location and scale parameters, and include general utility functions with only convexity or concavity restrictions. Wong (2006) extends both Meyer (1987) and Tobin (1958) by showing that the indifference curve is convex upwards for risk averters, concave downwards for risk lovers, and horizontal for risk neutral investors, to include the general conditions stated by Meyer (1987). In addition, Wong (2006) develops some properties among the first- and second-order SD efficient sets and the mean-variance efficient set.

Wong and Ma (2008) extend the work on the location-scale (LS) family with general n random seed sources in a multivariate setting. In addition, they develop some properties for general non-expected utility functions defined over the LS family, and characterize the shapes of the indifference curves induced by the location-scale expected utility functions and non-expected utility functions. Thereafter, they develop properties for well-defined partial orders and dominance relations defined over the LS family, including the first- and second-order stochastic dominances, the mean-variance rule, and location-scale dominance.

Broll, Egozcue, Wong, and Zitikis (2010) discuss prospect theory and establish general results concerning certain covariances from which they can, in turn, infer properties of indifference curves and hedging decisions within prospect theory.

### 2.6 Diversification

Wong and Li (1999) extend the theory of convex SD (Fishburn, 1974) by including any distribution function, developing the results for both risk seekers as well as risk averters, and including third-order stochastic dominance. Their results can be used to extend a theorem of Bawa et al. (1985) on comparisons between a convex combinations of several continuous distributions
and a single continuous distribution.

Li and Wong (1999) develop some results for the diversification preferences of risk averters and risk seekers. Egozcue and Wong (2010) incorporate both majorization theory and SD theory to develop a general theory and unifying framework for determining the diversification preferences of risk-averse investors, and conditions under which they would unanimously judge a particular asset to be superior. In particular, they develop a theory for comparing the preferences of different convex combinations of assets that characterize a portfolio to yield higher expected utility by second-order SD.

Egozcue et al. (2011) analyse the rankings of completely and partially diversified portfolios and also of specialized assets when investors follow so-called Markowitz preferences. Diversification strategies for Markowitz investors are more complex than in the case of risk-averse and risk-inclined investors, whose investment strategies have been investigated extensively in the literature. In particular, they observe that, for Markowitz investors, preferences toward risk vary depending on their sensitivities toward gains and losses.

For example, it can be shown that, unlike the case of risk-averse and risk-inclined investors, Markowitz investors might prefer investing their entire wealth in just one asset. This finding helps us to better understand some financial anomalies and puzzles, such as the well-known diversification puzzle, which notes that investors may concentrate on investing in only a few assets instead of choosing the seemingly more attractive complete diversification.

Lozza, Wong, Fabozzi, Egozcue (2018) provide a general valuation of the diversification attitude of investors. First, they empirically examine the diversification of mean-variance optimal choices in the US stock market during the 11-year period 2003-2013. Then they analyze the diversification problem from the perspective of risk-averse investors and risk-seeking investors.

Second, the authors prove that investors’ optimal choices will be similar if their utility functions are not too distant, independent of their tolerance (or aversion) to risk. Finally, they discuss
investors’ attitudes towards diversification when the choices available to investors depend on several parameters.

2.7 Risk measures

We have been developing properties for several risk measures to be used in finance, economics, and cognate disciplines, and discuss briefly the properties for some recent risk measures in this section.

2.7.1 VaR and CVaR

Ma and Wong (2010) establish some behavioral foundations for various types of VaR models, including VaR and conditional-VaR, as measures of downside risk. They establish some logical connections among VaRs, conditional-VaR, SD, and utility maximization. Though supported to some extent by unanimous choices by some specific groups of expected or non-expected-utility investors, VaRs as profiles of risk measures at various levels of risk tolerance are not quantifiable as they can only provide partial and incomplete risk assessments for risky prospects.

They also include in the discussion the relevant VaRs and several alternative risk measures for investors. These alternatives use somewhat weaker assumptions about risk-averse behavior by incorporating a mean-preserving-spread. For this latter group of investors, the authors provide arguments for and against the standard deviation versus VaR and conditional-VaR as objective and quantifiable measures of risk in the context of portfolio choice.

2.7.2 Omega ratio

Both SD and the Omega ratio can be used to examine whether markets are efficient, whether there is any arbitrage opportunity in the market, and whether there is any anomaly in the market. Guo, Jiang, and Wong (2017) analyse the relationship between SD and the Omega ratio. They find that second-order SD and/or second-order risk-seeking SD (RSD) alone for any two prospects is not sufficient to imply Omega ratio dominance, insofar as the Omega ratio of one asset is always
greater than that of the other. They extend the theory of risk measures by proving that the preference of second-order SD implies the preference of the corresponding Omega ratios only when the return threshold is less than the mean of the higher return asset.

On the other hand, the preference of the second-order RSD implies the preference of the corresponding Omega ratios only when the return threshold is larger than the mean of the smaller return asset. Nonetheless, first-order SD does imply Omega ratio dominance. Thereafter, they apply their theory to examine the relationship between property size and property investment in the Hong Kong real estate market, and conclude that the Hong Kong real estate market is not efficient as there are expected arbitrage opportunities and anomalies in the Hong Kong real estate market.

2.7.3 High-order risk measures

Niu, Wong, and Xu (2017) first show the sufficient relationship between the (n+1)-order SD and the n-order Kappa ratio. They clarify the restrictions for necessarily beating the target for the higher-order SD consistency of the Kappa ratios. Thereafter, the authors show that, in general, a necessary relationship between SD/RSD and the Kappa ratio cannot be established. They find that when the variables being compared belong to the same location-scale family or the same linear combination of location-scale families, they can obtain the necessary relationship between the (n+1)-order SD with the n-order Kappa ratio after imposing some conditions on the means.

2.8 Two-Moment decision model

Alghalith, Guo, Niu, and Wong (2017) analyze the impacts of joint energy and output prices uncertainties on input demands in a mean–variance framework. They show that an increase in the expected output price will cause the risk-averse firm to increase input demand, while an increase in expected energy prices will surely cause the risk-averse firm to decrease the demand for energy, but increase the demand for the non-risky inputs.
Furthermore, the authors investigate two cases with only uncertain energy price and only uncertain output price. In the case with only uncertain energy price, they determine that the uncertain energy price has no impact on the demands for the non-risky inputs. They also show that the concepts of elasticity and decreasing absolute risk aversion (DARA) play an important role in the comparative statics analysis.

