Validating the characteristic based asset pricing approach using a statistical clustering asset pricing model

Professor Steven Brown, New York University;
Professor Robert Faff, Accounting and Finance, Monash University;
Dr. Paul Lajbcygier, Faculty of Information Technology, Monash University;
Associate Professor Madhu Veeraraghavan, Accounting and Finance, Monash University

AIMS AND BACKGROUND

Aim

The part of finance theory dedicated to how risk is related to returns is known as asset pricing theory. Asset pricing theory should provide an accurate model which links return and risk, a link which is essential to finance. Currently, asset pricing theory is in the midst of a 'paradigm shift'. The prior dominant paradigm – the capital asset pricing model (CAPM) (Sharpe 1964; Litner 1965; Black 1972), linked return and risk via a simple one factor linear model. The CAPM, no longer appears to work. In its place, a new model – the factor asset pricing model (FAPM)(Fama and French 1993; Fama and French 1995), has become popular. The FAPM is a multi-factor linear model, stepped in a respected tradition of finance theory, which links risk to return, known as arbitrage pricing theory (APT) (Ross 1976).

A particular implementation of the FAPM which does not rely on APT– the characteristic based approach (CBA) (Daniel and Titman 1997), has appeared recently. Currently, the CBA appears as though it may become the new dominant paradigm in asset pricing.

We believe that this success may be unwarranted. Firstly, the CBA has not been carefully validated nor compared with its competitors. Secondly, the CBA is constructed in an ad-hoc fashion and consequently suffers many econometric problems that make it biased and consequently less accurate, out-of-sample, than its competitors.

Therefore, our aim is to validate the CBA. To achieve this aim we shall introduce an innovative asset pricing model validation methodology. We construct an asset pricing model that is estimated purely inductively — a statistical model. The reason for the introduction of the statistical model is that it does not suffer from the same econometric problems as the CBA. We shall compare the out-of-sample pricing accuracy of this statistical model with the CBA and FAPM. The statistical model may significantly out-perform the widely validated FAPM. However, due to the econometric problems induced by the ad-hoc implementation of the CBA, we believe that the statistical model will significantly out-perform the CBA. By comparing the performance of the FAPM and the CBA, with the statistical asset pricing model we can estimate the model bias induced by the ad-hoc implementation of the CBA.

Background

The capital asset pricing model (CAPM) is a one factor model that relates expected return to risk. Unfortunately, in the modern era (1969-1990) the capital asset pricing model (CAPM) fails. The three factor (Fama and French 1993; Fama and French 1996) model (FAPM) has been adopted as the initial alternative to the CAPM.

An alternative implementation of the FAPM, known as the CBA, which does not rely on the factor paradigm of (Ross 1976) and its risk interpretation, is to price assets by sorting and grouping securities on some exogenously determined criterion of interest, also known as

characteristics (i.e. size, book to market). The procedure is to sort the universe of stocks into a number of groups based on firm-size, book-to-market ratio and overall market-risk characteristics, allocate a particular stock to one of these groups and assign an expected return based on this membership — the characteristic based implementation (CBA) of the three factor Fama and French model (Daniel & Titman, 1997).

The problem with the CBA, as (Daniel, Titman et al. 2001) state:

> " The (Daniel and Titman 1997) results are clearly controversial; they reject a model that captures the central intuition of traditional asset pricing models in favour of a model that is ***almost completely ad-hoc***".

It is the ad-hoc nature of the CBA that is of most concern. It is our contention that the ad-hoc nature of the model leads to econometric problems, which in turn leads to model bias and consequently poor performance relative to its competitors.

Currently, the rivalry between the FAPM and the CBA is being interpreted as a contest between risk based asset pricing and non-risk based characteristic pricing. The weights on the FAPM factors are "risk" while the CBA, does NOT have any such weights and therefore no risk interpretation.

Furthermore, a number of econometric procedures required to estimate the model induce bias because they are ad-hoc and subsequent poor out-of-sample pricing accuracy ensues. (Daniel and Titman 1999) detail some of the econometric problems, levelled at the CBA, which include:

1. ***Truncation bias:*** the error that is induced due to an arbitrary division of data. In the context of the CBA, this occurs because of the arbitrary and ad-hoc division of each characteristic into groups.

2. ***Selection Bias:*** the bias that occurs when assets are selected for the construction of a portfolio. This is often done in asset pricing model testing as it reduces noise. In the context of the CBA this bias may occur because of the need to construct portfolios of securities that mimic characteristics (Black, Jensen et al. 1972).

3. ***Information loss/ Power loss:*** the problem associated with the loss of statistical power when testing for the accuracy of a model. In the context of the CBA, the power loss is due to the low dispersion of factor loadings which can occur because the number of groups in which a characteristic is divided is too many.

4. ***Data snooping bias:*** the error induced from fitting noise to a model (i.e. over fitting) due to visiting the data more than once. In the context of the CBA it may occur due to the use of characteristics similar to the FAPM 'risk' factors (Davis, Fama et al. 2000).

