Factor models in empirical research

Richard

July 2017

1 Fama-French

Fama and French (1992) test the CAPM. They find that *beta* seems to relate to size of firms. In particular, small firms have higher returns compared to large firms. Thus, they control for size when create the beta factor in their 1992 paper. After control and re-test the joint roles of market beta, size, E/P, leverage, and B/M, they find that:

- When measured it unrelated to size, β is not related to stock returns, even when it's the only explanatory variable.
- Used alone, size, E/P, leverage, and B/M equity have explanatory power
- Used combination, Size and B/M capture the cross-sectional variation of stock returns, and they reflect the multidimensional risks in asset pricing. What factors they proxy for? Distress risk or investment?
- Leverage effect is related to B/M effect
- E/P effect weakens after we control $size$ and B/M

Thus, Fama and French (1992) find out two common factors: $size$ and B/M.

Fama and French (1993) identifies 5 common risk factors in pricing stocks and bonds.

- For stocks: market factors, size, B/M
- For bonds: maturity $TERM$ (*term structure* = long-term gov bond - one month T-bill rate) and default risk DEF (default premium = long-term corporate bond - long-term gov bond)

In 1993, they expand their old paper by adding bond assets and other factors to explain bond returns such as *term structure*. In addition, instead of Fama-Macbeth approach, this paper uses Black, Jensen, Scholes (1972) time-series regression approach. Monthly stock and bond returns are regressed on the returns to a market portfolio of stocks and mimicking portfolio for size, B/M, and term structure risk factors in returns. The slopes are *factor loading or risk-factor sensitivities* for stocks and bonds. This method, later, will be the standard in testing factor models.

In the model, the dependent variable is excess returns (stock and bond returns minus one-month Treasury bill rate). Bonds include 2 government bonds and 5 corporate bond portfolios. Stocks include 25 of 5×5 size-BM portfolios. The idea is if the intercept from three-factor regressions (excess market return, mimicking size, and mimicking B/M factors) are close to zero, the model seems to do good job to explaining the cross-section of average stock returns.

Multifactor explanations of asset pricing anomalies The main finding of Fama and French (1996) is they propose 3-factor model ($FF3$) to solve most of anomalies *except for momentum*. The existing anomalies are: long-term reversal (DeBondt and Thaler, 1985), short-term momentum (Jegadeesh and Titman, 1993), size, B/M (Lakonishok, Shleifer, and Vishny, 1994), which cannot be explained by CAPM.

Form factors:

- At the end of June in year t, independently sort median ME of NYSE stocks and 3 groups of BE/ME (30-40-30%) of NYSE stocks. We got 6 portfolios.
- Value-weighted monthly returns on portfolios are calculated from July year t to June year $t+1$
- Two factors: $SMB = \frac{SL+SM+SH}{3} - \frac{BL+BM+BH}{3}$; $HML = \frac{SH+BH}{2} - \frac{SL+BL}{2}$

Three-factor models

$$R_i - R_f = a_i + b_i(R_m - R_f) + s_i SMB + h_i HML + e$$

If the three factors can proxy for all risk exposure to unobservable factors, we should see the intercepts are close to zero.

Fama and French often analyze by using two methods:

- Two-way sort
- Run factor model regression of 10 deciles (sort based on the anomalies) then test GRS that all intercepts are zero. If we accept the null, meaning that FF3 could explain the anomaly because all intercepts are zero. For example, if they test B/M, they first sort sample to 10 deciles of B/M then run factor models for 10 portfolios. Then, they test all 10 intercepts are zero by GRS.

A 5-factor model Fama and French (2015) update their model to include two more factors: investment and profitability. They propose a theoretical base that market value equity to book value equity of firms is:

$$\frac{M_t}{B_t} = \frac{\sum_{\tau=1}^{\infty} E(Y_{t+\tau} - dB_{t+\tau})/(1+r)^{\tau}}{B_t}$$

where Y is earning, B is book equity, $dB_{t+\tau} = B_{t+\tau} - B_{t+\tau-1}$ is change in book equity, r is long-term average expected stock return.