Alghalith, Niu, and Wong (2017) analyze the impacts of joint energy and output prices uncertainties on the inputs demands in a mean-variance framework. They find that the concepts of elasticities and variance vulnerability play important roles in the comparative statics analysis. If the firms’ preferences exhibit variance vulnerability, increasing the variance of energy price will necessarily cause the risk averse firm to decrease demand for the non-risky inputs.

Furthermore, the authors investigate two special cases with only uncertain energy price and only uncertain output price. In the case with only uncertain energy price, they show that the uncertain energy price has no impact on the demands for the non-risky inputs. If the firms’ preferences exhibit variance vulnerability, increasing the variance of energy price will surely cause the risk averse firm to decrease demand for energy.

With multiple additive risks, the mean–variance approach and the expected utility approach of risk preferences are compatible if all attainable distributions belong to the same location–scale family. Under this proviso, Guo, Wagener, and Wong (2018) survey existing results on the parallels of the two approaches with respect to risk attitudes, the changes thereof, and comparative statics for simple, linear choice problems under risk.

In the mean–variance approach, all effects can be couched in terms of the marginal rate of substitution between the mean and variance. They apply the theory developed in the paper to examine the behavior of banking firms, and study risk-taking behavior with background risk in the mean-variance model.

2.9 Dynamic models with background risk
Alghalith, Guo, Wong, and Zhu (2016) use a general utility function to present two dynamic models of background risk. They present a stochastic factor model with an additive background risk. Thereafter, they present a dynamic model of simultaneous (correlated) multiplicative background risk and additive background risk.

### 2.10 Regret-Aversion

Egozcue, Guo, and Wong (2015) examine the optimal output of a competitive firm for price uncertainty. Instead of assuming a risk-averse firm, the authors assume that the firm is regret-averse. They find that optimal output under uncertainty would be lower than under certainty, and prove that optimal output could increase or decrease as the regret factor varies.

Guo, Wong, Xu, and Zhu (2015) investigate regret-averse firms’ production and hedging behavior. They show that the separation theorem operates under regret aversion by proving that regret aversion is independent of the level of optimal production. On the other hand, the authors find that the full-hedging theorem does not always hold under regret aversion, as regret-averse firms take hedged positions differently from those of risk-averse firms in some situations. With more regret aversion, regret-averse firms will hold smaller optimal hedging positions in an unbiased futures market. Furthermore, contrary to the conventional expectations, they show that banning firms from forward trading affects their production level in both directions.

### 2.11 Covariances and copulas

Chebyshev’s integral inequality, also known as covariance inequality, is an important problem in economics, finance, marketing, management, and decision making in a wide range of cognate disciplines. Egozcue, Fuentes García, and Wong (2009) derive some covariance inequalities for monotonic and non-monotonic functions. The results can be useful in many applications in economics, finance, marketing, management, and decision making, and related disciplines where optimal decision making is desired.
Egozcue, Fuentes García, Wong, and Zitikis (2010) sharpen the upper bound of a Grüss-type covariance inequality by incorporating a notion of quadrant dependence between random variables, and also using the idea of constraining the means of the random variables.

Egozcue, Fuentes García, Wong, and Zitikis (2011a) show that Grüss-type probabilistic inequalities for covariances can be considerably sharpened when the underlying random variables are quadrant dependent in expectation (QDE). The established covariance bounds not only sharpen the classical Grüss inequality, but also improve upon Grüss-type bounds under the assumption of quadrant dependency (QD), which is stronger than QDE. The authors illustrate the general results with examples based on specially devised bivariate distributions that are QDE but not QD. Such results play important roles in decision making under uncertainty, and particularly in areas such as economics, finance, marketing, management, insurance and cognate disciplines in which optimal decision making is required.

A number of problems in economics, finance and insurance rely on determining the signs of the covariances of two transformations of a random variable. The classical Chebyshev’s inequality offers a powerful tool for solving the problem, but assumes that the transformations are monotonic, which is not always the case in applications.

For this reason, Egozcue, Fuentes García, Wong, and Zitikis (2011b) establish new results for determining the covariance signs and provide further insights into the area. Unlike many previous contributions, their method of analysis, which is probabilistic in nature, does not rely on the classical Hoffding’s representation of the covariances or on any of the numerous extensions and generalizations.

Egozcue, Fuentes García, Wong, and Zitikis (2012) establish the smallest upper bound for the p’th absolute central moment over the class of all random variables with values in a compact interval. Numerical values of the bound are calculated for the first ten integer values of p, and its asymptotic behaviour derived when p tends to infinity. In addition, the authors establish an analogous bound in the case of all symmetric random variables with values in a compact interval. Such results play
important roles in a number of areas, including actuarial science, economics, finance, marketing, management, operations research, and reliability.

It is well known that quadrant dependent (QD) random variables are also quadrant dependent in expectation (QDE). The recent literature has offered examples that establish rigorously the fact that there are QDE random variables which are not QD. The examples are based on convex combinations of specially chosen positive and negative QD copulas. Egozcue, Fuentes García, Wong, and Zitikis (2013) establish general results that determine when convex combinations of arbitrary QD copulas yield negative or positive QD/QDE copulas. In addition to being an interesting mathematical exercise, the established results are helpful from a practical perspective when modelling insurance and financial portfolios.

3. Statistical and Econometric Models

Another suggestion is to develop statistical and econometric models in the areas related to management information, decision sciences, economics, finance, and cognate disciplines. After developing mathematical models, one might consider developing related statistical and econometric models. We have developed several econometrics papers related to management information, decision sciences, economics, and finance, among others.

3.1 Portfolio optimization

We have developed some novel theoretical results on portfolio optimization. When the dimension of the data is large, the theoretical model of the classical MV portfolio optimization developed by Markowitz (1952) has been found to have problematic issues in estimation as substituting the sample mean and covariance matrix into the MV optimization procedure will result in a serious departure of the optimal return estimate. Moreover, the corresponding portfolio allocation estimates will deviate from their theoretical counterparts when the number of assets is large. We call this return estimate the “plug-in” return, and its corresponding estimate for the asset allocation the “plug-in allocation.”
Bai, Liu, and Wong (2009a) prove that this phenomenon is normal and call it “over-prediction”. In order to circumvent over-prediction, the authors use a new method by incorporating the idea of the bootstrap into the theory of a 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 over-prediction and drastically reduce the error. The authors 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 (2009a) 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 using a bootstrap approach which, as a result, is difficult for practitioners to adopt the estimation technique in practice. 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.

