

NBER WORKING PAPER SERIES

THE INVESTMENT CAPM

Lu Zhang

Working Paper 23226

<http://www.nber.org/papers/w23226>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

March 2017

I thank Hang Bai, Zhengyu Cao, Alex Edmans, Andrei Goncalves, Kewei Hou, Stephen Penman, Andreas Stathopoulos, and especially Chen Xue for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Lu Zhang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Investment CAPM  
Lu Zhang  
NBER Working Paper No. 23226  
March 2017  
JEL No. E13,E22,G12,G14,G31

**ABSTRACT**

A new class of Capital Asset Pricing Models (CAPM) arises from the first principle of real investment for individual firms. Conceptually as "causal" as the consumption CAPM, yet empirically more tractable, the investment CAPM emerges as a leading asset pricing paradigm. Firms do a good job in aligning investment policies with costs of capital, and this alignment drives many empirical patterns that are anomalous in the consumption CAPM. Most important, integrating the anomalies literature in finance and accounting with neoclassical economics, the investment CAPM succeeds in mounting an efficient markets counterrevolution to behavioral finance in the past 15 years.

Lu Zhang  
Fisher College of Business  
The Ohio State University  
2100 Neil Avenue  
Columbus, OH 43210  
and NBER  
zhanglu@fisher.osu.edu

# 1 Introduction

Consider a two-period stochastic general equilibrium model. The economy has three defining characteristics of neoclassical economics: (i) Agents have rational expectations; (ii) consumers maximize utility, and firms maximize their market value of equity; and (iii) markets clear.

There are two dates,  $t$  and  $t + 1$ . The economy is populated by a representative household and heterogeneous firms, indexed by  $i = 1, 2, \dots, N$ . The representative household maximizes its expected utility,  $U(C_t) + \rho E_t[U(C_{t+1})]$ , in which  $\rho$  is the time preference coefficient, and  $C_t$  and  $C_{t+1}$  are consumption expenditures in  $t$  and  $t + 1$ , respectively. Let  $P_{it}$  be the ex-dividend equity, and  $D_{it}$  the dividend of firm  $i$  at period  $t$ . The first principle of consumption says that:

$$P_{it} = E_t[M_{t+1}(P_{it+1} + D_{it+1})] \quad \Rightarrow \quad E_t[M_{t+1}r_{it+1}^S] = 1, \quad (1)$$

in which  $r_{it+1}^S \equiv (P_{it+1} + D_{it+1})/P_{it}$  is firm  $i$ 's stock return, and  $M_{t+1} \equiv \rho U'(C_{t+1})/U'(C_t)$  is the household's stochastic discount factor. Equation (1) can be rewritten as:

$$E_t[r_{it+1}^S] - r_{ft} = \beta_{it}^M \lambda_{Mt}, \quad (2)$$

in which  $r_{ft} \equiv 1/E_t[M_{t+1}]$  is the real interest rate,  $\beta_{it}^M \equiv -\text{Cov}(r_{it+1}^S, M_{t+1})/\text{Var}(M_{t+1})$  is the consumption beta, and  $\lambda_{Mt} \equiv \text{Var}(M_{t+1})/E_t[M_{t+1}]$  is the price of the consumption risk. Equation (2) is the consumption CAPM, first derived by Rubinstein (1976), Lucas (1978), and Breeden (1979). The classic CAPM, due to Treynor (1962), Sharpe (1964), Lintner (1965), and Mossin (1966), can be obtained as a special case of the consumption CAPM under quadratic utility or exponential utility with normally distributed returns (Cochrane 2005, p. 152–155).

On the production side, firms produce a single commodity to be consumed or invested. Firm  $i$  starts with the productive capital,  $K_{it}$ , operates in both dates, and exits at the end of date  $t + 1$  with a liquidation value of zero. The rate of capital depreciation is set to be 100%. Firms differ in capital,  $K_{it}$ , and profitability,  $X_{it}$ , both of which are known at the beginning of date  $t$ . The operating profits are given by  $\Pi_{it} \equiv X_{it}K_{it}$ . Firm  $i$ 's profitability at date  $t + 1$ ,  $X_{it+1}$ , is stochastic, and is subject to aggregate shocks affecting all firms simultaneously, as well as firm-specific shocks affecting only firm  $i$ . Let  $I_{it}$  be the investment for date  $t$ , then  $K_{it+1} = I_{it}$ . Investment entails quadratic adjustment costs,  $(a/2)(I_{it}/K_{it})^2 K_{it}$ , in which  $a > 0$  is a constant parameter.

Firm  $i$  uses its operating profits at date  $t$  to pay investment and adjustment costs. If the net

cash flow,  $D_{it} \equiv X_{it}K_{it} - I_{it} - (a/2)(I_{it}/K_{it})^2K_{it}$ , is positive, the firm distributes it back to the household. A negative  $D_{it}$  means external equity raised by the firm from the household. At date  $t+1$ , firm  $i$  uses capital,  $K_{it+1}$ , to obtain operating profits,  $X_{it+1}K_{it+1}$ , which is in turn distributed as dividends back to the household,  $D_{it+1} \equiv X_{it+1}K_{it+1}$ . With only two dates, firm  $i$  does not invest in date  $t+1$ ,  $I_{it+1} = 0$ , and the ex-dividend equity value,  $P_{it+1}$ , is zero.

Taking the household's stochastic discount factor,  $M_{t+1}$ , as given, firm  $i$  chooses  $I_{it}$  to maximize the cum-dividend equity value at the beginning of date  $t$ :

$$P_{it} + D_{it} = \max_{\{I_{it}\}} \left[ X_{it}K_{it} - I_{it} - \frac{a}{2} \left( \frac{I_{it}}{K_{it}} \right)^2 K_{it} + E_t[M_{t+1}X_{it+1}K_{it+1}] \right]. \quad (3)$$

The first principle of investment for firm  $i$  says that:

$$1 + a \frac{I_{it}}{K_{it}} = E_t[M_{t+1}X_{it+1}]. \quad (4)$$

Intuitively, the marginal costs of investment, consisting of the purchasing price (unity) and the marginal adjustment costs,  $a(I_{it}/K_{it})$ , must equal marginal  $q$ , which is the present value of the marginal benefits of investment given by the marginal product of capital,  $X_{it+1}$ .

The first principle of investment can be rewritten without the stochastic discount factor,  $M_{t+1}$  (Cochrane 1991). In particular, equation (3), when combined with  $D_{it} \equiv X_{it}K_{it} - I_{it} - (a/2)(I_{it}/K_{it})^2K_{it}$ , implies that the ex-dividend equity value at the optimum is:

$$P_{it} = E_t[M_{t+1}X_{it+1}K_{it+1}]. \quad (5)$$

The stock return can then be rewritten as:

$$r_{it+1}^S = \frac{P_{it+1} + D_{it+1}}{P_{it}} = \frac{X_{it+1}K_{it+1}}{E_t[M_{t+1}X_{it+1}K_{it+1}]} = \frac{X_{it+1}}{E_t[M_{t+1}X_{it+1}]}. \quad (6)$$

Combining equations (4) and (6) yields *the investment CAPM*:

$$r_{it+1}^S = \frac{X_{it+1}}{1 + a(I_{it}/K_{it})}. \quad (7)$$

Intuitively, firm  $i$  keeps investing until the date- $t$  marginal costs of investment,  $1 + a(I_{it}/K_{it})$ , equal the marginal benefits of investment at  $t+1$ ,  $X_{it+1}$ , discounted to date  $t$  with the stock return,  $r_{it+1}^S$ , as the discount rate. Equivalently, the ratio of the marginal benefits of investment at  $t+1$  divided by the marginal costs of investment at  $t$  equals the discount rate,  $r_{it+1}^S$ .

Most important, the investment CAPM, as asset pricing theory, gives rise to *cross-sectionally varying expected returns*. The model predicts that, all else equal, high investment stocks should earn lower expected returns than low investment stocks, and that stocks with high expected profitability should earn higher expected returns than stocks with low expected profitability. When expected returns vary cross-sectionally in equilibrium, stock prices will adjust in a way that connects expected returns to characteristics. In particular, stock prices will not adjust to conform to a cross-sectionally constant discount rate, meaning that characteristics do not predict returns. A cross-sectionally constant discount rate is equivalent to saying that all stocks are equally risky.

The intuition behind the investment CAPM is just the net present value rule in capital budgeting, which is a fundamental principle in modern corporate finance. Investment predicts returns because given expected profitability, high costs of capital imply low net present values of new capital and low investment, and low costs of capital imply high net present values of new capital and high investment. Profitability predicts returns because high expected profitability relative to low investment must imply high discount rates. The high discount rates are necessary to offset the high expected profitability to induce low net present values of new capital and low investment. If the discount rates were not high enough, firms would instead observe high net present values of new capital and invest more. Conversely, low expected profitability relative to high investment must imply low discount rates. If the discount rates were not low enough to counteract the low expected profitability, firms would instead observe low net present values of new capital and invest less.