Implications of the model is:

- M decreases and B/M increases: so stock returns r increase
- $E(Y_{t+\tau})$ increases (i.e., **profitability factor**) then stock returns r increase
- dB increases (i.e., **investment factor**) then stock returns r decreases

The empirical model is:

$$R_i - R_f = a_i + b_i(R_m - R_f) + s_i SMB + h_i HML + r_i RMW + c_i CMA + e$$

where the new two factors are:

- RMW (robust minus weak profitability, as ROE)
- CMA (conservative minus aggressive investment, as asset growth)

2 What factors that Fama-French model proxy for?

2.1 Innovations of variables relating to *investment opportunity*

It is known that to some extent, SMB and HML could explain some variation in cross-section returns. But Fama and French could not explain what are these factors proxy for. Specifically, FF (1993) suggest that HML and SMB might proxy for state variables that describe *time variation in the investment opportunity set.* Fama and French (1996) argue that the SMB and HML factors of the Fama and French (1993) model proxy for financial distress.

Petkova (2006-JF) propose an idea that these factors are correlated with ***innovations*** *(i.e., changes)* in variables that describe investment opportunities. In other words, the FF model is consistent with an ICAPM explanation.

In literature, some empirical findings show that FF factors relate to macroeconomic variables and business cycle fluctuations (FF factors predict future GDP growth or consumption growth). In this paper, Petkova uses a set of relevant state variables such as

short-term T-bill, term spread, aggregate dividend yield, and default spread. These state variables are chosen to model *two aspects* of the investment opportunity set, namely, ***the yield curve*** and the ***conditional distribution of asset returns***. In this paper, she does not assume that HML and SMB have predictive abilities for the excess market return. Rather, they proxy for innovations (i.e., changes) in variables that possess such ability.

A factor model with excess market return and innovations in the aggregate dividend yield, term spread, default spread, and one-month T-bill yield has a higher explanatory power than the FF three-factor model. FF factors are not significant explanatory variables for the cross-section of average returns in the presence of these innovation factors. Furthermore, this suggestive ICAPM model could be able to account for common timevarying patterns in returns (conditional asset pricing model).

What is ICAPM framework According to **(Merton, 1973)**, if investment opportunities change over time, then assets' exposures to these changes are important determinants of average returns in addition to the market beta. So, Petkova looks at the innovations in state variables that capture uncertainty about investment opportunities in the future.

$$R_{it} = a_i + \beta_{iM} R_{Mt} + \sum(\beta_{i,u^K}) u_t^K + e_{it}$$

where R_i is returns on 25 portfolios sorted by 5×5 size and B/M, R_M is excess market return, u^K is the innovation to state variable K. The innovation is the unexpected component of the variable.

For state variables, she adopts a VAR model that demeanded vector z_t follows a first-order VAR as $z_t = Az_{t-1} + u_t$, where the residual u_t is the innovation terms and is also the risk factor in the model above.

More details about state variables: choose a set of state variables to model two aspects of the investment opportunity set, these are

- To capture *yield curve slope*: short-term T-bill yield (RF) and the term spread (TERM)
- To capture *conditional distribution of asset returns*: aggregate dividend yield (DIV), the default spread (DEF), and interest rates
- She also adds SMB and HML of Fama-French as a comparison. Thus, the FF3 model is also adopted as a competing asset pricing model.

2.2 Default risk

Vassalou and Xing (2004) use Merton (1974) option pricing model to calculate default measure of individual firms. They find that size effect is a *default effect*, and this is also largely true for the book-to-market (BM) effect. The size effect *exists only* within the quintile with the *highest default risk*, which the spread between small and big firms is 45%/year. In the highest default risk quintile, the B/M spread is 30%/year. In low default risk segments, the size effect and B/M effects disappear. Small and value stocks have higher default risk than big and growth stocks and the relation is monotonic.

In addition, they find that high default-risk firms earn higher returns than low default-risk firms, only for small and high BM firms. It is clear that the highest returns are earned by stocks that are either both small in size and high default risk (DLI), or both high DLI and high BM. Finally, they find that default risk is systematic and therefore priced in the cross section of equity returns.

