Experimental–like: Factorial design of Multi-Factor Asset-Pricing Model with Randomization Approach

Proposed by Derek Cheng
PhD student

Hanken, School of Economics

Department of Finance, Helsinki, Finland

June 01, 2010

(1st draft)

Abstract

This paper proposes an Experimental–like: Factorial Design of Multi-Factor Asset-Pricing Model with Randomization Approach. Our design, incorporating randomization sampling procedure and full-factorial design, is to test the validity of standard CAPM and further extension to a multi-factor model. This proposed model is expected to 1) eliminate the problem of data-snooping , 2) provide unbiased statistical inference, 3) improve the power of predictability and 4) serve as a new practical guideline for empirical research in finance. More important, this model proposed is simple, reasonable and testable.

Key: ***Factorial Design, Randomization, CAPM; Data-snooping; Experimental Finance***

1. Introduction and Literature Reviews

"Natural selection is a mechanism for generating an exceedingly high degree of improbability."

Dangers of data–driven Inference in empirical portfolio model

Data-snooping is a problem of bias which occurs when the same set of dataset is used more than once for statistical inference. Data-snooping is fully aware in natural science. It involves problems of ethics, trust and integrity. Reusing same sources of datasets is highly "not" recommended in experimental design for hypothesis testing, especially in medical and pharmaceutical sciences. It is because rejection of null hypothesis can be easily manipulated by introducing data-snooping. However, problems of data-snooping seem not to be seriously alerted by empirical finance researchers in the past twenty years. Lo and Mackinlay (1990) have shown that:

> **"Tests of financial asset pricing models may yield misleading inferences when properties of the data are used to construct the test statistics. In particular, such tests are often based on returns to portfolios of common stock, where portfolios are constructed by sorting on some empirically motivated characteristic of the securities such as market value of equity. Analytical calculations, Monte Carlo simulations, and two empirical examples show that the effects of this type of data snooping can be substantial."**

In this paper, Lo and Mackinlay (1990) have also highlighted that data-snooping bias can be also interpreted as increasing the power of tests by introducing "induced order statistics". With the complementary relationship of type I and type II errors, increasing the power of tests is equivalent to increasing the Type I error. Unfortunately, in the standard CAPM model or multifactor asset-pricing model setting, alpha is equal to zero which is restricted by theory of finance. Testing the validity of CAPM model with Fame-French

Model Protocol is equivalent to testing the null hypothesis of H_0: α **=0**. Rejection of null hypothesis can be easily manipulated or controlled by introducing data-snooping implicitly, hence, increasing the power of tests, consequently, increasing the Type I error , and finally drawing the wrong conclusion. In other words, under the Fame-French 3 factor Model(FF3M) Protocol, it is so not difficult to reject the null hypothesis of the alpha in CAPM model by introducing data-snooping bias. How does the Fama-French Model Protocol introduce data-snooping bias implicitly?

Potential pitfalls of Fama-French 3-factor design Protocol

Fama and French (1993) propose that the 3-factor model is to capture cross-sectional variation in average stock return and can be expressed as

$$r_t - r_f = \alpha_{FF} + \beta_{t,MKT} r_{MKT} + \beta_{t,SMB} SMB + \beta_{t,HML} HML + \varepsilon_t \quad . \qquad (1)$$

Their design protocol is to adopt the technique of design of experiment (DOE) with two-way additive balanced ANOVA design protocol, plus extra independent variable (excess market risk premium), constituting as a 3-factor model. The second additive factor, SMB is constructed by first sort of firm size while the third additive factor, HML is constructed by second sort of B/M ratio. In this design protocol, they suggest that there is no interaction terms between these two risk factors.

To mimic these two risk independent factors and one dependent variable **(** $r_t - r_f$), Fama-French(1993) propose to construct three mimicking portfolios based on the two characteristics of firm's size and BE/ME ratio with the "6 abnormal rules", (in the sense that violate the assumptions of DOE design protocol):

- sorting out of sorts from all stocks by size and by B/M or visa- versa
- ranking between groups within ranked groups in terms of quartiles or deciles
- Averaging the values within groups and between group to form 25 portfolios
- differencing minimum and maximum portfolio value within groups to form SMB

- differencing maximum and minimum portfolio value between groups to form HML
- using the same sources of 25 portfolios as dependent variables for regression analysis

Reviewing to Fama-French 3-factor model protocol described above, we found that two-way additive ANOVA design technique becomes the "two-way sorts" mimicking strategy in their model setting. Besides, it is not so difficult to uncover that violation of reusing the same sources-dataset is so obvious in mimicking portfolios of dependent and independent variables.