The consumption CAPM and the investment CAPM are the two sides of the same coin in general equilibrium, delivering identical expected returns (Lin and Zhang 2013). To see this insight, combining the consumption CAPM in equation (2) and the investment CAPM in equation (7) yields:

$$r_{ft} + \beta_{it}^M \lambda_{Mt} = E_t[r_{it+1}^S] = \frac{E_t[X_{it+1}]}{1 + a(I_{it}/K_{it})}. \quad (8)$$

Intuitively, the consumption CAPM, which is derived from the first principle of consumption, connects expected returns to consumption betas. The consumption CAPM predicts that consumption betas are sufficient statistics for expected returns. Once consumption betas are controlled for, characteristics should not affect the cross section of expected returns. This prediction is the organizing framework for the bulk of empirical finance and capital markets research in accounting.

The investment CAPM, which is derived from the first principle of investment, connects expected returns to characteristics. It predicts that characteristics are sufficient statistics for expected

returns. Once characteristics are controlled for, consumption betas should not affect expected returns. As such, the consumption CAPM and the classic CAPM as its special case miss what neoclassical economics has to say about the cross section of expected returns from the investment side altogether. Derived from the two-period manager’s problem, the investment CAPM is the dual proposition to the classic CAPM, which is in turn derived from a two-period investor’s problem.

This essay aims at three objectives. First, I review the investment CAPM literature within a unified conceptual framework, identify its main strands, and clarify their interconnections. Second, I explain the big picture. I trace the origin of the investment CAPM to at least Fisher’s (1930) classic *The Theory of Interest*. I describe the reincarnation of the investment CAPM as an inevitable response to the anomalies literature and its challenge to efficient markets (Fama 1965, 1970) and rational expectations (Muth 1961; Lucas 1972). Third, to the extent that the investment CAPM thinks about asset pricing issues very differently from the traditional consumption CAPM, I hope to make the related literature accessible to doctoral students, economists who do not specialize in asset pricing, as well as investment management professionals. While mostly a review, many new insights and results are also furnished. Finally, due to space limit, this essay reviews mostly the empirical literature. A companion essay, Zhang (2017), reviews the related quantitative theories.

The rest of the paper unfolds as follows. Section 2 shows how the investment CAPM is tested via a factor model. Section 3 shows the structural estimation and tests of the multiperiod investment CAPM. Section 4 explains the big picture of the investment CAPM, and clarifies its relations with the consumption CAPM and behavioral finance. Finally, Section 5 discusses open questions.

## 2 The $q$ -factor Model

Hou, Xue, and Zhang (2015) implement the investment CAPM with Black, Jensen, and Scholes’s (1972) portfolio approach. In their  $q$ -factor model, the expected return of an asset in excess of the risk-free rate, denoted  $E[R_i - R_f]$ , is described by the sensitivities of its returns to the market factor, a size factor, an investment factor, and a profitability (return on equity, ROE) factor:

$$E[R_i - R_f] = \beta_{\text{MKT}}^i E[\text{MKT}] + \beta_{\text{ME}}^i E[r_{\text{ME}}] + \beta_{\text{I/A}}^i E[r_{\text{I/A}}] + \beta_{\text{ROE}}^i E[r_{\text{ROE}}], \quad (9)$$

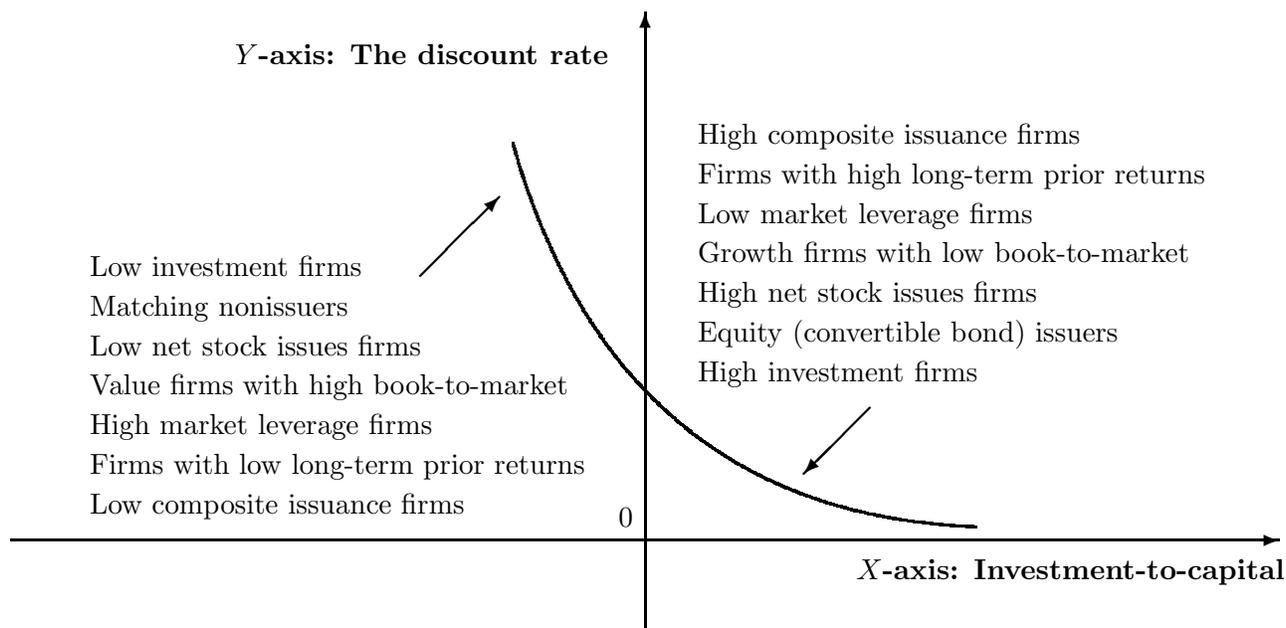
in which  $E[R_i - R_f]$  is the expected excess return,  $E[\text{MKT}]$ ,  $E[r_{\text{ME}}]$ ,  $E[r_{\text{I/A}}]$ , and  $E[r_{\text{ROE}}]$  are the expected factor premiums, and  $\beta_{\text{MKT}}^i$ ,  $\beta_{\text{ME}}^i$ ,  $\beta_{\text{I/A}}^i$ , and  $\beta_{\text{ROE}}^i$  are the factor loadings, respectively.

Strictly speaking, the investment CAPM in equation (7) only motivates the investment and ROE factors. Hou, Xue, and Zhang (2015) include the size factor to make the  $q$ -factor model more palatable for evaluating mutual funds, in which size is a popular investment style. However, Hou et al. show that while the size factor helps the  $q$ -factor model fit the average returns across the size deciles, its incremental role in a broad set of anomaly deciles is minimal. Finally, equation (7) is primarily a cross-sectional model, and the size, investment, and ROE factors are all zero-investment portfolios. As such, the market factor is also included to anchor the time-series average of expected returns, while the task of fitting the cross-sectional variation of expected returns is left to the  $q$ -factors.

## 2.1 Intuition

Figure 1 shows that many cross-sectional patterns, such as equity issuance, accruals, valuation ratios, and long-term reversal, are likely different manifestations of the investment premium.

**Figure 1. The Investment Premium**



The negative relation between average returns and equity issuance (Ritter 1991; Loughran and Ritter 1995) is consistent with the investment premium. The flow of funds constraint of firms implies that a firm's uses of funds must equal its sources of funds. As such, all else equal, issuers must invest more, and earn lower average returns than nonissuers (Lyandres, Sun, and Zhang 2008).

In addition, total asset growth predicts returns with a negative slope (Cooper, Gulen, and Schill 2008). However, asset growth is the most comprehensive measure of investment-to-capital, in which capital is interpreted as all productive assets, and investment is the change in total assets. As such, asset growth is the premier manifestation of the investment effect.

The value premium (Rosenberg, Reid, and Lanstein 1985) is also consistent with the investment premium. Combining equations (4) and (5) implies that the marginal costs of investment,  $1+a(I_{it}/K_{it})$ , equal market equity-to-capital,  $P_{it}/K_{it+1}$ . Without debt,  $P_{it}/K_{it+1}$  equals market-to-book equity. Intuitively, value firms with low  $P_{it}/K_{it+1}$  should invest less, and earn higher expected returns than growth firms with high  $P_{it}/K_{it+1}$ . More generally, firms with high valuation ratios have more growth opportunities, invest more, and earn lower expected returns than firms with low valuation ratios. The investment premium also manifests as long-term reversal (De Bondt and Thaler 1985). High valuation ratios of growth firms are often associated with a stream of positive shocks to prior stock returns, and low valuation ratios of value firms with a stream of negative shocks to prior stock returns. As such, firms with high long-term prior returns tend to be growth firms that invest more, and earn lower expected returns than firms with low long-term prior returns.

In addition to the investment premium, equation (7) also gives rise to the profitability premium. Controlling for investment, firms with high expected profitability should earn higher expected returns than firms with low expected profitability. For any portfolio sorts that produce wider cross-sectional expected return spreads associated with expected profitability than those with investment, their average returns can be interpreted with the common implied sort on expected profitability.