Li, Bai, McAleer, and Wong (2017) improve on the estimation by using the spectral distribution of the sample covariance. They develop the limiting behavior of the quadratic form with the sample spectral corrected covariance matrix, and explain the superior performance to the sample covariance as the dimension increases to infinity with the sample size proportionally. Moreover, the authors derive the limiting behavior of the expected return and risk on the spectrally corrected MV portfolio. They also illustrate the superior properties of the spectral corrected MV portfolio in practice.

In simulations, they compare the spectrally corrected estimates with the traditional and bootstrap-corrected estimates, and show the performance of the spectrally corrected estimates are superior in terms of the portfolio return as well as for the portfolio risk. They also compare the performance of the novel proposed estimation method with different optimal portfolio estimates for real S&P 500 data.
We note that portfolio optimization can be used for big data as well finite samples that might not be classified as big data. In the theory developed by Bai, Liu, and Wong (2009a, b), Leung, Ng, and Wong (2012), and Li, Bai, McAleer, and Wong (2017) have already mentioned their theory holds when the observations tend to infinity. Academics and practitioners can use portfolio optimization in their analysis for big data, and for finite samples. The literature for using portfolio optimization in their theoretical and empirical analyses includes Abid, Mroua, and Wong (2009, 1013), Abid, Leung, Mroua, and Wong (2014), Hoang, Lean, and Wong (2015), Hoang, Wong, and Zhu (2015), among others.

3.2 Testing investors’ behavioral models


We note that the testing Investors’ behavioral models developed by Lam, Liu, and Wong (2008, 2010), and Guo, McAleer, Wong, and Zhu (2017), can be used for big data as well as in finite samples. Fabozzi, Fung, Lam, and Wong (2013) have already developed three tests and use S&P data to test for the magnitude effect of short-term underreaction and long-term overreaction. Academics and practitioners can apply the tests developed in Fabozzi, Fung, Lam, and Wong (2013) for broader data sets, such as for international markets and over time, leading to dynamic panel data models, so that the tests can be used for big data as well as for finite samples.

Wong, Chow, Hon, and Woo (2018) conduct a questionnaire survey to examine whether the theory developed by Lam, Liu, and Wong (2008, 2010), and Guo, McAleer, Wong, and Zhu (2017), holds empirically by studying the conservative and representative heuristics of Hong Kong small investors who adopt momentum and/or contrarian trading strategies. It is worth noting that academics and practitioners could conduct a questionnaire survey for big data as well as finite samples associated with this topic.
3.3 Stochastic Dominance

Ng, Wong, and Xiao (2017) develop tests for stochastic dominance by proposing and translating the inference problem of stochastic dominance into parameter restrictions in quantile regressions. They are variants of the one-sided Kolmogorov-Smirnoff statistic with a limiting distribution of the standard Brownian bridge. The procedure to obtain the critical values of the proposed test statistics are provided. Simulation results show superior size and power compared with alternative procedures. They apply the estimation method to the NASDAQ 100 and S&P 500 indexes to investigate dominance relationship before and after major turning points. The empirical results show no arbitrage opportunities between the bear and bull markets.

Bai, Li, McAleer, and Wong (2015) derive the limiting process of stochastic dominance statistics for risk averters as well as for risk seekers, both for when the underlying processes are dependent or independent. They take account of the dependency of the partitions and propose a bootstrap method to determine the critical points. In addition, they illustrate the applicability of the stochastic dominance statistics for both risk averters and risk seekers to analyze the dominance relationship between the Chinese and US stock markets for the full sample period, as well as for the sub-periods before and after crises, including the internet bubble, the recent sub-prime crisis, and global financial crisis.

The empirical findings could be used to draw inferences on the preferences of risk averters and risk seekers in investing in the Chinese and US stock markets. The results also enable an examination as to whether there are arbitrage opportunities in these markets, whether these markets are efficient, and if investors are rational.

Bai, Li, Liu and Wong (2011) develop new statistics for both PSD and MSD of the first three orders. These statistics provide tools to examine the preferences of investors with S-shaped utility functions in prospect theory and investors with RS-shaped investors. They also derive the limiting distributions of the test statistics to be stochastic processes, propose a bootstrap method to decide the critical points of the tests, and prove the consistency of the bootstrap tests. The authors also
illustrate the applicability of their proposed statistics by examining the preferences of investors with the corresponding S-shaped and RS-shaped utility functions vis-a-vis returns on iShares, and vis-a-vis returns of traditional stocks and Internet stocks, before and after the internet bubble.


3.4 Risk measures

Leung and Wong (2008) apply the technique of the repeated measures design to develop the Multiple Sharpe ratio test statistic to test the hypothesis of the equality of the multiple Sharpe ratios. They also establish the asymptotic distribution of the statistic and its properties. In order to demonstrate the superiority of the proposed statistic relative to the traditional pairwise Sharpe ratio test, they illustrate their approach by testing the equality of the Sharpe ratios for eighteen iShares.

The pairwise Sharpe ratio test shows that the performance of all 18 iShares are indistinguishable, as they reject the equality of the Sharpe ratios in each year as well as for the entire sample. These empirical results imply that the 18 iShares perform differently in each year, as well as for the entire sample, with some tests outperforming others in the market.

Recent results in optimal stopping theory have shown that a ‘bang-bang’ (buy or sell immediately) style of trading strategy is, in some sense optimal, provided that the asset price dynamics follow certain familiar stochastic processes. Wong, Wright, Yam, and Yung (2012) construct a reward-to-variability ratio (specifically, the mixed Sharpe ratio) that is sufficient for purposes of implementing the strategy.
The use of the novel ratio for optimal portfolio selection is discussed, and evidence for it varying over time is established. The performances of the ‘bang-bang’ and ‘buy-and-hold’ trading strategies are compared, and the former is found to be significantly more profitable.

Bai, Wang, and Wong (2011) develop the mean-variance-ratio statistic to test the equality of the mean-variance ratios and prove that our proposed statistic is uniformly most powerful unbiased. In addition, they illustrate the applicability of our proposed test to compare the performances of stock indices.

Thereafter, Bai, Hui, Wong, and Zitikis (2012) propose and develop mean-variance-ratio (MVR) statistics for comparing the performance of prospects after the effect of the background risk has been mitigated. They investigate the performance of the statistics in large and small samples and show that, in the non-asymptotic framework, the MVR statistic produces a uniformly most powerful unbiased (UMPU) test.

The authors discuss the applicability of the MVR test in the case of large samples, and illustrate its superiority in the case of small samples by analyzing the Korea and Singapore stock returns after the impact of the US stock returns (which is viewed as the background risk) has been deducted. They find, in particular, that when samples are small, the MVR statistic can detect differences in asset performance while the Sharpe ratio, which is the mean-standard-deviation-ratio statistic, may not be able to do so.

