Factor models in empirical asset pricing

Peter Schotman
Maastricht University

25 September 2017

1 Schedule

This PhD minicourse will take place at the Swedish House of Finance, room Fama, March 5-9, 2018. Sessions are scheduled daily 10:00-12:00 and 14:00-16:30 on Monday-Thursday, and 14:00-16:00 on Friday.

2 Introduction

Factor models are the core of empirical asset pricing. For asset pricing they provide the link between covariances and expected returns. The asset pricing condition

$$\mathrm{E}[R_i] = \beta_i' \lambda,$$

links the expected return $\mathrm{E}[R_i]$ for asset i to the vector of its factor beta's and a vector of factor risk premiums λ.

In portfolio management factor investing aims at simplifying the problem of portfolio construction by reducing the dimension of the space to search in. Factor models reduce the task of searching among thousands of assets and trading strategies to the more tractable problem of finding the preferred risk-return tradeoff among a small number of factors.

While conceptually elegant, the empirical identification of factors is a difficult problem. One problem is that there are too many factors. Cochrane (2011) coined the term 'factor zoo' referring to the literally hundreds of factors discovered so far. Factors are found in various ways:

- The Fama-French factors are based on economic intuition for firm characteristics that may be indicators of expected returns.
- A pure statistical approach starts from estimating the principal components from a panel of returns.

- Many factors originate from a trading strategy that seemed to produce abnormal returns. Examples are momentum, price reversal, low beta, and idiosyncratic volatility.
- Much more fundamental are factor models derived from a macroeconomic equilibrium asset pricing model. Examples are long-run consumption growth, macroeconomic uncertainty, or past consumption habits.

Typical questions in empirical asset pricing all start with factors:

- How many factors are there? Given the many factors 'discovered' so far, 'taming the factor zoo' (the title of the Feng, Giglio and Xiu (2017) paper) is high on the agenda.
- Which factors are priced? For portfolio management it does not make sense to take factor risk for which there is no reward.
- How stable is the factor structure? This adds an additional layer of complexity, but also may reduce the number of factors. A single factor model with time-varying beta's is often equivalent to a multifactor model, and vice versa.
- What is the economic interpretation of the factors?

Factor models are not only relevant for stocks, but for bonds as well. Here the consensus is that there are three factors in bonds returns. But the search is for factors that explain both bonds as well as stocks. In fact, the search is for factors that explain multiple asset classes around the world.

The course will focus on new statistical methodologies for estimating factor models and their use in evaluating asset pricing models and in constructing portfolios. The emphasis will be on the dimensionality of both the number of potential factors and the large cross-section of available assets and potential dynamic trading strategies. The aim for this course is to bring you up to date with ideas and research methods.

3 Outline

The topic is in the interface of empirical asset pricing and econometrics. That means we will use insights from both fields. At the same time this is not a course where we go very deep into the details of either asset pricing theory or proofs of the properties of statistical methods. Motivated by the finance questions we will look at how different methods can help in finding answers.

In class we will mostly discuss recent (and some older) research papers. A lot of the finance and financial econometrics background information can be found in textbooks.

- Cochrane (2005) is the classic asset pricing text. Part I covers the fundamental relations between stochastic discount factors, factor models, and mean-variance efficient frontiers. The same material can also be found in other asset pricing texts.
- Even 20 years after publication, chapters 5 and 6 of Campbell, Lo and MacKinlay (1997) still contain the basics for estimating and testing linear asset pricing models.
- Ang (2014) contains the basics of factor investing, explaining why this makes sense as an investment approach.

During the course we explore the potential of advances in statistical learning for problems in empirical asset pricing and portfolio management. Hastie, Tibshirani and Friedman (2010) provides an overview of many techniques at an accessible intuitive level. We will discuss some of the material in the book in applications to financial returns data.

Teaching will as interactive as possible. Research papers sometimes offer different perspectives on the same problem. An important element of the discussions is a critical appraisal of the papers, especially when we look at unpublished working papers. For most sessions I will ask a subgroup of students to take the lead by giving a brief introduction based on one or two articles. These presentations are meant to start a discussion about the main ideas, arguments and evidence. What are the two or three empirical results that you think stand out? What, if any, are the main methodological improvements?

I will also encourage thinking about research ideas related to the papers we discuss in class. What would be interesting next questions? How could such questions potentially be addressed?

4 Topics

The literature for the topics is indicative. Since the literature is evolving, I may replace some of the articles referenced below by a recent working paper. For each session we will concentrate on two papers and have a more cursory look at others.