Earnings momentum winners that have recently experienced positive shocks to profitability tend to be more profitable, with higher expected profitability, than earnings momentum losers that have recently experienced negative shocks to profitability. The profitability effect then implies that earnings momentum winners should earn higher expected returns than earnings momentum losers. In addition, shocks to earnings are positively correlated with stock returns contemporaneously. Intuitively, firms with positive earnings shocks tend to experience immediate stock price increases, whereas firms with negative earnings shocks tend to experience immediate stock price decreases. As such, the profitability effect is also consistent with price momentum, i.e., stocks that have performed well recently continue to earn higher average returns in the subsequent six months than stocks that have performed poorly recently. Finally, firms that are more financially distressed are less profitable, and all else equal, should earn lower expected returns than firms that are less

financially distressed. As such, the profitability effect is also consistent with the distress anomaly.

The anomaly variables described so far are directly related to investment and profitability. Hou, Xue, and Zhang (2015) test the  $q$ -factor model with substantially more anomalies. As noted, the consumption CAPM and the investment CAPM are theoretically equivalent in equilibrium, delivering identical expected returns. In particular, the investment CAPM says that controlling for a few characteristics is sufficient to explain the cross section of expected returns. Hou et al. take this prediction seriously, and test the  $q$ -factor model with an exhaustive array of anomaly variables, including those that are not directly related to investment and profitability.

## 2.2 Evidence

Hou, Xue, and Zhang (2015) measure investment-to-capital,  $I/A$ , as the annual change in total assets divided by one-year-lagged total assets, and profitability, ROE, as quarterly earnings (income before extraordinary items) divided by one-quarter-lagged book equity. The  $q$ -factors are constructed from a triple  $2 \times 3 \times 3$  sort on size,  $I/A$ , and ROE. Hou et al. examine nearly 80 anomaly variables that cover all the major categories of anomalies, including momentum, value-versus-growth, investment, profitability, intangibles, and trading frictions. To form testing deciles, NYSE breakpoints and value-weighted decile returns are used to alleviate the impact of microcaps.

The  $q$ -factor model outperforms the Fama-French three-factor and Carhart models. Across the 35 significant high-minus-low deciles, the average magnitude of the  $q$ -model alphas is 0.2% per month, which is lower than 0.33% in the Carhart model and 0.55% in the three-factor model. Only five out of 35 high-minus-low deciles have significant  $q$ -model alphas at the 5% level. In contrast, 19 high-minus-low deciles have significant Carhart alphas, and 27 have significant Fama-French alphas. The  $q$ -factor model also has the lowest mean absolute value of alphas across all 35 sets of deciles, 0.11% per month, which is lower than 0.12% in the Carhart model and 0.16% in the three-factor model. The Gibbons, Ross, and Shanken (GRS, 1989) test on the null that the alphas are jointly zero across a given set of deciles rejects the  $q$ -factor model at the 5% level in 20 out of 35 sets of deciles, but the Carhart model in 24, and the three-factor model in 28 sets of deciles. Across the nine momentum anomalies, the average magnitude of the high-minus-low alphas is 0.19% per month in the  $q$ -factor model, 0.29% in the Carhart model, and 0.85% in the three-factor model.

Consistent with the broad implications of the investment CAPM, most anomalies turn out to be different manifestations of the investment and profitability effects. Price and earnings momen-

tum are mainly connected to the ROE factor. For the high-minus-low momentum deciles, all the ROE factor loadings are significant, but the investment factor loadings are mostly insignificant. The investment factor is mainly responsible for capturing the value-minus-growth anomalies. Their investment factor loadings are all more than five standard errors from zero, but the ROE factor loadings are mostly insignificant. The investment factor is mainly responsible for the investment anomalies, the ROE factor for the profitability anomalies, and a combination of the two factors accounts for the anomalies in the intangibles and trading frictions categories.

## 2.3 Debate

The  $q$ -factor model poses to end the almost quarter-century reign of the Fama-French three-factor model, along with its extension, the Carhart four-factor model, as the workhorse model for empirical asset pricing. Perhaps not surprisingly, the  $q$ -factor model touches off a firestorm of controversy.<sup>1</sup>

### 2.3.1 “Endorsement” from Fama and French (2015)

Subsequent to our work, Fama and French (2015) incorporate their own versions of the investment and profitability factors into their original three-factor model to form a five-factor model:<sup>2</sup>

$$E[R_i - R_f] = b_i E[\text{MKT}] + s_i E[\text{SMB}] + h_i E[\text{HML}] + r_i E[\text{RMW}] + c_i E[\text{CMA}]. \quad (10)$$

MKT, SMB, and HML are the market, size, and value factors that first appear in their three-factor model. The two new factors resemble the  $q$ -factors. RMW (robust-minus-weak) is the difference between the returns on diversified portfolios of stocks with robust and weak operating profitability. CMA (conservative-minus-aggressive) is the difference between the returns on diversified portfolios of low and high investment stocks. Operating profitability, OP, is revenues minus costs of goods sold, minus selling, general, and administrative expenses, minus interest expenses, all scaled by book equity (Novy-Marx 2013). Investment, Inv, is the change in book assets divided by one-year-lagged book assets (same as in Hou, Xue, and Zhang 2015). The factors are constructed from independent  $2 \times 3$  sorts by interacting size with, separately, book-to-market, OP, and Inv.

---

<sup>1</sup>Although first appearing in October 2012 as NBER working paper 18435, the Hou-Xue-Zhang (2015) article is a new incarnation of the previous work circulated as “Neoclassical factors” (as NBER working paper 13282, July 2007), “An equilibrium three-factor model” (January 2009), “Production-based factors” (April 2009), “A better three-factor model that explains more anomalies” (June 2009), and “An alternative three-factor model” (April 2010). The frequent title changes make it clear that the June 2009 title was fought against vigorously, albeit unsuccessfully. The insight that investment and profitability are fundamental forces in the cross section of expected stock returns in the investment CAPM first appears in Zhang (2005a, NBER working paper 11322, May 2005).

<sup>2</sup>Their June 2013 draft, Fama and French (2013), adds only a profitability factor into their three-factor model, and subsequent drafts, starting from the November 2013 draft, also add an investment factor.

Fama and French (2015) motivate their five-factor model from the Miller and Modigliani (1961) valuation theory. From the dividend discounting model, the market value of firm  $i$ 's stock,  $P_{it}$ , is the present value of its expected dividends,  $P_{it} = \sum_{\tau=1}^{\infty} E[D_{it+\tau}]/(1+r_i)^\tau$ , in which  $D_{it}$  is dividends, and  $r_i$  is the firm's long-term average expected stock return, i.e., the internal rate of return. The clean surplus relation says that dividends equal earnings minus the change in book equity,  $D_{it+\tau} = \Pi_{it+\tau} - \Delta B_{it+\tau}$ , in which  $\Delta B_{it+\tau} \equiv B_{it+\tau} - B_{it+\tau-1}$ . It follows that:

$$\frac{P_{it}}{B_{it}} = \frac{\sum_{\tau=1}^{\infty} E[\Pi_{it+\tau} - \Delta B_{it+\tau}]/(1+r_i)^\tau}{B_{it}}. \quad (11)$$

Fama and French (2015) argue that the valuation equation (11) makes three theoretical predictions. First, fixing everything except the current market value,  $P_{it}$ , and the expected stock return,  $r_i$ , a low  $P_{it}$ , or a high book-to-market equity,  $B_{it}/P_{it}$ , implies a high expected return. Second, fixing everything except the expected profitability and the expected stock return, high expected profitability implies a high expected return. Third, fixing everything except the expected growth in book equity and the expected return, high expected book equity growth implies a low expected return.

Fama and French (2015) use operating profitability as the proxy for the expected profitability and past investment as the proxy for the expected investment. Empirically, the five-factor model outperforms the original three-factor model in pricing testing portfolios formed on size, book-to-market, investment, and profitability. However, these portfolios are formed on the same variables underlying the factors. Fama and French (2016) extend their set of testing portfolios to include accruals, net share issues, momentum, volatility, and the market beta, which are still a small subset of the comprehensive universe of anomalies examined in Hou, Xue, and Zhang (2015).

### 2.3.2 Which Factor Model?

Hou, Xue, and Zhang (2016) compare the  $q$ -factor model with the Fama-French five-factor model on empirical grounds. To make the sample comparable to Fama and French (2015), who start their sample in July 1963, Hou et al. extend the sample for the  $q$ -factors back to January 1967.

Table 1 shows that, from January 1967 to December 2014, the size, investment, and ROE factors earn on average 0.32%, 0.43%, and 0.56% per month ( $t = 2.42, 5.08,$  and  $5.24$ ), respectively. SMB, HML, RMW, and CMA earn on average 0.26%, 0.36%, 0.27%, and 0.34% ( $t = 1.92, 2.57, 2.58,$  and  $3.63$ ), respectively. The five-factor model cannot explain the investment and ROE  $q$ -factors, leaving alphas of 0.12% ( $t = 3.35$ ) and 0.45% ( $t = 5.6$ ), respectively. However, the  $q$ -factor model explains

the HML, RMW, and CMA returns, with tiny alphas of 0.03%, 0.04%, and 0.01%, respectively ( $t$ -statistics all below 0.5). As such, RMW and CMA are noisy versions of the  $q$ -factors.