3.5 Economic and financial indicators

We have developed financial indicators and have applied some economic indicators to examine several important economic issues. For example, Wong, Chew, and Sikorski (2001) develop a new financial indicator to test the performance of stock market forecasts by using E/P ratios and bond yields. They also develop two test statistics to use the indicator and illustrate empirically the tests in several stock markets.

The empirical results show that the forecasts generated from the indicator would enable investors to escape most of the crashes and catch most of the bull runs. The trading signals provided by the indicator can also generate profits that are significantly superior to the buy-and-hold strategy.

Exploring the characteristics associated with the formation of bubbles that occurred in the Hong Kong stock market in 1997 and 2007, and the 2000 dot-com bubble of Nasdaq, McAleer, Suen, and Wong (2016) establish trading rules that not only produce returns that are significantly greater than the buy-and-hold strategies, but also produce greater wealth compared with TA strategies without trading rules.

Chong, Cao, and Wong (2017) develop a new market sentiment index for the Hong Kong stock market, one of the largest stock markets in the world by using the turnover ratio, short-selling volume, money flow, HIBOR, and returns of the US and Japanese markets, and the Shanghai and Shenzhen Composite indices.

Thereafter, they incorporate the threshold regression model with the sentiment index as a threshold variable to capture the state of the Hong Kong stock market. The authors find that the practical trading rule which sells (buys) the HSI or S&P/HKEx LargeCapIndex when the sentiment index is above (below) the upper threshold value can beat the buy-and-hold strategy.

Sethi, Wong, and Acharya (2018) examine the sectoral impact of disinflationary monetary policy by calculating the sacrifice ratios for several OECD and non-OECD countries. Sacrifice ratios
calculated through the episode method reveal that disinflationary monetary policy has a differential impact across three sectors in both OECD and non-OECD countries. Of the three sectors, the industry and service sectors show significant output loss due to a tight monetary policy in OECD and non-OECD countries.

The agricultural sector shows a differential impact of disinflation policy, namely a negative sacrifice ratio in OECD countries, thereby indicating that output growth is insignificantly affected by a tight monetary policy. Non-OECD countries yield positive sacrifice ratios, suggesting that the output loss is significant. Furthermore, it is observed that sacrifice ratios calculated from aggregate data are different from ratios that are calculated using sectoral data.

Financial and economic indicators can be used for big data, and for large data sets that might not be classified as big data. For example, Wong, Chew, and Sikorski (2001) use their indicator to test in markets for the USA, UK, Japan, Germany, and Singapore. This is not especially big data. Academics and practitioners could use it to test for stock markets for a large number of international markets using dynamic panel data models. The authors can use it to test not only for stock markets, but also for any financial products, and also to test for big data sets.

Similarly, Sethi, Wong, and Acharya (2018) apply the sacrifice ratios to examine the sectoral impact of disinflationary monetary policy for several OECD and non-OECD countries. This is not especially associated with big data. However, academics and practitioners can apply the sacrifice ratios to examine the sectoral impact of disinflationary monetary policy for a large number of countries worldwide, which would be classified as a large data set.

3.6 Contagion

Wan and Wong (2001) provide a simple example of a refinancing game with incomplete information, where the lack of transparency is both necessary and sufficient for the propagation of local financial distress across disjoint financial networks. The authors note that contagion is an important topic in both economics and finance.
There are some tests for contagion, for example, the test developed by Fry, Martin and Tang (2010), and Fry-McKibbin and Hsiao (2015). The tests can be used for big data, and also for large data sets that might not be characterized as big data.

3.7 Technical analysis

The new financial indicator introduced by Wong, Chew, and Sikorski (2001) to test the performance of stock market forecasts can be classified as technical analysis. Substantial research have been undertaken in technical analysis. For example, Wong, Manzur, and Chew (2003) use technical analysis in signalling the timing of stock market entry and exit.

The authors introduce test statistics to test the performance of the most established of the trend followers, namely the Moving Average, and the most frequently used counter-trend indicator, namely the Relative Strength Index. Using Singapore data, the empirical results indicate that the indicators can be used to generate significantly positive return. It is found that member firms of the Singapore Stock Exchange tend to enjoy substantial profits by applying relatively simple technical indicators.

Wong, Du, and Chong (2005) examine the profitability of applying technical analysis that signals the entry and exit from the stock market in three Chinese stock markets, namely the Shanghai, Hong Kong and Taiwan Stock Exchanges. Applying the trading signals generated by the MA family to the Greater China markets, generate significantly positive returns that outperform the buy-and-hold strategy. The cumulative wealth obtained also surpasses that of the buy-and-hold strategy, regardless of transaction costs.

In addition, the authors analyse the performance of the MA family before and after the 1997 Asian Financial Crisis, and find that the MA family works well in both sub-periods, as well as in different market conditions of bull runs, bear markets and mixed markets. The empirical observation that technical analysis can forecast the directions in these markets implies that the three China stock markets are not efficient. Lam, Chong, and Wong (2007) examine whether a day’s surge or plummet in stock price serve as a market entry or exit signal. Returns of five trading rules based
on 1-day and intraday momentum are estimated for several major world stock indices. It is found that the trading rules perform well in the Asian indices, but not in those of Europe and the USA.

Kung and Wong (2009a) investigate whether these measures have led to less profitability for those investors who employ technical rules for trading stocks. Their results show that the three trading rules consistently generate higher annual returns for 1988-1996 than those for 1999-2007. Furthermore, they generally perform better than the buy-and-hold (BH) strategy for 1988-1996, but perform no better than the BH strategy for 1999-2007. These findings suggest that the efficiency of the Singapore stock market has been considerably enhanced by the measures implemented after the financial crisis.

Kung and Wong (2009b) use two popular technical trading rules to assess whether the gradual liberalization of Taiwan’s securities markets has improved the efficiency in its stock market. The results show that the two rules have considerable predictive power for 1983-1990, become less predictive for 1991-1997, and cannot predict the market for the period 1998-2005. These empirical results indicate that the efficiency of the Taiwan stock market has been greatly enhanced by the liberalization measures implemented in the past 20 years.

The above studies examine technical analysis for reasonably big data sets. In addition, academics and practitioners can apply technical analysis to examine the performance of a larger number of stock markets, as well as other financial market for larger data sets.

3.8 Cost of capital

Wong and Chan (2004) extend the theory by developing estimators of the reliability, and prove that the estimators are consistent. Estimation of the cost of equity capital and its reliability can be used for both big data, and large data sets that might not be classified as big data.