Although two-way ANOVA design is a common scientific technique to investigate the factor effects between groups and within groups, a high power of predictability requires an unbiased experimental design platform without data-snooping setting. Assumptions of normal distribution, homogeneity in population, randomization of data sampling and dependent variables (response variables) not known before empirical design should be obeyed; otherwise t-test or F-test statistic is biased, invalid and meaningless.

An article, "Sorting out sorts", published by Berk (2000) have also pointed out that by using sufficient sorting and grouping, it is possible for researchers either 1) to reduce the explanatory power in order to keep control alpha equal to zero or 2) to increase the explanatory power in order to keep control alpha not equal to zero. Berk (2000) also emphasizes that

> **"Sorting mechanism determines what part is attributable to the between-sample variation. This apportioning of the variation directly affects the power and size of the statistical test, thereby affecting the statistical inference."**

Unlike natural sciences, empirical asset pricing modeling predominantly studies in non-experimental data of realized returns. Experimental design is practically inflexible in empirical finance setting. Sorting mechanism is a very common but dangerous practice in empirical asset pricing model testing. Data-snooping bias seems not to be alerted by financial researchers. Many researchers mindlessly adopt the so-called "benchmark"

Fama-French 3-factor Model Protocol to mimic portfolios for hypothesis testing in Capital asset-pricing model (CAPM), or further extend the FF3M Protocol with incorporating another theory (such as inversely proportional relationship) in order to seek for the high explanatory power for CAPM anomalies by trading-off with non-zero alpha.

For instance, Chui and Wei (1998), mindlessly adopting the procedures of FF3 model protocol, and claim that a strong size effect in four Pacific-Basic emerging markets (Hong Kong, Malaysia, Korea and Thailand). Arshanapalli, Coggin, and Doukas (1998) use the similar Fama-French mimicking portfolio protocol to draw the biased conclusion that higher book-to-market stocks always outperform lower book-to-market stocks in their analysis of eighteen international equity markets. More recently, Chen and Zhang (2009), also adopt the similar protocol of FF3M to claim that Tobin Q-theory can be incorporated in FF3M Protocol to explain CAPM anomalies. It is not surprising that over the past 10 years many research papers and literatures in empirical asset pricing still mindlessly follow or replicate the similar Fama-French Protocol to mimic portfolios for CAPM testing with or without incorporating new theory of finance. More surprisingly, those research papers followed FF3M Protocol, are only to justify their claims on how high explanatory power of their factor model in term of R square or how derivate the alpha from zero with F test on CAPM anomalies. Their papers are totally neglected any discussion on “how valid their design models are”. Scientifically speaking, if their design models (adopted FF3M Protocol) are not valid, there is no reason to trust the biased statistical figure of R Square or F-test on alpha (with standard OLS approach without any diagnostic correction in their linear n-factor model), thereby, there is no reason to believe their final claims that their n-factor model can parsimoniously summarize the cross-sectional average return, because their models are usually lacking of predictive power. We agree that in empirical asset pricing test, GRS test statistics is one of the robust test for the case of non-normality, however, there is no literature mentioning that GRS test statistics is also robust or immunized under data-snooping bias design protocol. GRS test statistics does a good job for joint testing under non-normality but it does not help to signify the problem of data-snooping bias and even not to signify the invalidity of linear 3-factor model.

White (2000) points out that

> **"Whenever a 'good' forecasting model is obtained by an extensive specification search, there is always the danger that the observed good performance results not from the actual forecasting ability, but is instead just luck."**

In other word, some so-called better models have better explanatory power for "special" datasets, however, their better explanatory power are solely "by chance" within the "special" sorting or grouping sample data, but "not by statistical inference". Usually such kinds of invalid models have no (or very low) predictive power of (at least one-step ahead) forecasting.

2. Purpose of the research proposal

There are various arguments for and against the validity of FF3M Protocol. Most empirical researchers in Finance believe that mimicking portfolio with same sources of datasets is the inevitable way in financial portfolio modeling. They also enforce their excuse that empirical finance is non-experimental and cannot keep financial market fixed or controlled. They claim that they no better choice except the biased FF3M model. Therefore, extensive literatures viscously focus on biased conclusion whether "CAPM is dead" or "Fama-French multi-factor model is alive", neglecting the fundamental problems of data-snooping implicitly pre-existed in their design setting. Scientifically speaking, any argument and conclusion drawn from this non-scientific design are not strong enough to address the empirical findings of theory of CAPM model.