Hou, Mo, Xue, and Zhang (2017) raise four concerns on Fama and French's (2015) motivation of the five-factor model. First, Fama and French derive the relations between book-to-market, investment, and profitability only with the internal rate of return, but argue that the difference between the one-period-ahead expected return and the internal rate of return is not important. However, Hou et al. estimate the internal rate of returns for RMW and CMA with accounting-based models (Gebhardt, Lee, and Swaminathan 2001), and show that these estimates deviate greatly from their one-month-ahead average returns. Whereas the average returns of RMW are significantly positive, its estimates of the internal rate of returns are often significantly negative.

Second, Fama and French (2015) argue that the value factor is a separate factor in valuation theory, but find it to be redundant in describing average returns in the data. However, the investment CAPM implies that the value premium is just a different manifestation of the investment effect. The first principle of investment says that the marginal costs of investment, which rise with investment, equal marginal  $q$ , which is in turn closely related to market-to-book equity. As such, the value factor is redundant in the presence of the investment factor in the investment CAPM. Without the redundant HML, the five-factor model collapses to a noisy version of the  $q$ -factor model.

Third, Fama and French (2015) motivate their investment factor, CMA, from the negative relation between the expected investment and the internal rate of returns in valuation theory. However, the valuation equation can be reformulated to show that the relation between the one-period-ahead expected return and the expected investment is more likely to be positive. Combining the definition of return,  $P_{it} = (E_t[D_{it+1}] + E_t[P_{it+1}])/(1 + E_t[r_{it+1}])$ , with the clean surplus relation yields:

$$P_{it} = \frac{E_t[\Pi_{it+1} - \Delta B_{it+1}] + E_t[P_{it+1}]}{1 + E_t[r_{it+1}]} \quad (12)$$

Dividing both sides by  $B_{it}$  and rearranging yield:

$$\frac{P_{it}}{B_{it}} = \frac{E_t \left[ \frac{\Pi_{it+1}}{B_{it}} \right] - E_t \left[ \frac{\Delta B_{it+1}}{B_{it}} \right] + E_t \left[ \frac{P_{it+1}}{B_{it+1}} \left( 1 + \frac{\Delta B_{it+1}}{B_{it}} \right) \right]}{1 + E_t[r_{it+1}]}, \quad (13)$$

$$\frac{P_{it}}{B_{it}} = \frac{E_t \left[ \frac{\Pi_{it+1}}{B_{it}} \right] + E_t \left[ \frac{\Delta B_{it+1}}{B_{it}} \left( \frac{P_{it+1}}{B_{it+1}} - 1 \right) \right] + E_t \left[ \frac{P_{it+1}}{B_{it+1}} \right]}{1 + E_t[r_{it+1}]} \quad (14)$$

Fixing everything except  $E_t[\Delta B_{it+1}/B_{it}]$  and  $E_t[r_{it+1}]$ , high  $E_t[\Delta B_{it+1}/B_{it}]$  is likely to be associ-

**Table 1 : Factor Spanning Tests, January 1967 to December 2014**

$r_{ME}$ ,  $r_{I/A}$ , and  $r_{ROE}$  are the size, investment, ROE factors in the  $q$ -factor model, respectively. SMB, HML, RMW, and CMA are the size, value, profitability, and investment factors from the Fama-French five-factor model, respectively. The data for SMB and HML in the three-factor model, SMB, HML, RMW, and CMA in the five-factor model, as well as UMD are from Kenneth French's Web site.  $m$  is the average return,  $\alpha_C$  is the Carhart alpha,  $\alpha_q$  the  $q$ -factor alpha,  $a$  is the five-factor alpha, and  $b, s, h, r$ , and  $c$  are five-factor loadings. The  $t$ -statistics in parentheses are adjusted for heteroscedasticity and autocorrelations.

Panel A: The Hou-Xue-Zhang $q$ -factors							
	$m$	$\alpha_C$	$\beta_{MKT}$	$\beta_{SMB}$	$\beta_{HML}$	$\beta_{UMD}$	$R^2$
$r_{ME}$	0.32 (2.42)	0.01 (0.25)	0.01 (1.08)	0.97 (67.08)	0.17 (7.21)	0.03 (1.87)	0.94
$r_{I/A}$	0.43 (5.08)	0.29 (4.57)	-0.06 (-4.51)	-0.04 (-1.88)	0.41 (13.36)	0.05 (1.93)	0.52
$r_{ROE}$	0.56 (5.24)	0.51 (5.58)	-0.04 (-1.39)	-0.30 (-4.31)	-0.12 (-1.79)	0.27 (6.19)	0.40
	$a$	$b$	$s$	$h$	$r$	$c$	$R^2$
$r_{ME}$	0.05 (1.39)	0.00 (0.39)	0.98 (68.34)	0.02 (1.14)	-0.01 (-0.21)	0.04 (1.19)	0.95
$r_{I/A}$	0.12 (3.35)	0.01 (0.73)	-0.05 (-2.86)	0.04 (1.60)	0.07 (2.77)	0.82 (26.52)	0.84
$r_{ROE}$	0.45 (5.60)	-0.04 (-1.45)	-0.11 (-2.69)	-0.24 (-3.54)	0.75 (13.46)	0.13 (1.34)	0.52
Panel B: The Fama-French five factors							
	$m$	$\alpha_C$	$\beta_{MKT}$	$\beta_{SMB}$	$\beta_{HML}$	$\beta_{UMD}$	$R^2$
SMB	0.26 (1.92)	-0.02 (-1.24)	0.00 (0.96)	1.00 (89.87)	0.13 (8.07)	0.00 (0.11)	0.99
HML	0.36 (2.57)	-0.00 (-1.79)	0.00 (1.79)	-0.00 (-1.69)	1.00 (13282.85)	-0.00 (-0.87)	1.00
RMW	0.27 (2.58)	0.33 (3.31)	-0.04 (-1.32)	-0.28 (-3.20)	-0.00 (-0.03)	0.04 (0.81)	0.19
CMA	0.34 (3.63)	0.19 (2.83)	-0.09 (-4.42)	0.03 (0.86)	0.46 (13.52)	0.04 (1.51)	0.55
		$\alpha_q$	$\beta_{MKT}$	$\beta_{ME}$	$\beta_{I/A}$	$\beta_{ROE}$	$R^2$
SMB		0.05 (1.48)	-0.00 (-0.17)	0.94 (62.40)	-0.09 (-4.91)	-0.10 (-5.94)	0.96
HML		0.03 (0.28)	-0.05 (-1.33)	0.00 (0.03)	1.03 (11.72)	-0.17 (-2.17)	0.50
RMW		0.04 (0.42)	-0.03 (-0.99)	-0.12 (-1.78)	-0.03 (-0.35)	0.53 (8.59)	0.49
CMA		0.01 (0.32)	-0.05 (-3.63)	0.04 (1.68)	0.94 (35.26)	-0.11 (-3.95)	0.85

ated with high  $E_t[r_{it+1}]$ , because  $P_{it+1}/B_{it+1} - 1$  tends to be positive in the data. More generally, leading equation (14) by one period at a time and recursively substituting  $P_{it+1}/B_{it+1}$  in the same equation implies a positive  $E_t[\Delta B_{it+\tau}/B_{it}] - E_t[r_{it+1}]$  relation for all  $\tau \geq 1$ .<sup>3</sup>

Finally, after arguing for the negative relation between the expected investment and the expected return, Fama and French (2015) use past investment as the proxy for the expected investment. However, while past profitability forecasts future profitability, past investment does not forecast future investment. Hou, Mo, Xue, and Zhang (2017) document that in annual cross-sectional regressions of future book equity growth on current asset growth, the average  $R^2$  starts at 5% in year one, drops quickly to zero in year four, and remains at zero through year ten. The low persistence of micro-level investment is well known in the lumpy investment literature (Dixit and Pindyck 1994; Doms and Dunne 1998; Whited 1998). In all, the last two concerns imply that the investment factor can only be derived from market-to-book in the valuation equation (11), augmented with the investment-value linkage, which is in turn a key insight from the investment CAPM.

### 2.3.3 Which ROE Factor?

Fama and French (2006) test the expected profitability and expected investment effects predicted by valuation theory. Cross-sectional regressions are used to predict profitability and asset growth one, two, and three years ahead, and the fitted values from these first-stage regressions are used as explanatory variables in second-stage cross-sectional regressions of returns. Contrary to the hypothesized negative relation between the expected investment and expected returns, the average slopes on the expected investment in second-stage regressions are insignificantly positive. Also, with a long list of predictors for future profitability in the first-stage, including lagged fundamentals, returns, analysts forecasts, and the default probability, the expected profitability shows only insignificant, albeit positive, average slopes in the second-stage regressions.