### 3.9 Robust estimation

Bian and Dickey (1996) develop a robust Bayesian estimator for the vector of regression coefficients using a Cauchy-type g-prior. This estimator is an adaptive weighted average of the least squares estimator and the prior location, and is robust with respect to flat-tailed sample distributions.

Wong and Bian (1997) develop an alternative approach to estimate the regression coefficients. Wong and Bian (2000) introduce the robust Bayesian estimator developed by Bian and Dickey (1996) to the estimation of the Capital Asset Pricing Model (CAPM), in which the distribution of the error component is widely known to be flat-tailed.

In order to support their proposal, the authors apply both the robust Bayesian estimator and the least squares estimator in simulations of CAPM, and also in the analysis of CAPM for US annual and monthly stock returns. The simulation results show that the Bayesian estimator is robust and superior to the least squares estimator when the CAPM is contaminated by large normal and non-normal disturbances, especially with Cauchy disturbances.

In their empirical study, the authors find that the robust Bayesian estimate is uniformly more efficient than the least squares estimate in terms of the relative efficiency of one-step ahead forecast mean square errors, especially in small samples. They introduce the robust Bayesian estimator developed by Bian and Dickey (1996) as this robust Bayesian estimator is adaptive and robust with respect to flat-tailed sample distribution. However, few papers have used this estimator in practice.

This estimator is adaptive and robust in the sense that if the sample does not contain outliers, the estimator will rely more on the sample information. On the other hand, if there are many outliers...
in the sample, the robust Bayesian estimator will use more information arising from the prior. To the best of our knowledge, only the estimator in Bian and Dickey (1996) has this feature, and so this estimator is recommended. It should be noted that the robust Bayesian estimator can be used for big data, and for large data sets that might not be interpreted as such.

3.10 Unit roots, cointegration, causality tests, and nonlinearity

We have applied several tests related to unit roots, cointegration, and causality, including for higher moments, specifically a simple test for causality in volatility (see Chang and McAleer, 2017), and discuss a few of the innovations below.

Tiku and Wong (1998) develop a unit root test to accommodate data that follow an AR(1) process. We use the three moment chi-square and four moment F approximations to test for unit roots in an AR(1) model when the innovations have one of a wide family of symmetric Student’s t distributions. In cointegration analysis, vector error-correction models (VECMs) have become an important means of analysing long run cointegrating equilibrium relationships.

The usual full-order VECMs assume all nonzero entries in their coefficient matrices. However applications of VECMs to economic and financial time series data have revealed that zero entries are indeed possible. If indirect causality or Granger non-causality exists among the variables, the use of a full-order VECM will incorrectly conclude only the existence of Granger causality among these variables.

In addition, the statistical and numerical accuracy of the cointegrating vectors estimated in a mis-specified full-order VECM will be problematic. It has been argued that the zero-non-zero (ZNZ) patterned VECM is a more straightforward and effective means of testing for both indirect causality and Granger non-causality. Wong, Penn, Terrell, and Lim (2004) present simulations and an application that demonstrate the usefulness of the ZNZ patterned VECM.

Lam, Wong, and Wong (2006) develop some properties on the autocorrelation of the k-period returns for the general mean reversion (GMR) process, in which the stationary component is not
restricted to the AR(1) process but takes the form of a general ARMA process. The authors derive some properties of the GMR process and three new nonparametric tests that compare the relative variability of returns over different horizons to validate the GMR process as an alternative to a random walk. The authors examine the asymptotic properties of the novel tests, which can be used to identify random walk models from the GMR processes.

The traditional linear Granger causality test has been widely used to examine linear causality among several time series in bivariate settings, as well as in multivariate settings. Hiemstra and Jones (1994) develop a nonlinear Granger causality test in a bivariate setting to investigate the nonlinear causality between stock prices and trading volume. Bai, Wong, and Zhang (2010) extend the work by developing a nonlinear causality test in multivariate settings.

Bai, Li, Wong, and Zhang (2011) discuss linear causality tests in multivariate settings, and thereafter develop a nonlinear causality test in multivariate settings. A Monte Carlo simulation is conducted to demonstrate the superiority of the proposed multivariate test over its bivariate counterpart. In addition, the authors illustrate the applicability of the proposed test to analyze the relationships among different Chinese stock market indices.

Hui, Wong, Bai, and Zhu (2017) propose a simple and efficient method to examine whether a time series process possesses any nonlinear features by testing dependence remaining in the residuals after fitting the data with a linear model. The advantage of the proposed nonlinearity test is that it is not required to know the exact nonlinear features and the detailed nonlinear forms of the time series process. It can also be used to test whether the hypothesized model, including linear and nonlinear components of the variable being examined, is appropriate as long as the residuals of the model being used can be estimated.

The simulation study shows that the proposed test is stable and powerful. The authors apply the proposed statistic to test whether there is any nonlinear feature in sunspot data, and whether the S&P 500 index follows a random walk. The conclusion drawn from the proposed test is consistent with results that are available from alternative tests.
An early development in testing for causality (technically, Granger non-causality) in the conditional variance (or volatility) associated with financial returns was the portmanteau statistic for non-causality in the variance of Cheng and Ng (1996). A subsequent development was the Lagrange Multiplier (LM) test of non-causality in the conditional variance by Hafner and Herwartz (2006), who provided simulation results to show that their LM test was more powerful than the portmanteau statistic for sample sizes of 1000 and 4000 observations.

Although the LM test for causality proposed by Hafner and Herwartz (2006) is an interesting and useful development, it is nonetheless arbitrary. In particular, the specification on which the LM test is based does not rely on an underlying stochastic process, so the alternative hypothesis is also arbitrary, which can affect the power of the test.

Chang and McAleer (2017) derive a simple test for causality in volatility that provides regularity conditions arising from the underlying stochastic process, namely a random coefficient AR process, and a test for which the (quasi-) maximum likelihood estimates have valid asymptotic properties under the null hypothesis of non-causality. The simple test is intuitively appealing as it is based on an underlying stochastic process, is sympathetic to Granger’s (1969, 1988) notion of time series predictability, is easy to implement, and has a regularity condition that is not available in the LM test.

We note that using cointegration, causality and nonlinearity tools is very useful in analyzing many important issues and explains many financial and physiological phenomena well. For example, using cointegration, causality and nonlinearity tools, Batai, Chu, Lv, and Wong (2017) examine the factors that maintain a long-run equilibrium, short-run impact, and causality with the exchange rate of Mongolia over China to shed light on exchange rate determination.