The Fama–French (FF) factors SMB and HML *contain some default-related information*, but this is *not the main reason* that the FF model can explain the cross section of equity returns. SMB and HML appear to *contain important priced information*, unrelated to default risk.

While in literature, default risk on equities focus on default spread (long-term BAA corporate bonds minus long-term AAA or US Treasury bonds) from bond market, they estimate the *default likelihood indicators (DLI)* from equity data. In Merton model, equity of a firm is viewed as a call option on firm asset. Because equity holders are residual claimants on the firm's assets after all other obligations have been met.

If V_A is assets value, X is book value of debt, market value of equity is the call option of V_A with time to expiration equal to T. Apply the Black and Scholes (193) model:

$$V_E = V_A N(d_1) - Xe^{-rT} N(d_2)$$

where r is risk-free rate, N is cdf of the standard normal distribution.

The default probability is the probability that the firm assets will be less than the book value of firm liabilities.

$$P_{def} = Pr\left(V_A \leq X|V_A\right) = Pr(ln(V_A) \leq ln(X_t)|V_A)$$

And distance to default: $DD = \frac{ln(V_A/X)+\left(\mu-\frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$

Default occurs when the *ratio of the value of assets to debt is less than 1*, or its log is negative. The DD tells us by *how many standard deviations* the log of this ratio needs to *deviate from its mean* in order for default to occur.

3 A history of factor model

3.1 A time-line of important studies

Year	*Authors*	*Descriptions*
1972	Black, Jensen, Scholes	Support Black-version CAPM
1973	Fama and Macbeth	Support CAPM
1973	Merton	Intertemporal or ICAPM
1976	Ross	Arbitrage Pricing Theory or APT
1986	Chen, Roll, Ross	Industrial Production, Default spread, and Term Spread
1991	Cochrane	Use aggregate investment's return and relate it expected stock returns
1992	Fama and French	Reject CAPM, support size and BE/ME
1993	Fama and French	3 factor model for equity and 2 factor for bond (Term and Default)
1993	Jegadeesh and Titman	Short term price momentum
1994	Lakonishok, Shleifer, Vishny	Value premium
1996	Fama and French	3 factor model solves size, B/M anomalies, but fail to solve the Momentum
1996	Chan, Jegadeesh, Lakonishok	Earning momentum
1996	Jagannathan and Wang	Conditional CAPM
1997	Carhart	A 4-factor model: 3 factor of FF3 and Momentum
2004	Titman, Wei, Xie	Capex and negative returns
2004	Vassalou and Xing	Default risk as FF3, but not reject FF3
2005	Zhang	Value premium
2006	Petkova	ICAPM idea, reject FF3
2008	Cooper et al.	Asset Growth and negative returns
2010	Wu, Zhang, Zhang	Add investment factor to traditional model to explain accruals
2011	Chen, Novy-Marx, Zhang	Alternative 3-factor q-factor model: Market, I/A, and ROE
2013	Belo et al.	A supply model to compare between actual Tobin's q and predicted q
2015	Hou, Xue, Zhang	A 4-factor q-factor model: Market, Size, I/A, ROE
2015	Fama and French	5-factor model: add RMW and CMA

#Red ink means investment-related models.

3.2 Test of CAPM

The CAPM could be written as $R_i = \alpha_i + \beta_i R_M + e_i$ where R_i is excess return.

In classical papers, first, authors run time-series regression to get $\hat{\beta}_i$ for each individual stock. Second, they rund cross-sectional regression $R_i = \gamma_0 + \gamma_1 \hat{\beta}_i + v_i$ and test if $\hat{\gamma}_0 = 0$ and $\hat{\gamma}_1 = \bar{R}_M$. However, there are several problems with this test:

- Measurement errors
- Error in variables
- Lacking power in tests

Black, Jensen, and Scholes (1972) attempt to solve these problems by: (i) group stocks based on beta, (ii) use instrument variable which is closey related to current beta but can be independently observed (past beta); and (iii) increase number of observation (use monthly return over longer time period).