Fama and French (2008) also report an insignificant profitability premium. Profitability is measured as income before extraordinary items (Compustat annual item IB), minus preferred dividends (item DVP), if available, plus income account deferred taxes (item TXDI), if available, divided by contemporaneous book equity. From annual sorts, Fama and French report that profitability sorts “produce the weakest average hedge portfolio returns,” and suggest that with controls for size and

---

<sup>3</sup>Lettau and Ludvigson (2002) show that high aggregate risk premiums forecast high long-term aggregate investment growth rates. Using cross-sectional regressions, Fama and French (2006) report a weakly positive relation between the expected investment and the expected returns. However, Aharoni, Grundy, and Zeng (2013) report a negative relation, and attribute the difference to firm-level variables, as opposed to per share variables in Fama and French.

book-to-market, “hedge returns do not provide much basis for the conclusion that there is a positive relation between average returns and profitability (p. 1663).

Novy-Marx (2013) shows that a different profitability measure, gross profits-to-assets, produces a significant premium in annual sorts. It is argued that gross profits are a cleaner accounting measure of economic profitability than earnings. Expensed investments, such as research and development, advertising, and employee training, reduce earnings, but do not increase book equity. These expenses give rise to higher future economic profits, and are better captured by gross profits. Fama and French (2015) base their profitability factor, RMW, on the gross profitability effect.

A cleaner measure of economic profits might not be the whole story. Panel A of Table 2 reports factor regressions for the annually formed gross profits-to-assets deciles. Consistent with Novy-Marx (2013), the high-minus-low decile earns an average return of 0.38% per month ( $t = 2.62$ ) and a Carhart alpha of 0.49% ( $t = 3.39$ ). The Fama-French five-factor alpha is 0.19% ( $t = 1.46$ ), and the  $q$ -factor alpha is 0.18% ( $t = 1.24$ ). More important, Panel B replicates the tests, but scales gross profits with one-year-lagged assets, not contemporaneous assets. The average high-minus-low return falls to only 0.16% per month ( $t = 1.04$ ). Intuitively, the gross profits-to-assets ratio equals the gross profits-to-lagged assets ratio divided by asset growth (contemporaneous assets-to-lagged assets). As such, the gross profitability effect is contaminated by a hidden investment effect. Once this investment effect is purged, the gross profitability effect largely disappears.

Which deflator should be used to scale profits, lagged or contemporaneous assets? Because in Compustat, both profits and assets are measured at the end of a period, economic logic implies that profits should be scaled by one-period-lagged assets. Intuitively, profits are generated by the one-period-lagged assets. Contemporaneous assets at the end of the period in Compustat are accumulated via the investment process over the course of the current period. With one-period time-to-build, contemporaneous assets can start to generate profits only at the beginning of next period.

In contrast to the small, insignificant profitability premium from annual sorts on gross profits-to-lagged assets, Panel C shows that monthly sorts per Hou, Xue, and Zhang (2015) on the same variable revive the profitability premium. Due to limited coverage for the cost of goods sold (Compustat quarterly item COGSQ) and total assets (item ATQ), the sample starts from January 1976. The high-minus-low decile earns an average return of 0.51% per month ( $t = 3.4$ ) and a Carhart alpha of 0.56% ( $t = 3.81$ ). The five-factor alpha is 0.3% ( $t = 2.14$ ), but the  $q$ -factor alpha is 0.2% ( $t = 1.41$ ).

**Table 2 : Deciles on Different Gross Profitability Measures**

All deciles are based on NYSE breakpoints and value-weighted returns.  $m$  is average excess returns, and  $t_m$  is its  $t$ -statistics.  $\alpha_C, a$ , and  $\alpha_q$  are the alphas from the Carhart model, the Fama-French five-factor model, and the  $q$ -factor model, and  $t_C, t_a$ , and  $t_q$  are their  $t$ -statistics, respectively. m.a.e. (mean absolute error) is the average magnitude of alphas across the testing deciles. The GRS  $p$ -value testing that all ten alphas are jointly zero is in brackets beneath the m.a.e. for a given factor model. All the  $t$ -statistics are adjusted for heteroscedasticity and autocorrelations. In Panel A, at the end of June of year  $t$ , stocks are split into deciles based on gross profits, measured as total revenue (Compustat annual item REVT) minus cost of goods sold (item COGS), divided by total assets (item AT), all for the fiscal year ending in calendar year  $t - 1$ . Monthly decile returns are calculated from July of year  $t$  to June of  $t + 1$ , and the deciles are rebalanced in June of  $t + 1$ . Panel B replicates Panel A, except that the denominator of the sorting variable is total assets for the fiscal year ending in calendar year  $t - 2$ . In Panel C, at the beginning of each month  $t$ , stocks are split into deciles based on total revenue (Compustat quarterly item REVTQ) minus cost of goods sold (item COGSQ), divided by one-quarter-lagged assets (item ATQ), all from the fiscal quarter ending at least four months ago. Monthly returns are calculated for month  $t$ , and the deciles are rebalanced at the beginning of month  $t + 1$ .

	Low	2	3	4	5	6	7	8	9	High	H-L	m.a.e.
Panel A: Gross profits-to-current assets, annual sorts, January 1967–December 2014												
$m$	0.35	0.40	0.46	0.49	0.62	0.57	0.51	0.49	0.61	0.73	0.38	
$t_m$	1.66	2.04	2.21	2.32	3.05	2.76	2.28	2.23	2.91	3.41	2.62	
$\alpha_C$	-0.15	-0.19	-0.12	-0.08	0.06	0.03	0.14	0.18	0.24	0.35	0.49	0.15
$t_C$	-1.51	-2.53	-1.42	-0.86	0.74	0.38	1.52	1.98	3.05	3.63	3.39	[0.00]
$a$	0.07	-0.16	-0.10	-0.14	0.01	0.01	0.03	0.07	0.16	0.27	0.19	0.10
$t_a$	0.85	-1.98	-1.13	-1.57	0.17	0.16	0.36	0.77	2.25	2.58	1.46	[0.04]
$\alpha_q$	0.02	-0.16	-0.11	-0.10	0.04	0.08	0.16	0.19	0.15	0.20	0.18	0.12
$t_q$	0.21	-1.76	-1.21	-1.13	0.50	1.02	1.65	1.65	1.82	1.90	1.24	[0.11]
Panel B: Gross profits-to-lagged assets, annual sorts, January 1967–December 2014												
$m$	0.46	0.45	0.49	0.54	0.63	0.61	0.50	0.49	0.60	0.61	0.16	
$t_m$	2.30	2.36	2.37	2.55	3.28	2.91	2.42	2.21	2.89	2.69	1.04	
Panel C: Gross profits-to-lagged assets, monthly sorts, January 1976–December 2014												
$m$	0.39	0.59	0.46	0.51	0.62	0.71	0.65	0.58	0.72	0.91	0.51	
$t_m$	1.56	2.94	1.98	2.22	2.81	3.21	2.75	2.58	3.22	3.87	3.40	
$\alpha_C$	-0.19	-0.03	-0.24	-0.20	-0.06	0.11	0.12	0.12	0.23	0.37	0.56	0.17
$t_C$	-1.66	-0.38	-2.57	-2.14	-0.71	1.35	1.24	1.53	2.63	3.76	3.81	[0.00]
$a$	0.03	0.05	-0.19	-0.24	-0.17	0.00	0.01	-0.03	0.14	0.33	0.30	0.12
$t_a$	0.26	0.70	-2.00	-2.51	-1.94	0.00	0.08	-0.31	1.55	2.96	2.14	[0.02]
$\alpha_q$	0.05	0.04	-0.16	-0.19	-0.08	0.07	0.16	0.00	0.14	0.25	0.20	0.11
$t_q$	0.40	0.47	-1.54	-2.03	-0.84	0.78	1.27	-0.04	1.47	2.12	1.41	[0.12]

**Table 3 : The ROA and ROE Deciles, Monthly Sorts, January 1976–December 2014**

All deciles are based on NYSE breakpoints and value-weighted returns.  $m$  is average excess returns, and  $t_m$  is its  $t$ -statistics.  $\alpha_C, a$ , and  $\alpha_q$  are the alphas from the Carhart model, the Fama-French five-factor model, and the  $q$ -factor model, and  $t_C, t_a$ , and  $t_q$  are their  $t$ -statistics, respectively. m.a.e. is the average magnitude of alphas across the testing deciles. The GRS  $p$ -value is in brackets beneath the m.a.e. for a given factor model. The  $t$ -statistics are adjusted for heteroscedasticity and autocorrelations. At the beginning of each month  $t$ , stocks are split into deciles based on quarter ROA and ROE, measured as income before extraordinary items (Compustat quarterly item IBQ) scaled by one-quarter-lagged assets (item ATQ) or one-quarter-lagged book equity, respectively. Quarterly earnings are used immediately after the most recent quarterly earnings announcement dates (item RDQ). Monthly decile returns are calculated for month  $t$ , and the deciles are rebalanced at the beginning of month  $t + 1$ .