4.1 Factor investing

In class we will discuss various attempts to define the space of factor portfolios. After the celebrated Fama-French 3-factor model for US equities, Fama and French (2015) extend their model to 5-factors. Hou, Xue and Zhang (2015, 2017) propose new factors and assess how well different factor models explain many anomalies. Fama and French (2012) explore if global Size and Book-to-Market factors span the space of equity returns around the world. Asness, Moskowitz and Pedersen (2013) take the search for factors beyond equity. They find evidence that value and momentum may be two of such global factors that work across asset classes.

Chen, Roll and Ross (1986) introduced the seminal set of macro factors in empirical asset pricing. Cochrane (2017) reviews factors that emerge from macro-economic theories.

Factor investing and its relation to machine learning has become an important focus for industry practice. The large asset management firms all advertise factor investing (or smart beta) approaches. These strategies are further often labeled by terms such as 'machine intelligence robo strategies'. Vazalet and Roncalli (2014) and Clarke, de Silva and Thorley (2016) discuss some of the practical issues.

4.2 Large N

We look at econometric methods for determining the number of factors among a large universe of individual assets and trading strategies. Bai and Ng (2012) develop information criteria to estimate the size of the factor space in case we have a large cross-section of assets. Recent work is discussed in Bai and Ng (2017). There is a lively literature on ways to put more structure on the models using networks or graphs. An example is Fan, Liao and Wang (2016), who relate factor loadings to observable variables to reduce the dimensionality. In the same spirit, Kozak, Nagel and Santosh (2017) specify the stochastic discount factor as a linear function of the entire cross-section of stock returns and use Bayesian priors for regularization to reduce the dimensionality.

Feng, Giglio and Xiu (2017) search for priced factors that explain the covariance structure of returns. Other studies look at which of many characteristics explain the cross-section of returns. Freyberger, Neuhierl and Weber (2017) is an example using Lasso techniques.

With large N the notion that an asset pricing model correctly prices every asset and every trading strategy also becomes questionable. Building on distance measures for misspecification developed by Hansen and Jagannathan (1997), Lönn and Schotman (2017) apply boosting methods to evaluate and compare alternative asset pricing models in a large universe of assets.

4.3 Statistical Pitfalls

Everyone searches for the pot of gold: a profitable trading strategy. What is published are successes. The many failures are underreported. Therefore, part of what we see in the Kenneth French data library[1] are many portfolio sorts and trading strategies that appeared to work very well at some point and therefore became of interest to follow. McLean and Pontiff (2016) analyze what happens in the years after a trading strategy is discovered. One reason more and more factors are being discovered is that the number of assets searched

[1] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

for profitable strategies is expanding. More firms get listed, more countries have good stock market data.

Harvey, Liu and Zhu (2016) look at the problem from a statistical testing perspective. How do we deal with testing multiple hypotheses. What is the statistical evidence when you find success after trying many times. How many of the discovered successful trading strategies are 'false positives'? Harvey. Liu and Zhu (2016) list more than 300 factors, and express doubt on the methodology and statistical tests underlying new factors. Harvey and Liu (2015) apply these ideas to discounting Sharpe ratios. Related techniques for performance measurement of mutual funds have been developed in Barras, Scaillet and Wermers (2010).

White's (2000) 'reality check' was an earlier econometric technique to correct tests for searching among many similar trading strategy designs. For example, in constructing a momentum strategy we have the freedom to specify the number of months looking backward to select stocks with momentum, the number of months to hold a portfolio, and the number or percentage of stocks included in the momentum portfolios. Adding further screens on size or particular industries adds even more scope. In the end we are optimizing over hundreds of possible momentum strategies. Simply reporting the performance of the best will not give the proper statistical significance.

More generally, what we can learn from the machine learning literature are techniques to avoid overfitting and methods that can correct performance measures for data mining. An example is cross-validation, which can be helpful in discounting optimized Sharpe ratios of trading strategies.

4.4 Time-varying covariances

In a parallel literature financial econometrics has made much progress in accurately estimating covariance matrices using high frequency data. Starting with improved univariate 'realized volatility' measures, many current developments address the much harder problem of estimating a large dimensional covariance matrix with high frequency data. An example of the current state of the art is Hautsch and Voigt (2017), which with a 300 dimensional matrix is still far from real big N.

An intermediate approach, combining daily time series and monthly cross-sectional data is Cosemans, Frehen, Schotman and Bauer (2016). They propose an efficient estimator for time-varying beta's for individual stocks. They apply the estimator to asset pricing, portfolio choice and risk management, and show the benefits of both using the large cross-section of individual stocks as well as allowing for the strong time variation in beta's.