In order to remedy the problem of data-snooping, and to provide a practical guideline of constructing mimicking portfolios, this paper proposes an experimental–like: "Factorial

Design of Multi-Factor Asset-Pricing Model with Randomization Approach".

This new design platform, incorporating randomization sampling procedure, is to test the validity of standard CAPM model and further extension to a multi-factor model. The proposed model with randomization approach is to eliminate the problem of data-snooping, provide unbiased statistical inference and improve the power of predictability. More important, our proposed model is simple, reasonable and testable.

The rest of the paper is organized as follows: Section 3 outlines Fisher's Randomization and Portfolios construction Algorithm. Section 4 discusses the procedure of implementation. The final section is concluding remarks.

3. Fisher's Randomization and Portfolios construction Algorithm

The requirement of randomization in experimental design is first stated by Ronald. A. Fisher, statistician and geneticist, in 1925 in his book "Statistical Methods for Research Workers". Randomization of sampling serves two purposes: it eliminates bias and enables a valid test of significance.

An example of outlining the factorial design of portfolios construction algorithm for L-level in each factor, where L=2, is the example shown in Appendix 1. The general design proposed in this paper is as follows:

1. Pre-clustering population size of total stocks "without" ranks and sorts
2. Formation of L^K portfolios, where K is number of factors with L-level in each.
3. Randomization sampling of 1% of the stocks from the total population for each level.
4. Construction of row factor portfolios within groups.
5. Construction of column factor portfolios between groups.
6. Construction of corresponding portfolios of response variable with 1% of the total population of stock.

In this design, we propose to use full-factorial design (L^K) for formation of portfolios where K is numbers of factor with L-level in each. The advantage of full-factorial design is to capture the effects of all independent variables in each level simultaneously.

Randomization of sampling is to eliminate the effect of data-snooping.

The decision rule of how to test standard CAPM and multi-factor asset pricing model is shown in Appendix 2. To test the standard CAPM, we suggest the procedures as follows:

1. Run a time-regression of standard CAPM with this 25 randomization portfolios, if the intercept terms are insignificant (α=0) for all 25 test, then Standard CAPM is valid; otherwise, standard CAPM is invalid.

2. If standard CAPM is invalid, then, add two additional factors in previous regression model. Re-do a new randomization sampling process to form L^K portfolios for two factors with L-level in each. Run L^K portfolios time-series regression separately. If intercept terms of all cases are still significant, repeat the previous procedure and add one more factor into the model, therefore, L^{K+1} portfolio are formed with L-level in each.

3. Repeat the same process and, until all intercept terms are insignificant and stop iterations.

4. If all the intercept terms are insignificant, then finally run a cross-section regression to test whether all factors are priced or not.

5. In order to preserve the randomization behavior, a constraint of total portfolios $L^k \leq 25\%$ of total stock population is imposed. That is, 1% of sampling from total population for each group is reduced or adjusted accordingly.

4.Implementation

In this section, we propose the procedure of implementation in our new design. Our objective is to run a pilot test for U.S. markets at the first stage and then extend to the test to other countries in later stage.

In the pilot stage, we propose all data extracted from CRSP and Compustat database. The total population of stocks (N) should include active, non-active and dead stocks. Including non-active and dead stocks is to reflect the real information in U.S financial market at that time point of frame. The date range is assigned from January 1960 to June

2010. All variables are in monthly figures. For this design, we choose a 3-level for each factor. A total of 3^K portfolios are formed for the multi-factor asset pricing test. All variables should not rank and sort. The level of each factor assumes to be the exogenous property of each stock. Pre-clustering of the total population of stock without ranks and sorts is for data stratification's purposes. Each listed Company has a unique "permno" code in CRSP database. This code can be served as a random number for sampling purposes. We also follow the Fama and French's variable definitions for apple-to-apple based comparison. For example, the firm size (FS) is defined as the Market equity of June of year t (ME (t_{atJune})) whereas for accounting variables, like Book-to-market (BE/ME), it uses a firm's market equity at the end year ($t\text{-}1_{\text{at December}}$) of December for computation. With 6-month time gap setting, it is to ensure all the accounting variables are available and known for investors who have sufficient time to respond in the financial markets. All returns are calculated in scale of natural logarithm. To proxy risk free rate, the 1-month TBill is used. The NYSE/AMEX/NASDAQ/ARCA equal weight index is chosen to proxy market return portfolio.