	Low	2	3	4	5	6	7	8	9	High	H–L	m.a.e.
Panel A: ROA (quarterly earnings-to-one-quarter-lagged assets)												
$m$	0.20	0.44	0.57	0.68	0.57	0.75	0.72	0.63	0.70	0.78	0.58	
$t_m$	0.56	1.58	2.25	3.30	2.87	3.61	3.58	3.05	3.25	3.41	2.53	
$\alpha_C$	-0.42	-0.15	-0.03	0.07	-0.04	0.12	0.11	0.08	0.13	0.27	0.68	0.14
$t_C$	-2.70	-1.38	-0.35	0.80	-0.57	1.48	1.59	0.94	1.82	3.61	3.61	[0.00]
$a$	-0.14	-0.16	-0.13	0.06	-0.06	0.03	0.02	-0.11	0.04	0.24	0.39	0.10
$t_a$	-1.30	-1.33	-1.38	0.75	-0.73	0.35	0.27	-1.56	0.60	3.07	2.85	[0.10]
$\alpha_q$	0.16	0.10	0.10	0.17	-0.04	0.06	0.03	-0.07	-0.01	0.13	-0.02	0.09
$t_q$	1.39	1.01	0.98	1.74	-0.49	0.61	0.43	-0.85	-0.10	1.60	-0.21	[0.20]
Panel B: ROE (quarterly earnings-to-one-quarter-lagged book equity)												
$m$	0.16	0.43	0.50	0.56	0.68	0.58	0.68	0.71	0.68	0.88	0.72	
$t_m$	0.45	1.57	2.02	2.69	3.36	2.68	3.34	3.37	3.17	3.93	2.79	
$\alpha_C$	-0.49	-0.15	-0.04	-0.03	0.08	-0.03	0.09	0.12	0.10	0.31	0.80	0.14
$t_C$	-3.09	-1.40	-0.38	-0.30	0.99	-0.32	1.30	1.60	1.46	3.42	3.72	[0.00]
$a$	-0.24	-0.15	0.00	-0.02	0.05	-0.03	0.02	0.04	0.10	0.15	0.39	0.08
$t_a$	-2.02	-1.29	0.01	-0.17	0.55	-0.32	0.23	0.48	1.45	1.71	2.68	[0.33]
$\alpha_q$	0.09	0.13	0.28	0.10	0.11	0.00	0.05	-0.05	0.03	0.03	-0.05	0.09
$t_q$	0.75	1.41	2.99	1.05	1.24	0.01	0.63	-0.59	0.37	0.39	-0.40	[0.06]

Gross profits are largely irrelevant in monthly sorts. Table 3 reports deciles formed monthly on ROA (earnings-to-lagged assets) and, separately, on ROE. To ease comparison, the sample starts at January 1976 as in Panel C of Table 2. Panel A shows that monthly ROA sorts yield an average high-minus-low return of 0.58% per month ( $t = 2.53$ ). From Panel B, the average return for the high-minus-low ROE decile is even higher, 0.72% ( $t = 2.79$ ). As such, despite much maligned in annual sorts, earnings perform better than gross profits in monthly sorts.

### 2.3.4 Independent Comparison

Several studies have independently compared the  $q$ -factor model with the Fama-French (2015) five-factor model. Their empirical results are broadly in line with those in Hou et al. (2016). Barillas

and Shanken (2015) develop a Bayesian test procedure that allows model comparison, i.e., the computation of model probabilities for the collection of all possible pricing models based on subsets of the given factors. Applying this new methodology to the head-to-head contest between the  $q$ -factor model and the five-factor model, Shanken (2015) reports model probabilities that are overwhelmingly in favor of the  $q$ -factor model (97% versus 3%).

Stambaugh and Yuan (2016) independently verify the two key results in Hou, Xue, and Zhang (2016). First, the  $q$ -factor model explains the Fama-French (2015) five-factor returns in time series regressions, but the five-factor model cannot explain the  $q$ -factor returns. Second, the  $q$ -factor model outperforms the five-factor model in explaining a wide array of anomalies in the cross section. Stambaugh and Yuan also propose a “mispricing” factor model. The mispricing factors are formed by aggregating a stock’s information across 11 anomalies with average rankings within two anomaly clusters that exhibit the greatest comovement in high-minus-low returns. Empirically, the mispricing factor model and the  $q$ -factor model are “evenly matched” in factor spanning tests, but the mispricing model has an edge in explaining anomalies, especially in the set of 11. The edge is not surprising, however, as the two mispricing factors are basically principal components extracted ex post from the 11 anomalies. The breakpoints for the mispricing factors are 20-60-20, as opposed to the more standard 30-40-30 from Fama and French (1993). Most important, one of the mispricing factors has a correlation about 0.7 with the investment factor, and another also has a correlation of 0.7 with the ROE factor in the  $q$ -factor model. As such, the principal component analysis essentially uncovers the  $q$ -factors that are in turn motivated from the economic theory in the investment CAPM.

## 2.4 Notes

The  $q$ -factor model is built on a rich empirical literature in finance and accounting. Most important, as the  $q$ -factor model uses the Fama-French three- and five-factor models as a straw man, it is only natural to acknowledge our enormous intellectual debt to Fama and French. Fama and French (1992) show that size and book-to-market combine to describe average returns in cross-sectional regressions, and that the relation between the market beta and average returns is flat, even when beta is used alone. Fama and French (1993) propose the three-factor model to replace the CAPM as the workhorse for estimating expected stock returns. Fama and French (1996) show that, except for momentum, the three-factor model summarizes the cross section of expected returns as of the mid-1990s. Carhart (1997) augments the three-factor model with a momentum factor. Perhaps

in response to the challenge by Hou, Xue, and Zhang (2015), Fama and French (2015) upgrade their three-factor model with two new factors that resemble the investment and ROE factors in the  $q$ -factor model. It is evident that the intellectual designs of the  $q$ -factor model, including its econometric tests, factor construction, formation of testing portfolios, and above all, the taste of the economic question, are all deeply influenced by Fama and French.

#### 2.4.1 Related Literature on the Investment Premium

**Different Forms of the Investment Premium** The first group of studies documents the investment effect in various forms, and shows how it relates to other cross-sectional patterns.

Titman, Wei, and Xie (2004) show that firms that increase capital investment subsequently earn lower average returns than firms that decrease capital investment. This effect is also stronger among firms with higher cash flows, implying higher investment discretion. Titman et al. interpret the evidence as investors underreacting to the empire building implications of increasing investment. Empire building means that managers invest for their own private benefits rather than the benefits for shareholders (Jensen 1986). However, the investment CAPM, derived without empire building, is consistent with the evidence that the investment effect is stronger in firms with higher cash flows. Taking the first-order derivative of equation (7) with respect to investment-to-capital yields  $\partial r_{it+1}^S / \partial (I_{it}/K_{it}) = -aX_{it+1}/[1 + a(I_{it}/K_{it})]^2$ . Its magnitude rises with profitability,  $X_{it+1}$ , meaning that the investment effect should be stronger among firms with higher cash flows.

Titman, Wei, and Xie (2004) measure abnormal investment as the ratio of capital expenditure divided by sales, scaled by the prior three-year moving average of this ratio. Dividing investment by sales makes the ratio closer to profitability than to investment. Hou, Xue, and Zhang (2016) show that the high-minus-low abnormal investment decile earns on average  $-0.31\%$  per month ( $t = -2.2$ ) from 1967 to 2014. The  $q$ -factor alpha is  $-0.17\%$  ( $t = -1.05$ ), along with an insignificant investment factor loading of 0.13 ( $t = 1.04$ ) but a significant ROE factor loading of  $-0.2$  ( $t = -2.26$ ).

Anderson and Garcia-Feijóo (2006) show that sorting on book-to-market provides a large spread in investment growth across extreme deciles, and that firms with high investment growth earn significantly lower subsequent returns on average than firms with low investment growth. Xing (2008) shows that a low-minus-high investment growth factor contains similar information as the value factor, and can price the 25 size and book-to-market portfolios as well as the value factor. Both studies interpret their evidence as consistent with the investment CAPM.

Lyandres, Sun, and Zhang (2008) show that the investment effect helps interpret the new issues puzzle (Ritter 1991; Loughran and Ritter 1995). Lyandres et al. show that adding an investment factor into the CAPM and the Fama-French three-factor model reduces a substantial amount of the underperformance following initial public offerings, seasoned equity offerings, and convertible debt offerings, as well as the composite issuance effect (Daniel and Titman 2006). Also, Lyandres et al. document that equity issuers invest much more relative to their assets than nonissuers matched on size and book-to-market, despite similar profitability across issuers and matching nonissuers.

Cooper, Gulen, and Schill (2008) document the strong predictive power of the annual growth rates of total assets in the cross section. Their key insight is that the investment effect is a pervasive phenomenon, going beyond specific components of investment explored in prior studies. In particular, their Table II shows that from 1968 to 2002, the high-minus-low asset growth decile earns on average a whopping equal-weighted return of  $-1.73\%$  per month and a value-weighted return of  $-1.05\%$ . Cooper et al. argue that “bias in the capitalization of new investments leads to a host of potential investment policy distortions,” and interpret their evidence as suggesting that “such potential distortions are present and economically meaningful (p. 1648).”