5. ***Measurement/ Errors in variable bias:*** the model bias which occurs because of the uncertainty in measuring model parameters. In asset pricing models error-in-variable bias occurs because of the need to include some measure of risk and return for individual stocks and also a measure of market return and these are measured with error. The CBA requires this information too, and is therefore prone to the same error (Miller and Scholes 1972).

There is a debate in the literature as to how carefully constructed the CBA has been. On the one hand, (Berk 2000) argues that the way in which the CBA is constructed means that it is prone to all of the econometric problems discussed above. Whereas (Daniel and Titman 1999) respond to these criticisms, arguing that the CBA methodology has addressed some, if

not all, of these potential problems. It is not clear who is correct and this justifies further analysis.

Also, studies which validate the CBA are limited, so it is not clear if the CBA is more accurate out-of-sample than its competitors. Whilst considerable effort has been spent on disentangling which asset pricing model is more accurate: the FAPM or the CBA, surprisingly, far less work has directly considered how well the CBA forecasts cross-sectional excess returns.

SIGNIFICANCE AND INNOVATION

This project tackles one of the most debated and central issues in modern finance: asset pricing theory. Asset pricing theory is vital to finance. It is used for decisions regrading trillions of dollars in asset allocation and capital expenditure.

Our aim is to validate a new and influential asset pricing model: the CBA, by comparing its pricing accuracy with a 'purely' statistically based asset pricing model. This approach to asset pricing validation is highly innovative. The statistical asset pricing model will use k-means clustering to find asset pricing clusters (i.e. asset prices that have similar historical returns); allocate an equity to one of these clusters; and, use the corresponding cluster mean as an ex-ante prediction of the asset's return in the next time period. This approach has been used successfully in the slightly different context (Brown and Goetzmann 1997) of mutual fund style analysis.

In comparison, the CBA procedure is to sort or 'bin' the universe of stocks in an ad-hoc way into a number of groups based on characteristics, such as: firm-size, book-to-market ratio and overall market-risk characteristics (e.g. momentum); allocate a particular stock to one of these groups and assign an expected return based on this membership or average 'bin' return.

The CBA and the statistically based asset pricing model are similar. They both predict out-of-sample returns using similar concepts of group and/or cluster membership. The CBA construction is ad-hoc. Due to its statistically sound construction, the new statistical asset pricing model (which is also consistent with the theoretical return generating process) is immune to the econometric problems of the CBA. In particular, its clusters and their number are obtained endogenously (i.e. without the need to be chosen a priori) thus eliminating truncation bias, selection bias and measurement error (1, 2 and 4 above); no ad-hoc division of characteristics into groups is required eliminating information loss (3 above); and a careful model fit is obtained, so as to reduce data snooping (5 above).

By comparing the performance of the FAPM and the CBA with the statistical asset pricing model we can gauge the size of the model bias induced by the ad-hoc implementation of the CBA. In this way we may be able to warn the academic community away from the CBA. On the other hand, if our hypothesis is wrong, and the CBA out-performs the statistical asset pricing model then this would offer new evidence that the CBA holds promise.

REFERENCES

Brown, S. and W. Goetzmann (1997). "Mutual fund styles." Journal of Financial Economics **43**: 373-399.

Chen, N., R. Roll, et al. (1986). "Economic forces and the stock market." Journal of Business **59**: 383-403.

Daniel, K. and S. Titman (1997). "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." The Journal of Finance **52**(1): 1-33.

Daniel, K. and S. Titman (1999). Sorting out "Sorting out sorts", Kellog School of Management, Northwestern University**:** 4.

Daniel, K., S. Titman, et al. (2001). "Explaining the Cross-Section of stock returns in Japan: Factors or Characteristics?" The Journal of Finance **56**(2): 743-766.

Fama, E. and K. French (1992). "The cross-section of Expected Stock Returns." The Journal of Finance **47**(2): 427-465.

Fama, E. and K. French (1993). "Common risk factors in the returns on stocks and bonds." Journal of Financial Economics **33**: 3-56.

Fama, E. and K. French (1995). "Size and book-to-market factors in earnings and returns." The Journal of Finance **50**(1): 131-155.

Fama, E. and K. French (1996). "Multifactor explanations of asset pricing anomolies." Journal of Finance **51**: 55-84.

Jegadeesh, N. and S. Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." Journal of Finance **48**: 65-91.

Lakonishok, J., A. Shleifer, et al. (1993). "Contrarian investment, extrapolation and investment risk." The Journal of Finance **49**: 1541-1578.

Ross, S. (1976). "The Arbitrage Theory of Capital Asset Pricing." Journal of Economic Theory **13**: 341-360.

Sharpe, W. (1964). "Capital asset prices: a theory of market equilibrium under condiions of risk." Journal of Finance **19**: 425-442.