They propose a time series regression method that:

- First, use past 5 years to run regression and estimate the beta: $R_i = \alpha_i + \beta_i R_M + e_i$. Use this beta to sort stocks to 10 portfolios. Repeat for 35 years.
- Second, run a time series regressions for each of 10 portfolios: $R_i = \alpha_i + \beta_i R_M + e_i$. Then get 10 $\hat{\beta}_i$
- Get the beta from time series then estimate the cross section regression for 10 observations: $\bar{R}_i = \gamma_0 + \gamma_1 \hat{\beta}_1$ then test $\gamma_0 = 0$ and $\gamma_1 = \bar{R}_M$

BJS (1972) support the Black-version of CAPM.

Next, in **Fama and Matbeth (1973)**, they also support the CAPM. Their method is a 3-step method:

- First, use 7 years (formation period) to find beta and sort stocks into 20 beta portfolios
- Second, use next 5 years (estimation period) to estimate $\hat{\beta}_p$ and $\hat{s}_p$
  - Run individual time-series regression: $R_i = \alpha_i + \beta_i R_M + e_i$, take s_i as standard deviation of residuals
  - Then take average of β_i and s_i of each portfolio as $\hat{\beta}_p$ and $\hat{s}_p$
- Third, for next one year (testing period), run cross-sectional regression for each month
  - $R_{pt} = \gamma_0 + \gamma_1 \hat{\beta}_{p,t-1} + \gamma_2 \hat{\beta}^2_{p,t-1} + \gamma_3 \hat{s}_{p,t-1}$

- Repeat for next 3 years, adding one more year in estimation period

- Finally, they have a time-series of γ and they want to test:
  - $\gamma_0 = r_f$
  - $\gamma_1 > 0$: beta is priced
  - $\gamma_2 = 0$: not priced in non-linear in beta and returns
  - $\gamma_3 = 0$: non-beta factor is not priced

The results support CAPM as their predictions in hypotheses, just using a t-test to test.

Fama and French (1992) apply a Fama-Macbeth process to test the CAPM. They find that the beta appears to be related to size. Thus, they control size when create the beta. Their results show that:

- Even when it is the only explanatory variable, beta fails to explain the cross-section returns. This rejects the CAPM.
- The main factors affect cross-section returns are size and BE/ME.
- Other factors such as leverage and E/P lose their explanatory powers after controlling for size and BE/ME

ICAPM and APT: are they the same? This is quite an interesting question. We refer the Chapter 9 of Cochrane (2005) and Chapter 6 of Campbell et al. (1997) to answer this question.

APT is more general than CAPM because it allows for multiple risk factors. APT assumes that the market is competitive and frictionless so return of asset is:

$$\begin{aligned} R_i &= a_i + b_i' f + e_i \\ E[e_i|f] &= 0 \\ E[e_i^2] &= \sigma_i^2 = \Sigma \end{aligned}$$

The factors f account for the common variation in asset returns so that the disturbance term e for large well-diversified portfolios vanishes. Thus, there is a SDF $m = a + b'f$ which linear in f and prices the returns. Chen, Roll, and Ross (1986) use financial theory to add some factors to model: maturity spread, inflation, industrial production, default spread. The results show that some notable factors affect stock returns include: market return, industrial production growth, unanticipated inflation, change in risk premium, yield curve.

ICAPM model is proposed by Merton (1973) that the conditional distribution of returns delivers a multifactor model. The current wealth (market return) is one factor and state variables serve as additional factors. The state variables are factors arise from investors' demand to hedge uncertainty about future investment opportunities. Petkova (2006) again revisits the ICAPM and propose a set of innovations from state variables. She chooses some state variables to explain two properties of the change in growth opportunities: yield curve slope (risk free and term spread) and conditional distribution of the asset returns (aggregate dividend yield, default spread, and interest rate). Using a VAR model: $z_t = Az_{t-1} + u_t$, then u_t is the residuals are used as the innovations of the state variables. Then she propose a model:

$$R_{it} = a_i + \beta_{im} R_{mt} + \sum (\beta_{i,u^k}) u_t^k + e_{it}$$

and use this model to compare with FF3. She proves that her model perform better than the conventional FF3 model.