Cooper, Gulen, and Schill (2008) have clearly influenced the  $q$ -factor model and the Fama-French five-factor model. Both form the investment factor on total asset growth. However, Cooper et al.’s evidence on the predictive power of asset growth is exaggerated by excessively weighting on microcaps. Their deciles are formed with all breakpoints, rather than NYSE breakpoints, and the portfolio returns are equal-weighted to give microcaps disproportionately large weights. Hou, Xue, and Zhang (2016) show that from 1967 to 2014, the high-minus-low asset growth decile earns  $-0.46\%$  per month ( $t = -2.92$ ) with NYSE breakpoints and value-weighted returns and  $-0.69\%$  ( $t = -4.74$ ) with all-but-micro breakpoints and equal-weighted returns. Also, the investment CAPM predicts that high investment implies low subsequent returns. As such, this evidence does not necessarily mean value-destroying investment distortions such as empire building.

Butler, Cornaggia, Grullon, and Weston (2011) propose a clever identification strategy to disentangle investment and behavioral market timing explanations for the underperformance following security issuance. The investment CAPM says that issuers invest more, and are associated with lower costs of capital than nonissuers. The market timing explanation says that managers issue more equity relative to debt when equity is overvalued, and repurchase more equity relative to debt when equity is undervalued. While the investment CAPM says that only the amount of net financing

forecasts returns, the market timing hypothesis predicts that the composition of net financing (equity relative to debt) is more important. Empirically, Butler et al. report pervasive evidence that conditional on the amount of net financing, the composition of financing does not forecast returns.

Cooper and Priestley (2011) report extensive evidence that the negative relation between investment and average returns is related to macroeconomic risk. First, low investment firms have substantially higher loadings on the Chen, Roll, and Ross (1986) macroeconomic risk factors than high investment firms, especially on the growth rate of industrial production and the term spread. Second, the large loading spreads on these two macro risk factors, combined with their large estimated risk premiums from two-pass cross-sectional regressions, account for 60–80% of the average return spreads across extreme investment deciles. Third, macro risk loadings fall during high investment periods, but rise during disinvestment periods. Finally, the investment factor returns contain information about future real growth rates of industrial production, gross domestic product, aggregate corporate earnings, and aggregate investment. However, Cooper and Priestley do not conduct similar risk analysis for the positive relation between profitability and average returns.

**The Cross-sectional Variation of the Investment Premium** Li and Zhang (2010) argue that investment frictions should strengthen the investment effect. Intuitively, when investment is frictionless (adjustment costs are zero), investment would be infinitely elastic to changes in the discount rate. As investment becomes more frictional, investment becomes less elastic to the discount rate changes. In other words, a given magnitude of change in investment corresponds to a higher magnitude of change in the discount rate, meaning that the investment effect in the average returns becomes stronger as investment becomes more frictional.

The impact of investment frictions on the investment effect is similar to limits to arbitrage (Shleifer and Vishny 1997). Trading frictions from the investors' side make arbitrage activities costly and incomplete, allowing mispricing to persist in the marketplace. Attributing the investment effect to mispricing, limits to arbitrage imply that it should be stronger among firms with stronger limits to arbitrage. Because the investment CAPM and mispricing emphasize different types of frictions that likely coexist in the data, their effects are not mutually exclusive. In addition, to the extent that firms with stocks that are more costly to trade could also face higher investment costs, investment frictions and limits to arbitrage could both be at work for the same set of firms simultaneously.

Empirically, using proxies for financial constraints (asset size, payout ratio, and credit rating) to

measure investment frictions, Li and Zhang (2010) report only weak evidence that the investment effect is stronger in firms with stronger investment frictions. In contrast, idiosyncratic volatility and dollar trading volume, which are common proxies for limits to arbitrage, dominate proxies of financial constraints in characterizing the cross-sectional variation of the investment effect.

Lam and Wei (2011) conduct more comprehensive tests by using ten proxies of limits to arbitrage, including idiosyncratic volatility, analyst coverage, analyst forecast dispersion, cash flow volatility, the number of institutional shareholders, stock price, bid-ask spread, institutional ownership, absolute return-to-volume, and dollar trading volume, as well as four proxies of financial constraints, firm age, asset size, payout ratio, and credit rating. Lam and Wei show that the proxies for limits to arbitrage are highly correlated with those for investment frictions. The evidence from equal-weighted returns is consistent with both limits to arbitrage and investment frictions hypotheses, but the evidence from value-weighted returns is weaker. Finally, in direct comparisons, each hypothesis is supported by a fair and similar amount of evidence.

Lipson, Mortal, and Schill (2011) refute Fama and French's (2008) conclusion that the investment effect exists only in small firms. Lipson et al. show that Fama and French's investment measure excludes the part of asset growth related to equity issues, which are a major source of financing for large firms. Also, the investment effect is stronger in firms with high idiosyncratic volatility, and is concentrated around earnings announcement dates. Finally, analysts forecast errors are higher for faster growing firms, consistent with the mispricing interpretation.

Titman, Wei, and Xie (2013) study the investment effect in international markets, and find that it is stronger in countries with more developed markets than in countries with less developed markets. This evidence lends support to the investment CAPM because financial market development aligns managers' incentives with shareholders, and the investment effect arises from maximizing the shareholder value. Titman et al. also show that the investment effect is not related to corporate governance or trading costs across countries, inconsistent with the mispricing interpretation.

Watanabe, Xu, Yao, and Yu (2013) conduct a more comprehensive study of the investment effect in international equity markets. If the investment effect is due to mispricing, it should be stronger in countries in which stocks are less efficiently priced and mispricing is more difficult to arbitrage away. In contrast, if the investment effect is due to the investment CAPM, it should be stronger in countries in which stocks are more efficiently priced. Watanabe et al use four country-level proxies for market efficiency: (i) stock return synchronicity, which is the average  $R^2$  from firm-level market

regressions within each country, and is negatively related to the amount of firm-specific information impounded into individual stock prices; (ii) future earnings response coefficients, a measure of the information content of stock prices for future earnings; (iii) the developed market status from the International Finance Corporation; and (iv) the importance of stock market to the economy, which is measured by total market capitalization-to-gross domestic product, as well as the number of publicly traded companies and initial public offerings scaled by population. Empirically, the investment effect is stronger in countries with lower stock return synchronicity and higher future earnings response coefficients, in developed markets, and in economies in which stock markets are more important. The evidence lends support to the investment CAPM, but casts doubt on mispricing.

**Explaining the Accrual Anomaly** Accounting accruals (earnings minus cash flows) allow a firm to measure its performance by recognizing economic events regardless of the timing of cash transactions. The basic idea is to match revenues to expenses at the time of a transaction, not a cash payment. Sloan (1996) shows that firms with high accruals earn lower average returns than firms with low accruals, and interprets the evidence as investors overreacting to the persistence of the accrual component of earnings in forming earnings expectations. These naive investors are subsequently surprised, when realized earnings of firms with high accruals fall short of, and those of firms with low accruals exceed, prior expectations. Richardson, Sloan, Soliman, and Tuna (2005) rank components of accruals by their degree of reliability, and show that less reliable components lead to less persistence in earnings and stronger accrual anomaly than more reliable components.

Several studies link accruals to investment. Fairfield, Whisenant, and Yohn (2003) decompose the growth in net operating assets into accruals and the growth in long-term net operating assets, and show that after controlling for current profitability, both components forecast future stock returns with a negative slope. Fairfield et al. suggest that investors overreact to both components equally, and that the accrual anomaly is a special case of a more general growth anomaly. Zhang (2007) documents that accruals are positively correlated with employee growth, and that the magnitude of the accrual anomaly monotonically increases with the accruals-employee growth covariation.

Interpreting accruals as working capital investment, Wu, Zhang, and Zhang (2010) apply the investment CAPM to the accrual anomaly. The investment CAPM helps explain Richardson, Sloan, Soliman, and Tuna's (2005) key finding that less reliable accruals lead to stronger accrual anomaly. Empirically, less reliable accruals are more correlated with investment than more reliable accruals.

Conceptually, less reliable changes in noncash working capital and changes in net noncurrent operating assets represent direct forms of investment in short-term and long-term capital, respectively. In contrast, more reliable changes in net financial assets contain diverse components, including marketable securities and financial liability, which are less correlated with investment.