When allowing for time-varying factor loadings, the number of factors per se is not always a well-defined concept. A single factor model with time-varying factor loadings is often equivalent to a multifactor model with constant factor loadings. For example, suppose individual firm betas's differ with the state of the business cycle z_t. Then the single factor CAPM model with time-varying beta's $\beta_{i,t-1} = b_{0i} + b_{1i} z_{t-1}$ is equivalent to a 2-factor model with constant factor loadings where the two factors are the return on the market portfolio $y_{M,t}$ and the 'managed portfolio' return $z_{t-1} y_{M,t}$. The same effect could also be achieved by forming portfolios. Individual stocks have time-varying beta depending on the book-to-market ratio. Sorting stocks on their book-to-market ratio may then result in portfolios with stable beta's. This works for many trading strategies. A trading strategy often involves rebalancing a portfolio to maintain a desired profile. The factor loadings of the trading strategy may then be constant even though the individual stock beta's are not.

In empirical tests it is therefore important to be precise about the space of traded assets. Are we interested in individual stocks or in explaining the performance of a trading strategy? In the language of Cochrane's (2005) *Asset Pricing* the difference is between conditional and unconditional models, and in realizing that a trading strategy with time-varying portfolio weights is just another asset. The Kenneth French database contains a large number of equity portfolios sorted on many different firm characteristics, previous returns and risk measures.

4.5 Portfolio management

Having developed a factor model the next step is to construct a factor based portfolio. Much of portfolio management is implicit in the idea of factor investing. Once the factors have been identified, the optimal portfolio invests in the factor mimicking portfolios.

For portfolio management factors need to be translated to trading strategies. For some factors this is easy, since the factor is defined as the trading strategy that leads to the abnormal returns. For economically motivated factors, such as inflation or an interest rate factor, a trading strategy that replicates the factor needs to be constructed. Lamont (2001) pioneered regression models for constructing factor mimicking portfolios. But when the number of assets, and even more so number of dynamic trading strategies, becomes large, standard regression models become problematic. In such cases we need to impose flexible restrictions that allow regressions where the number of potential explanatory variables exceeds the number of time series observations. Implementing this requires regularization techniques, such as Lasso or Boosting, or Bayesian priors. Lönn and Schotman (2017) construct an efficient frontier from economic tracking portfolios based on a large universe of assets. They build on the tracking portfolio methodology of Lamont (2001) to identify portfolios of stocks that can mimic an economic factor.

Portfolio optimization with many assets is notoriously difficult. Naive mean-variance optimization, and even some advanced methods, generally leads to portfolios with very poor out-of-sample performance. As shown in DeMiguel, Garlappi and Uppal (2009) an equally weighted benchmark is hard to beat. Factor models are one way to deal with large scale portfolio problems. Alternative approaches also exist. Common to successful methods is some form of regularization to avoid extreme portfolio weights. In finance Jagannathan and Ma (2003) have shown that short-sell constraints can be a valuable restriction on optimized portfolios. More generally, machine learning techniques have had some success in regularizing large scale portfolio problems such that performance improves. One example is Fan, Zhang, and Yu (2012) who apply the LARS algorithm. Heaton, Polson and Witte (2016) take this a step further and apply 'deep learning' techniques to portfolio optimization. Technically this will be beyond what we will cover in the course. For those interested Goodfellow, Bengio and Courville (2016) is a recent textbook on 'deep learning'.

Other approaches to regularization of portfolio weights have been motivated by model uncertainty. Factor models again prove useful. In an early contribution Goldfarb and Iyengar (2003) apply robust optimization to an uncertain factor model. In a setting with multiple competing models, Lutgens and Schotman (2010) show that shrinkage towards factor portfolios strongly improves portfolio performance.

5 Literature

Ang, A. (2014), *Asset Management: A Systematic Approach to Factor Investing*, Oxford University Press.

Asness, C., Moskowitz, T. and Pedersen, L. (2013), 'Value and Momentum Everywhere', *Journal of Finance* **68**, 929–985.

Bai, J. and Ng, S. (2002), 'Determining the Number of Factors in Approximate Factor Models', *Econometrica* **70**, 191–221.

Bai, J. and Ng, S. (2017), Principal Components and Regularized Estimation of Factor Models. arXiv working paper 1708.08137v1.

Barras, L., Scaillet, O. and Wermers, R. (2010), 'False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas', *Journal of Finance* **65**, 179–216.