Although they are quite similar, APT and ICAPM have some distinguished properties:

Properties	***APT***	***ICAPM***
R^2 of regress R^i on f	High	Not need high
f as covariance matrix of return	Yes	No
Factors orthogonal or iid	Yes	No
Inspiration of factors	Analyze the covariance matrix of returns and find portfolios that characterize common factors	From the conditional distribution of future asset returns

The Chen, Roll, Ross (1986) describle itself as an APT but the factors are from financial literature, so it is more likely an ICAPM model. In Fama and French (1993), they propose that SMB and HML proxy for state variables that describe time variation in the investment opportunity set (quite similar to ICAPM). However, Fama and French (1996) construct model from sort returns based on factors, so it is more likely an APT model rather than an ICAPM.

A short review of CAPM and factor model

Classical CAPM

Markowitz (1958) laid a very first groundwork for CAPM. He proposes that investors would optimally hold a mean-variance frontier portfolio: a portfolio gives the highest expected return for a given level of variance. Sharpe (1964) and Lintner (1965) develope this idea that if investors are homogeneous expectations and optimally hold mean-variance frontier portfolios, then the market portfolio will be itself a mean-variance portfolio. The Sharpe-Lintner CAPM model is:

$$\begin{aligned} E[R_i] &= R_f + \beta_{im}\left(E[R_m] - R_f\right) \\ \beta_{im} &= cov\left(R_i, R_m\right)/var(R_m) \end{aligned}$$

where the R_m is return on market portfolio, R_f is risk-free rate. The model could be written in excess return too:

$$E[R_i - R_f] = \beta_{im}\left(E[R_m] - R_f\right)$$

Black (1972) propose a more general Black-version CAPM:

$$E[R_i] = R_{0m} + \beta_{im}\left(E[R_m] - E[R_{0m}]\right)$$

where R_{0m} is return on zero-beta portfolio associated with m.

Early empirical evidence

Early evidence such as Black, Jensen, Scholes (1972) and Fama and Macbeth (1973) support the CAPM. Using time-series regression approach for 10 beta-sort portfolios, Black, Jensen, Scholes (1972) support the CAPM. Fama and Macbeth (1973) use a cross-section method to test the CAPM. Their method includes two main steps:

1. For each t in T periods, assuming beta is known, run a cross-section regression: $E[R_i - R_f] = \gamma_{0t} + \gamma_{1t}\beta_{im}$

2. Given the time series of estimates γ_{0t} and γ_{1t}, use an usual t-test that $\gamma_0 = 0$ and $\gamma_1 > 0$

$$t - statistic = \frac{\bar{\hat{\gamma}}_j}{s(\hat{\gamma}_j)}$$

Again, their evidence supports the CAPM.

Anomalies

However, CAPM fails to explain stock returns of some portfolio formed from specific factors. First, Fama and French (1992) find that beta tends to be related to size, small stocks earn higher returns than large stocks. Thus, they control for size when forming the beta. Then, applying the Fama-Macbeth procedure, they find that beta cound not explain the cross-section returns, even when it is the only explanatory variable. However, size and ratio of book equity to market equity (B/M) have significant power to exlain stock returns. Other factors (such as leverage, earning E/P) fail to explain stock returns, after controlling size and B/M.

Debondt and Thaler (1985) find a long-term reversal in stock returns, while Jegadeesh and Titman (1993) find a short-term momentum. Momentum effect refers to a portfolio formed by buying past winners and selling past losers has a higher average return than the CAPM predicts. Lakonishok, Shleifer, and Vishny (1994) find the value premium that value stocks outperform the growth stocks. Value stocks refer to stock with low price relative to fundamental value such as high B/M, E/P, C/P (cash), and low SG (sale growth). Thus, a contrarian strategy that sells growth stocks and buys value stocks could earn positive returns.

Factor models

Fama and French (1993) show 5 common risk factors to explain equity return and bond return. For stock returns, they propose three factors: market excess return, size, and B/M. For government and corporate bond return, they propose two additional factors for maturity spread (TERM) and default spread (DEF). Using a time-series approach as in Black, Jensen, Scholes (1972), they find that these mimicking factors formed based on these factors can capture stock return and bond return.