5.Concluding Remarks

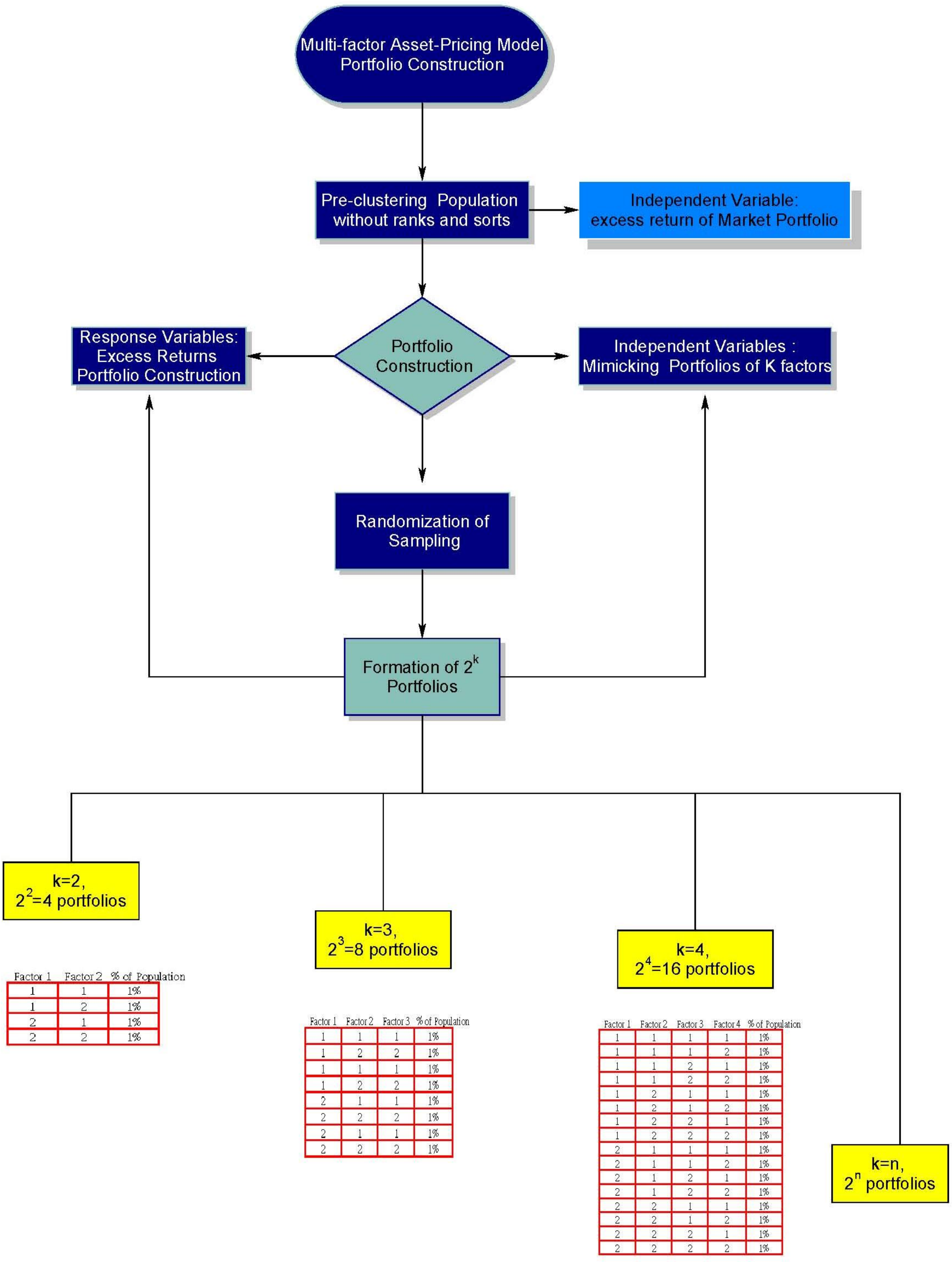
This paper proposes an experimental–like: "Factorial Design of Multi-Factor Asset-Pricing Model with Randomization Approach". Our new design platform, incorporating randomization sampling procedure and full-factorial design, is to test the validity of standard CAPM and further extension to a multi-factor model. This proposed model is expected to 1) eliminate the problem of data-snooping, 2) provide unbiased statistical inference, 3) improve the power of predictability and 4) serve as a new practical guideline for empirical research in finance. More important, this model proposed is simple, reasonable and testable.