The authors find that, in the long run, the gross domestic product (GDP) of China and the index of world price have significantly positive effects, while Mongolia’s GDP and the Shanghai stock index have significantly negative effects on the Mongolian exchange rate.
The research also reveals the existence of a short run dynamic interaction, and highly significant linear and nonlinear multivariate causality from all the explanatory variables to the Mongolian exchange rate. The authors observe that there is strong linear causality from each of the GDPs of Mongolia and China and the index of world price to Mongolian exchange rate, but not from the index of world price. Moreover, there is strongly significant nonlinear causality from the Shanghai stock index to the Mongolian exchange rate, and weakly significant nonlinear causalities from both the GDP of China and the index of world price on the Mongolian exchange rate, but not from Mongolia’s GDP. The empirical findings are useful for investors, manufacturers, and traders for their investment decision making, and for policy makers for their decisions regarding both monetary and fiscal policies that could affect the Mongolian exchange rate.

Academics and practitioners can apply unit root, cointegration, causality, and nonlinearity tests in many different areas for big data, and large data sets, as in empirical finance that uses nano-tick data, and dynamic panel data models with both large cross section and time series components. The literature in applying unit root, cointegration, causality and nonlinearity tests includes Wong, Penn, Terrell, and Lim (2004), Wong, Khan, and Du (2006), Qiao, Liew, and Wong (2007), Foo, Wong, and Chong (2008), Qiao, Smyth, and Wong (2008), Qiao, Chiang, and Wong (2008), Chiang, Qiao, and Wong (2009), Qiao, McAleer, and Wong (2009), Qiao, Li, and Wong (2011), Vieito, Wong, and Zhu (2015), and chang and McAleer (2017), among many others.

### 3.11 Confidence intervals

Homm and Pigorsch (2012) use the Aumann and Serrano index to develop a new economic performance measure (EPM), which is well known to have advantages over alternative measures. Niu, Guo, McAleer, and Wong (2018) extend the theory by constructing a one-sample confidence interval of EPM, and construct confidence intervals for the difference of EPMs for two independent samples. The authors also derive the asymptotic distribution for EPM and for the difference of two EPMs when the samples are independent. They conduct simulations to show the proposed theory performs well for one and two independent samples.
The simulations show that the proposed approach is robust in the dependent case. The theory developed is used to construct both one-sample and two-sample confidence intervals of EPMs for the Singapore and USA stock indices. It is worth noting that estimation of the confidence intervals can be used for big data, and large finite samples that are not regarded as big data.

The theory of confidence intervals for EPM developed in Niu, Guo, McAleer, and Wong (2018) can be used to develop the theory of confidence intervals for any risk measure or economic indicator which, in turn, could be used to construct confidence intervals for big data, large finite data samples that are not otherwise classified as big data.

### 3.12 Other econometrics models and tests

The literature provides numerous alternative econometric/statistic models/tests, several of which have been used in a number of cognate disciplines, including economics, finance, management, marketing and statistics. Some of these are discussed below.

Wong and Miller (1990) develop a theory and methodology for repeated time series (RTS) measurements on autoregressive integrated moving average noise (ARIMAN) process. The theory enables a relaxation of the normality assumption in the ARIMAN model, and to identify appropriate models for each component series of the relevant stochastic process. The authors discuss the properties, estimation, and forecasting of RTS ARIMAN models and illustrate with examples.

Wong, Miller, and Shrestha (2001) extend the theory and methodology of Wong and Miller (1990) by allowing the error variance, as well as the number of repetitions, to change over time. They show that the model is identified, and derive the maximum likelihood estimator using the Kalman filter technique.

Tiku, Wong, Vaughan, and Bian (2000) consider AR(q) models in time series with non-normal innovations represented by a member of a wide family of symmetric distributions (Student's t). Since the ML (maximum likelihood) estimators are intractable, we derive the MML (modified
maximum likelihood) estimators of the parameters and show that they are remarkably efficient. We use these estimators for hypothesis testing, and show that the resulting tests are robust and powerful.

Tiku, Wong, and Bian (1999a) extend the methods by considering AR(q) models in time series with asymmetric innovations represented by two families of distributions, namely (i) gamma with support IR: (0, ∞), and (ii) generalized logistic with support IR: (-∞, ∞). As the maximum likelihood estimators (MLE) are intractable, the authors derive modified maximum likelihood (MML) estimators of the parameters and show that they are very easy to compute and are also efficient. The authors investigate the efficiency properties of the classical LS (least squares) estimators. Their efficiencies relative to the proposed MML estimators are very low.

Tiku, Wong, and Bian (1999b) estimate coefficients in a simple regression model in the presence of autocorrelated errors. The underlying distribution is assumed to be symmetric, namely one of Student’s t family for illustration. Closed form estimators are obtained and shown to be remarkably efficient and robust.

Wong and Bian (2005) extend the results to the case where the underlying distribution is a generalized logistic distribution. The generalized logistic distribution family represents very wide skewed distributions ranging from highly right skewed to highly left skewed. Analogously, the authors develop MML estimators as the ML estimators are intractable for the generalized logistic data. The authors examine the asymptotic properties of the proposed estimators and conduct simulations to establish small sample properties of small size and high power.

Bian and Dickey (1996) develop a robust Bayesian estimator for the vector of regression coefficients using a Cauchy-type g-prior. This estimator is an adaptive weighted average of the least squares estimator (LSE) and the prior location, and is robust to fat-tailed sample distributions. Wong and Bian (2000) introduce the robust Bayesian estimator to the estimation of the Capital Asset Pricing Model (CAPM) in which the distribution of the error component is well-known to be fat-tailed.
In order to support their proposal, the authors apply both the robust Bayesian estimator and the least squares estimator (LSE) in simulations of CAPM, and also in analysing CAPM for US annual and monthly stock returns. The simulation results show that the Bayesian estimator is robust and superior to LSE when CAPM is contaminated by large normal and/or non-normal disturbances, especially by Cauchy disturbances.

In the empirical study, the authors find that the robust Bayesian estimate is uniformly more efficient than the LSE in terms of the relative efficiency of one-step ahead forecast mean square errors, especially in small samples. Bian, McAleer, and Wong (2013) develop a modified maximum likelihood (MML) estimator for the multiple linear regression model with underlying Student’s t distribution.

The authors obtain a closed form solution of the estimators, derive the asymptotic properties, and demonstrate that the MML estimator is more appropriate for estimating the parameters of the Capital Asset Pricing Model (CAPM) by comparing its performance with LSE for monthly returns of US portfolios. The empirical results reveal that the MML estimators are more efficient than LSE in terms of the relative efficiency of one-step-ahead forecast mean square errors in small samples.