Campbell, J., Lo, A. and MacKinlay, C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Cazalet, Z. and Roncalli, T. (2014), Facts and Fantasies About Factor Investing. SSRN working paper 2524547.

Chen, N., Roll, R. and Ross, S. (1986), 'Economic Forces and the Stock Market', *Journal of Business* **59**, 383–403.

Clarke, R., de Silva, H. and Thorley, S. (2016), 'Fundamentals of Efficient Factor Investing (corrected May 2017)', *Financial Analysts Journal* **72**, 9–26.

Cochrane, J. (2011), 'Discount Rates', *Journal of Finance* **56**, 1047–1108.

Cochrane, J. (2017), 'Macro-Finance', *Review of Finance* **21**, 1–42.

Cochrane, J. H. (2005), *Asset Pricing*, Princeton University Press.

Cosemans, M., Frehen, R., Schotman, P. and Bauer, R. (2016), 'Estimating Security Betas Using Prior Information Based on Firm Fundamentals', *Review of Financial Studies* **29**, 1072–1112.

DeMiguel, V., Garlappi, L. and Uppal, R. (2007), 'Optimal versus Naive Diversification:

How Inefficient Is the 1/N Portfolio Strategy?', *Review of Financial Studies* **22**, 1915–1953.

Fama, E. F. and French, K. R. (2012), 'Size, value, and momentum in international stock returns', *Journal of Financial Economics* **105**, 457–472.

Fama, E. F. and French, K. R. (2015), 'A Five-Factor Asset-Pricing Model', *Journal of Financial Economics* **116**, 1–22.

Fan, J., Liao, Y. and Wang, W. (2016), 'Projected Principal Component Analysis in Factor Models', *Annals of Statistics* **44**, 219–254.

Fan, J., Zhang, J. and Yu, K. (2012), 'Vast Portfolio Selection With Gross-Exposure Constraints', *Journal of the American Statistical Association* **107**, 592–606.

Feng, G., Giglio, S. and Xiu, D. (2017), Taming the Factor Zoo. SSRN working paper 2934020.

Freyberger, J., Neuhierl, A. and Weber, M. (2017), Nonparametric Dissection of the Cross-Section of Expected Stock Returns. NBER working paper 23227.

Goldfarb, D. and Iyengar, G. (2003), 'Robust Portfolio Selection Problems', *Mathematics of Operations Research* **28**, 1–38.

Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, MIT Press.

Hansen, L. and Jagannathan, R. (1997), 'Assessing Specification Errors in Stochastic Discount Factor Models', *Journal of Finance* **52**, 557–590.

Harvey, C. and Liu, Y. (2015), 'Backtesting', *Journal of Portfolio Management* **Fall**, 13–28.

Harvey, C., Liu, Y. and Zhu, H. (2016), '... and the Cross-section of Expected Returns', *Review of Financial Studies* **29**, 5–68.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning*, Springer.

Hautsch, N. and Voigt, S. (2017), Adaptive Large-Scale Portfolio Allocation under Transaction Costs: Mixing High-Frequency and Low-Frequency Data. WU Vienna working paper.

Heaton, J., Polson, N. and Witte, J. (2016), 'Deep Learning for Finance: Deep Portfolios', *Applied Stochastic Models in Business and Industry* **33**, 3–12.

Hou, K., Xue, C. and Zhang, L. (2015), 'Digesting Anomalies: An Investment Approach', *Review of Financial Studies* **28**, 650–705.

Hou, K., Xue, C. and Zhang, L. (2017), A Comparison of New Factor Models. SSRN working paper 2520929.

Jagannathan, R. and Ma, T. (2003), 'Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps', *Journal of Finance* **58**, 1651–1683.

Kozak, S., Nagel, S. and Santosh, S. (2017), Shrinking the Cross Section. SSRN working paper 2945663.

Lamont, O. (2001), 'Economic Tracking Portfolios', *Journal of Econometrics* **105**, 161–184.

Lönn, R. and Schotman, P. (2017), Empirical asset pricing with many assets and short time series. Working Paper, Maastricht University.

Lutgens, F. and Schotman, P. (2010), 'Robust Portfolio Optimisation with Multiple Experts', *Review of Finance* **14**, 343–383.

McLean, R. and Pontiff, J. (2016), 'Does Academic Research Destroy Stock Return Predictability?', *Journal of Finance* **71**, 5–31.

White, H. (2000), 'A Reality Check for Data Snooping', *Econometrica* **68**, 1097–1126.