Further advantages of this design can save the cost of resources and reduce the redundant procedures of sorting and ranking suggested by Fama and French model (1993). The weakness of our design is not suitable for a small or short history of financial market. It is highlighted that the feature of our design is to optimally balance between predictable power and explanatory power whereas FF3M Protocol is to maximize explanatory power solely, regardless of fundamental concept of statistical inference: "Unbiasedness" in the econometrical modeling.

The foreseeable impact of our proposal is to introduce a new research channel, linking up empirical finance and experimental finance. Theory of experimental finance has a potential for growth. Application of DOE techniques in experimental finance is valuable for further extension in portfolio and behavior finance.

Appendix 1: An Example of Factorial Design (Level =2)

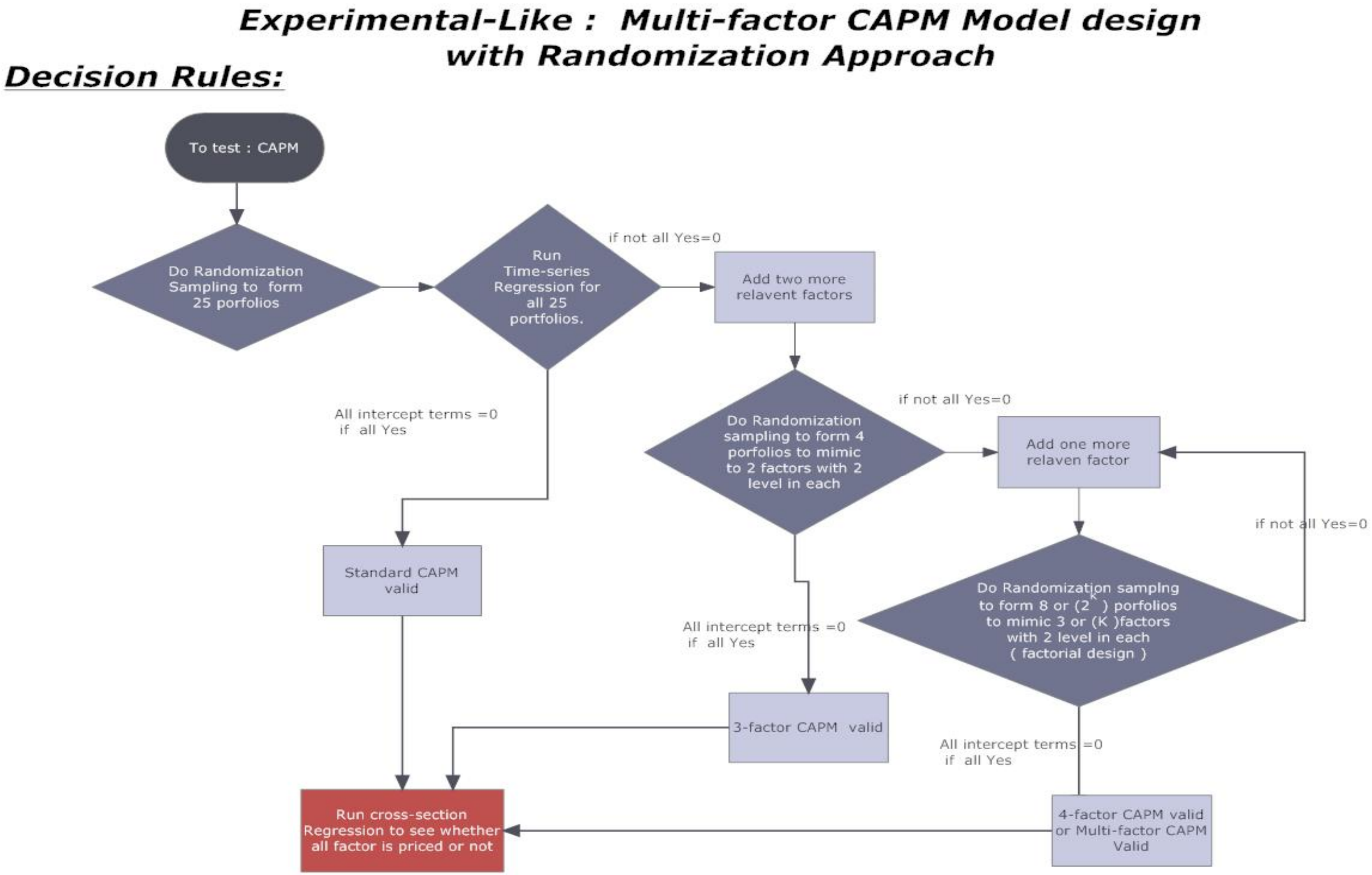
Portfolio construction Algorithm



Factor 1	Factor 2	% of Population
1	1	1%
1	2	1%
2	1	1%
2	2	1%

Factor 1	Factor 2	Factor 3	% of Population
1	1	1	1%
1	2	2	1%
1	1	1	1%
1	2	2	1%
2	1	1	1%
2	2	2	1%
2	1	1	1%
2	2	2	1%