Bian, McAleer, and Wong (2011) develop a new test, namely the trinomial test, for pairwise ordinal data samples to improve the power of the sign test by modifying its treatment of zero differences between observations, effectively increasing the use of sample information. Simulations demonstrate the power superiority of the proposed trinomial test statistic over the sign test in small samples in the presence of tied observations.

The authors also show that the proposed trinomial test has substantially higher power than the sign test in large samples and also in the presence of tied observations, as the sign test ignores information from observations resulting in ties.

It is worth noting that all of the above estimation and testing procedures can be used for big data, as well as for finite samples that might not be classified as big data.
4. Simulations

After developing statistical theories for Big Data, Computational Science, Economics, Finance, Marketing, Management, Psychology, and cognate disciplines, academics and practitioners could consider conducting simulations to examine whether the estimators and hypothesis tests that have been developed have good size and high power. We conduct simulations to examine the performance in finite samples of small dimension. For example, Tiku and Wong (1998) conduct simulations to examine whether their unit root test have good size and high power.

Tiku, Wong, and Bian (1999) consider AR(q) models in time series with asymmetric innovations represented by the gamma and generalized logistic distributions. They derive MML (modified maximum likelihood) estimators of the parameters and show that they are remarkably efficient. The authors conduct simulations to examine whether their estimators have small size and high power.

Tiku, Wong, and Bian (1999a) develop the theory to estimate the coefficients in a simple regression model with autocorrelated errors under the Student’s t distribution. The authors conduct simulations to examine whether the estimators have small size and high power.

Tiku, Wong, Vaughan, and Bian (2000) consider AR(q) models in time series with non-normal innovations represented by a member of a wide family of symmetric distributions (Student’s t). The authors derive the MML estimators of the parameters and show that they are efficient. The authors use the estimators for hypothesis testing, and conduct simulations to show that the resulting tests are robust and powerful.

After checking the theory on estimation and hypothesis testing lead to test that have small size and high power, academics and practitioners could then apply the theory and estimation and testing methods to analyze some interesting and important issues for big data and in large finite samples that are deemed not to be big data.
5. Empirical Studies

After developing theoretical models, as well as econometric and statistical models to estimate the parameters, academics and practitioners could then apply the theories to analyse some interesting issues in the seven disciplines and cognate areas.

5.1 Applications in economics and finance

Readers may refer to Chang, McAleer, and Wong (2016a) for applying the theoretical models, and econometric and statistical models, to behavioural, finance, health and medical economics; Chang, McAleer, and Wong (2016b) for applying the theory, and econometric and statistical models, to informatics, data mining, econometrics and financial economics; Chang, McAleer, and Wong (2016c) for applying the theory, and econometric and statistical models, to management science, economics and finance; and Chang, McAleer, and Wong (2017) for applying the theory, and econometric and statistical models, to management information, decision sciences, and financial economics.

Academics and practitioners could apply their theories to other financial economic problems. For example, Raza, Sharif, Wong, and Karim (2016) investigate the empirical influence of tourism development (TD) on environmental degradation in a high-tourist-arrival economy (that is, USA) using the wavelet transformation framework. This new methodology enables the decomposition of time series at different time frequencies.

In the paper, the authors use the maximal overlap discrete wavelet transform (MODWT), wavelet covariance, wavelet correlation, continuous wavelet power spectrum, wavelet coherence spectrum and wavelet-based Granger causality analysis, to analyse the relationship between TD and CO$_2$ emissions in the USA by using monthly data for the period 1996(1) to 2015(3). The results indicate that TD has a significant positive influence over CE in the short, medium and long run. The authors find unidirectional influences of TD on CE in the short, medium and long run in the USA.
In addition, SD can be used to examine income inequality. For example, Chow, Lui, Valenzuela, and Wong (2015) apply SD tests to analyze the relative welfare levels of income distributions for the poor and rich in different groups of individuals. Bai, Valenzuela, Wong, and Zhu (2016) extend the theory by applying MSD and PSD to develop SD tests for the poor (test for poorness), the rich (test for richness), and middle class (test for middle class) to achieve a more robust analysis of relative welfare levels in the analysis of income distributions.

5.2 Applications in psychology

In this sub-section, we discuss applications in marketing, management, and psychology. We first discuss applications in psychology. Chang, Wong, Teo, and Fam (1997) hypothesize that, when measured as an intrinsic need rather than as an inference from actual achievement and/or achievement-related images, the motivation to achieve may be a transcultural construct.

The authors use the Work and Family Orientation Scale (WOFO) to tap this intrinsic need. They conducted a series of surveys in Singapore with four samples of students and employed workers, comprising a total sample size of 1147. Factor analyses of the results with PROMAX rotation revealed three oblique factors that are similar in content to those of reported US data. Comparisons of the factor structure from subsamples in Singapore revealed good reliability.

Confirmatory factor analysis showed a high degree of correspondence. LISREL modeling indicated that this three-dimensional model: work ethics, mastery and competition, is an appropriate model for the Singaporean data. Predictive validity of WOFO was established by differentiating two groups of students with varying achievement levels. Convergence validity was supported by correlations with the Social and Individual-Oriented Achievement Scale developed in Taiwan, especially for the Chinese.

Chang, Wong, and Teo (2000) conducted a survey on two hundred and seventeen male (45%) and female (55%) Singaporean Chinese secondary school students (mean age = 16), with the Work and Family Orientation Scale (WOFO) designed to measure the intrinsic motivation to achieve and the Individual-oriented and Social-oriented Achievement Motivation Scale (10AM-SOAM),
a culture-specific measure of Chinese achievement tendencies. They use the causal model to explore the relationships between the different dimensions of WOFO and 10AMSOAM.

Chang, Wong, and Koh (2003) conducted a series of three studies to test the internal structure of the Chinese value hierarchy (CVH) in Singapore. Study 1 identified the empirically best–fit model with six factors, namely Prudence, Industry, Civic–Harmony, Moral Development, Social Power and Moderation. The relative magnitudes and interfactor correlations suggested that these factors could be further grouped into two superordinate clusters, namely (i) The Modern factor, with significantly higher magnitude, consisting of Prudence, Industry, Civic–Harmony and Moral Development; and (ii) the Tradition factor, with lower magnitude, consisting of Social Power and Moderation.