Fama and French (1996) again use the 3 factor model (hereafter FF3) includes market excess return, size, and B/M to explain anomalies in literature.

$$R_i - R_f = a_i + b_i(R_m - R_f) + s_i SMB + h_i HML$$

where SMB is small-minus-big mimicking portfolio return, HML is high-minus-low.

The idea is if the FF3 can explain the portfolio return of anomalies, the intercept will be close to zero. The FF3 successfully explain size, B/M, and long-term reversal anomalies, but fail to explain the momentum. Thus, Carhart (1997) add momentum and construct a 4-factor Carhart model:

$$R_i - R_f = a_i + b_i(R_m - R_f) + s_i SMB + h_i HML + u_i UMD$$

where UMD is winner-minus-loser.

Conditional CAPM

While traditional CAPM assumes a static beta, Jagannathan and Wang (1996) relax two assumptions: (i) allow beta and market premium depend on a set of information at specific time point and vary over time and (ii) add an additional human capital return together with market return. Their model is as follows:

$$E[R_i] = c_0 + c_{size} ln(ME) + c_{vw}\beta^{vw} + c_{prem}\beta^{prem} + c_{labor}\beta^{labor}$$

where $\beta^{vw} = cov(R_i, R_{vw})/var(R_{vw})$ and R_{vw} is value-weighted market return;

$\beta^{prem} = cov(R_i, R_{prem})/var(R_{prem})$ and R_{prem} is defaul spread between BAA and AAA rated bonds;

$\beta^{labor} = cov(R_i, R_{labor})/var(R_{labor})$ and R_{labor} is growth rate in per capital labor income.

Next question is which factors that FF3 proxies for? This question is quite interesting because in Fama and French (1993), they propose that SMB and HML proxy for state variables that describe time variation in the investment opportunity set (quite similar to ICAPM). Fama and French (1996) argue that SMB and HML might proxy for financial distress.

APT and ICAPM

Ross (1976) propose the APT model that its factors f account for the common variation in asset returns so that the disturbance term e for large well-diversified portfolios vanishes. Thus, there is a SDF $m = a + b'f$ which linear in f and prices the returns. Chen, Roll, and Ross (1986) use financial theory to add some factors to model: maturity spread, inflation, industrial production, default spread. The results show that some notable factors affect stock returns include: market return, industrial production growth, unanticipated inflation, change in risk premium, yield curve.

Vassalou and Xing (2004) use default risk to check whether FF3's factors proxy for default risk. They find that size effect and B/M effect exist only in subsample of stocks with high default risk. In addition, the stock with high default risk earn higher stock return than those with low default risk only in subsample of small stock or high B/M ratio. They conclude that FF3 contain some default-related information but they don't reject the FF3. They argue that FF3 contain other important priced information that unrelated to default risk.

Merton (1973) proposes the ICAPM model that is a multifactor model that includes current wealth (market return) and other factors explain the future investment opportunities. Petkova (2006) applies this idea to extract the innovations u^K from state variables that describible the yield curve (risk free and TERM) and conditional distribution of asset returns (interest rate, aggregate dividend yield DIV, and default spread DEF). Then she propose an ICAPM:

$$R_{it} = a_i + \beta_{im} R_{mt} + \sum (\beta_{i,u^k}) u_t^k + e_{it}$$

and use this model to compare with FF3. She proves that her model perform better than the conventional FF3 model.

A recent 5-factor model

As an attempt to fix their FF3 that fails to explain a bunch of anomalies, Fama and French (2015) finally add two additional factor profitability and investment to their old FF3.

$$R_i - R_f = a_i + b_i(R_m - R_f) + s_i SMB + h_i HML + r_i RMW + c_i CMA$$

where RMW is robust-minus-weak (profitability) and CMA is conservative-minus-aggressive (investment) factors.

References

Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993. ISSN 0304-405X. doi: 10.1016/0304-405X(93)90023-5. URL `http://www.sciencedirect.com/science/article/pii/0304405X93900235`.

Eugene F Fama and Kenneth R French. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, 51(1):55–84, 1996.

Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, April 2015. ISSN 0304-405X. doi: 10.1016/j.jfineco.2014.10.010. URL `http://www.sciencedirect.com/science/article/pii/S0304405X14002323`.