Factor 1	Factor 2	Factor 3	Factor 4	% of Population
1	1	1	1	1%
1	1	1	2	1%
1	1	2	1	1%
1	1	2	2	1%
1	2	1	1	1%
1	2	1	2	1%
1	2	2	1	1%
1	2	2	2	1%
2	1	1	1	1%
2	1	1	2	1%
2	1	2	1	1%
2	1	2	2	1%
2	2	1	1	1%
2	2	1	2	1%
2	2	2	1	1%
2	2	2	2	1%


Experimental-Like : Multi-factor CAPM Model design with Randomization Approach
Decision Rules:
To test : CAPM
Do Randomization Sampling to form 25 porfolios
Run Time-series Regression for all 25 portfolios.
if not all Yes=0
Add two more relavent factors
All intercept terms =0 if all Yes
Standard CAPM valid
Do Randomization sampling to form 4 porfolios to mimic to 2 factors with 2 level in each
if not all Yes=0
Add one more relaven factor
if not all Yes=0
All intercept terms =0 if all Yes
Do Randomization sampling to form 8 or (2^K) porfolios to mimic 3 or (K)factors with 2 level in each (factorial design)
3-factor CAPM valid
All intercept terms =0 if all Yes
Run cross-section Regression to see whether all factor is priced or not
4-factor CAPM valid or Multi-factor CAPM Valid

. Reference

Arshanapalli, B., Coggin, T. D., & Doukas, J. (1998). Multifactor asset pricing analysis of international value investment strategies. Journal of Portfolio Management, 24,10−24.

Chui C. W., & Wei, K. C. (1998). Book-to-market, firm size, and the turn-of-the-year effect: Evidence from Pacific-basin emerging markets. Pacific-Basic Finance Journal, 6, 275−293

David Afshartous, Michael Wolf (2007), Avoiding 'data snooping' in multilevel and mixed effects models, Journal of the Royal Statistical Society: Series A (Statistics in Society) 170 (4), 1035–1059.

Eugene F. Fama and Kenneth R. French , Common risk factors in the returns on stocks and bonds, Journal of Financial Economics, Volume 33, Issue 1, February 1993, 3-56

Fisher RA. The Design of Experiments. Edinburgh: Oliver and Boyd, 1935

Hansen,P. R.(2005). A test for superior predictive ability, Journal of Business and Economic Statistics, 23, 365–380.

Holm,S.(1979). A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics, 6, 65–70.

Hsu, P. and C. Kuan (2005). Reexamining the portability of technical analysis with data snooping checks, Journal of Financial Econometrics, 3, 606–628.

J.S.Huxley, A.C.Hardy and E.B.Ford, London: Allen and Unwin, 1954, Reported by J.S.Huxley in Evolution as a Process.

Keith S. K. Lam, The relationship between size, book-to-market equity ratio, earnings–price ratio, and return for the Hong Kong stock market, Global Finance Journal, Volume 13, Issue 2, 2002, 163-179

Liu, W.,2006. A liquidity augmented capital asset pricing model. Journal of Financial Economics 82, 631–671.

LO, A.W. and A.C. MACKINLAY, Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies, 1990, Volume 3, 431-468.

Politis, D.N. and J.P. Romano (1994), "The Stationary Bootstrap", Journal of American Statistical Association, 89, 1303-1313.

Liew, J., Vassalou, M.,2000. Can book-to-market, size and momentum be risk factors that predict economic growth? Journal of Financial Economics 57, 221–245.

Sullivan, R A. Timmermann, and H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap, Journal of Finance, 54, 1647–1691.

Sullivan, R. A. Timmermann, and H. White (2001). Dangers of data mining: The case of calendar effects in stock returns, Journal of Econometrics, 105, 249–286.

Verardo, M., 2008. Heterogenous beliefs and momentum profits. Journal of Financial and Quantitative Analysis forthcoming.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.