Study 2 surveyed university students with differential preference for language usage, namely English or Chinese. Both language groups were equally high on the Modern factor, but the Chinese-language-preferred group showed a significantly higher endorsement for the Tradition factor, Chinese Worldview (CWV) and Chinese Health Beliefs (CHB).

Further convergence validation for the Modern and Tradition factors was obtained by investigating their correlations with traditional Chinese beliefs and practices for the two language groups separately. Study 3 tested generational differences in CVH. University participants (Self) were compared with their parents (Parents) and friends (Friends).

There were no differences between Self and Friends on both the Modern and Tradition factors, CWV and individual differences of modernity. Parents and Self did not differ on individual differences of modernity. Parents, however, were higher on the Modern factor, the Tradition factor and CWV. The results were discussed to support the concept of ‘multiple modernity’ in Asian societies, and the ‘revised convergence hypothesis’.

**5.3 Applications in marketing and management**
We now discuss applications in marketing and management. Liao and Wong (2008) explore empirically the major considerations associated with Internet-enabled e-banking systems and systematically measure the determinants of customer interactions with e-banking services. The results suggest that the perceived usefulness, ease of use, security, convenience and responsiveness to service requests significantly explain variations in customer interactions. Exploratory factor analysis and reliability tests indicate that these constructs are relevant and reliable.

Confirmatory factor analysis confirms that they possess significant convergent and discriminatory validities. Both perceived usefulness and perceived ease of use were found to have significant impacts on customer interactions with Internet e-banking services. Perceived security, responsiveness and convenience also represent primary avenues that influence customer interactions. In particular, stringent security control is critical to Internet e-banking operations. Prompt responses to service requests can also encourage customers to use Internet e-banking services.

The empirical findings have managerial implications for enhancing extant Internet e-banking operations and developing viable Internet e-banking services. Liao, Shi, and Wong (2012) explore empirically consumer perceptions of the smartcard as e-cash for purchasing goods and services at retail outlets in Hong Kong. The authors design a multi-attribute model to test the hypotheses using the survey data collected from individual consumers.

The empirical results show that perceived ease of use, convenience, automatic add-value service, compact design, security, reliability, and merchant support have significant effects on perceived usefulness of the smartcard for micro e-payment. The findings contribute to the literature of consumer behavior with regard to the applications of information technology in retailing, and have implications for implementing emerging technology to enhance retail services in different contexts.

Liao, Shi, and Wong (2014) examine the sustainability of smartcard payments in retailing and consumer services. The analytical results of the survey data suggest that usefulness, ease of use, convenience, automatic add-value service, security, reliability, and participation of popular service providers have considerable effects on a continuous use of smartcard payment.
The authors identify empirically and justify the key determinants of sustainable smartcard payment from the perspective of consumers. The findings provide managerial insights for the implementation of cutting-edge technology to enhance sales and service operations and make important contributions to research and practice in technology-based service innovation and service operations management.

Moslehpour, Wong, Aulia, and Pham (2017) examine the key factors that influence Taiwanese consumers to repurchase Korean beauty products. They use a quantitative approach to test the proposed hypotheses using structural equation modelling. A causal research design is used in the research to identify a cause-and-effect relationship among the constructs. Primary data collection is used for the empirical analysis.

This contribution provides a better understanding of the key factors that influence Taiwanese consumers’ repurchase intentions (RI) of Korean cosmetics products. They find that perceived price (PP) and country of origin (COO) significantly influence word-of-mouth (WOM), while PP, COO and WOM significantly influence RI. WOM is the most influential variable toward RI, followed by COO and PP.

Very few studies have examined a general construct of RI related to beauty products. The empirical findings imply several practical directions for marketers of beauty product industries, specifically for Taiwanese consumers. The analysis helps to understand the factors that form a basic consideration for Taiwanese consumers in repurchasing Korean beauty products. Second, it underscores the role of WOM between the independent variables (PP and COO) and RI as the dependent variable.

The mediating role of customer satisfaction has been widely discussed in the existing literature. However, to the best of our knowledge, there is still a lack of studies that focus on the low-cost airline industry, especially in Vietnam. For this reason, Moslehpour, Wong, Lin, and Huyen (2018) investigate the factors that influence purchase intentions and the mediating role of customer satisfaction in VietJet Air in Vietnam. A quantitative research method is applied, with the data
being collected through an online questionnaire from three main regions in Vietnam, namely the North, Center, and South.

The empirical results indicate that customer satisfaction mediates the relationship between the independent variables (customer expectation/perceived value) and the dependent variable (purchase intention) in the case of VietJet Air in Vietnam. In general, the analysis not only enriches the existing literature, but might also be an invaluable reference to VietJet Air and similar low cost Vietnamese carrier managers to consider their strategic marketing plans.

Moslehpour, Pham, Wong, and Bilgiçli (2018) propose a new model that combines personality traits (PT) and Technology Acceptance Model (TAM) to examine the influences of personality characteristics (conscientiousness, openness), and perception of technology (perceived usefulness, perceives ease of use) on e-purchase intentions by using a questionnaire survey to collect a similar sample of Taiwanese online consumers.

The authors find that conscientiousness (CON) (personality attribute) significantly influences perceived usefulness (PU), perceived ease of use (PEOU) (technology perception attributes), and openness to experience (OPE). PU, PEOU, and OPE have significant impacts on online purchase intention (INT). PEOU has the strongest positive impact on (INT). In addition, PU, PEOU, and OPE combined mediate the relationship between CON and INT. Further analysis of mediation shows that PU and PEOU (separately) are each significant mediators. However, OPE alone is not a significant mediator.

**6. Concluding Remarks**

In this paper, we discussed different types of utility functions, stochastic dominance, mean-risk models, portfolio optimization, and others as these topics are important in Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology in terms of theory and econometric and statistical analysis. Authors could extend their work to link the seven cognate disciplines.
Although we have discussed the contributions in SD, MR, and PO related to Big Data, Computational Science, Economics, Finance, Marketing, Management, and Psychology, there are theoretical contributions in other areas that could also be useful in these cognate disciplines. Readers may refer to Chang, McAleer, and Wong (2016a, b, c) and Chang, McAleer, and Wong (2017) for contributions in other cognate areas that might be useful in theory and practice.
References


Moslehpour, M., Pham, V.K., Wong, W.K., Bilgiçli, I., 2018, Online purchase intention of Taiwanese consumers: Sustainable effects of personality traits and technology perception attributes. Forthcoming in Sustainability.


Qiao, Z., Wong, W.K., 2015, Which is a better investment choice in the Hong Kong residential property market: A big or small property? Applied Economics, 47(16), 1670-1685.