Table 4 reports factor regressions of deciles formed on operating accruals, measured with the balance sheet approach of Sloan (1996), scaled by average total assets. The high-minus-low decile earns an average return of  $-0.29\%$  per month ( $t = -2.13$ ). Consistent with Wu, Zhang, and Zhang (2010), a two-factor model with the market and investment factors yields a high-minus-low alpha of  $-0.1\%$  ( $t = -0.66$ ). The investment factor loading is  $-0.53$ , which is more than four standard errors from zero. However, as shown in Hou, Xue, and Zhang (2015), the full-fledged  $q$ -factor model fails to capture the accrual anomaly with an alpha of  $-0.42\%$  ( $t = -2.96$ ). The culprit is a large, positive ROE factor loading,  $0.31$  ( $t = 4.08$ ), which goes in the wrong way to explain the average return.<sup>4</sup>

Following Ball, Gerakos, Linnainmaa, and Nikolaev (2016), Table 4 also experiments with an alternative  $q$ -factor model with a cash-based ROE factor. The cash-based ROE is quarterly cash-based operating profits divided by one-quarter-lagged book equity.<sup>5</sup> The sample starts in January 1972 due to limited coverage in Compustat quarterly files in earlier years. The construction of the cash-based  $q$ -factors is identical to that for the benchmark  $q$ -factors, except that at the beginning of each month  $t$ , stocks are split on the cash-based ROE for the fiscal quarter ending at least four months ago, as opposed to the ROE calculated with the latest announced earnings.

The cash-based ROE factor earns on average  $0.59\%$  per month ( $t = 7.5$ ) from January 1972 to December 2014. Its Carhart alpha is  $0.57\%$  ( $t = 7.83$ ), and the Fama-French five-factor alpha  $0.54\%$  ( $t = 6.81$ ). The investment factor earns on average  $0.33\%$  ( $t = 3.28$ ). Its Carhart alpha is  $0.17\%$  ( $t = 2.09$ ), and the five-factor alpha is close to zero. The average size factor return is  $0.28\%$  ( $t = 1.9$ ). The correlation between the benchmark and cash-based ROE factors is only  $0.56$ . The

---

<sup>4</sup>Hou, Xue, and Zhang (2015) also show that the investment factor loading for the high-minus-low accrual decile is close to zero, in contrast to  $-0.47$  ( $t = -5.19$ ) in Table 4. The difference arises because Hou et al. measure operating accruals, starting from 1988, as net income (Compustat annual item NI) minus net cash flow from operations (item OANCF) from the statement of cash flows. In untabulated results, from July 1989 to December 2014, the high-minus-low decile on this cash flows-based operating accruals earns an average return of  $-0.38\%$  per month ( $t = -2.12$ ). The  $q$ -factor alpha is  $-0.51\%$  ( $t = -2.64$ ), the investment factor loading  $-0.07$ , and the ROE factor loading  $0.28$  ( $t = 3.06$ ).

<sup>5</sup>Cash-based operating profits are revenues (Compustat quarterly item REVTQ) minus cost of goods sold (item COGSQ) minus selling, general, and administrative expenses (item XSGAQ) plus research and development expenditure (item XRDQ, zero if missing) minus change in accounts receivable (item RECTQ) minus change in inventory (item INVTQ) plus change in deferred revenue (item DRCQ plus item DRLTQ) plus change in trade accounts payable (item APQ). All balance sheet changes are quarterly changes, and are set to zero if missing.

**Table 4 : The Balance Sheet Operating Accruals Deciles**

Operating accruals are defined as  $(dCA-dCASH)-(dCL-dSTD-dTP)-DP$ , in which  $dCA$  is the change in current assets (Compustat annual item ACT),  $dCASH$  is the change in cash or cash equivalents (item CHE),  $dCL$  is the change in current liabilities (item LCT),  $dSTD$  is the change in debt included in current liabilities (item DLC),  $dTP$  is the change in income taxes payable (item TXP, zero if missing), and  $DP$  is depreciation and amortization (item DP). At the end of June of each year  $t$ , stocks are split into deciles with NYSE breakpoints on operating accruals for the fiscal year ending in calendar year  $t - 1$  scaled by average total assets (item AT) for the fiscal years ending in  $t - 1$  and  $t - 2$ . Monthly value-weighted decile returns are calculated from July of year  $t$  to June of  $t + 1$ , and the deciles are rebalanced in June of  $t + 1$ .  $m$  is average excess returns,  $t_m$  is its  $t$ -statistics, and m.a.e. (mean absolute error) is the average magnitude of alphas across the testing deciles. The GRS  $p$ -value is in brackets beneath the m.a.e. for a given factor model. All the  $t$ -statistics are adjusted for heteroscedasticity and autocorrelations.

	Low	2	3	4	5	6	7	8	9	High	H-L	m.a.e.
Panel A: Average returns, January 1967–December 2014												
$m$	0.61	0.55	0.56	0.63	0.62	0.60	0.62	0.42	0.44	0.32	-0.29	
$t_m$	2.40	2.61	3.06	3.42	3.09	3.19	3.19	2.00	1.93	1.14	-2.13	
Panel B: Two-factor regressions, the market and investment factors, January 1967–December 2014												
$\alpha$	0.17	0.06	0.05	0.11	0.10	0.06	0.22	-0.03	0.05	0.07	-0.10	0.09
$\beta_{MKT}$	1.09	0.99	0.93	0.93	0.93	0.91	0.92	0.96	1.05	1.15	0.07	[0.01]
$\beta_{I/A}$	-0.26	-0.04	0.09	0.10	0.10	0.16	-0.13	-0.09	-0.31	-0.79	-0.53	
$t_\alpha$	1.48	0.67	0.78	1.41	1.35	0.88	3.54	-0.35	0.62	0.78	-0.66	
$t_{MKT}$	27.84	37.17	56.71	35.95	42.16	52.43	43.58	47.76	41.83	43.73	1.46	
$t_{I/A}$	-2.45	-0.29	1.77	1.79	1.71	1.93	-2.66	-1.28	-4.32	-12.70	-4.03	
Panel C: The benchmark $q$ -factor regressions, January 1967–December 2014												
$\alpha_q$	0.33	0.16	0.13	0.09	0.11	-0.04	0.14	-0.12	-0.03	-0.09	-0.42	0.12
$\beta_{MKT}$	1.07	1.00	0.94	0.94	0.95	0.94	0.94	0.97	1.05	1.10	0.03	[0.00]
$\beta_{ME}$	-0.06	-0.12	-0.12	-0.03	-0.08	-0.03	-0.04	0.05	0.05	0.37	0.43	
$\beta_{I/A}$	-0.28	-0.05	0.08	0.10	0.10	0.17	-0.13	-0.07	-0.30	-0.75	-0.47	
$\beta_{ROE}$	-0.22	-0.10	-0.07	0.04	0.02	0.17	0.13	0.13	0.09	0.10	0.31	
$t_q$	2.65	1.39	1.81	1.01	1.36	-0.48	2.10	-1.57	-0.35	-1.00	-2.96	
$t_{MKT}$	29.98	38.57	52.20	36.98	45.00	57.17	43.81	45.38	37.99	46.56	0.86	
$t_{ME}$	-1.21	-2.39	-3.57	-0.94	-2.14	-0.86	-1.21	1.72	1.42	11.77	7.60	
$t_{I/A}$	-3.11	-0.45	1.62	1.76	1.78	2.41	-3.01	-1.25	-4.48	-15.72	-5.19	
$t_{ROE}$	-2.55	-1.27	-2.09	0.80	0.46	3.28	2.87	3.10	1.78	2.03	4.08	
Panel D: The $q$ -factor regressions with the cash-based ROE factor, January 1972–December 2014												
$\alpha_q$	0.14	0.12	0.12	0.10	0.15	-0.02	0.13	-0.07	0.02	-0.03	-0.17	0.09
$\beta_{MKT}$	1.11	1.02	0.94	0.94	0.92	0.91	0.93	0.95	1.04	1.11	0.00	[0.07]
$\beta_{ME}$	0.03	-0.09	-0.09	-0.03	-0.10	-0.06	-0.07	0.00	0.01	0.29	0.26	
$\beta_{I/A}$	-0.16	-0.05	0.03	0.10	0.06	0.12	-0.10	-0.03	-0.26	-0.64	-0.48	
$\beta_{ROE}$	-0.09	-0.03	0.00	0.05	-0.03	0.21	0.11	0.02	-0.04	-0.14	-0.05	
$t_q$	1.07	1.16	1.44	1.15	1.73	-0.22	1.87	-0.74	0.18	-0.37	-1.09	
$t_{MKT}$	26.54	39.37	51.24	34.02	46.76	49.24	38.93	42.14	33.86	41.07	0.07	
$t_{ME}$	0.69	-1.95	-3.22	-1.20	-2.57	-1.89	-2.09	-0.13	0.40	8.13	4.50	
$t_{I/A}$	-1.51	-0.44	0.66	1.97	1.14	2.12	-2.45	-0.59	-4.20	-11.79	-4.26	
$t_{ROE}$	-0.92	-0.24	0.02	0.75	-0.43	2.17	1.93	0.34	-0.46	-2.52	-0.48	

cash-based ROE factor has an alpha of 0.42% ( $t = 5.1$ ) in the  $q$ -factor model. However, the cash-based  $q$ -factor model cannot subsume the benchmark  $q$ -factors, with alphas of 0.15% ( $t = 2.94$ ), 0.15% ( $t = 3.55$ ), and 0.23% ( $t = 2.37$ ) for the size, investment, and ROE factors, respectively.

Panel D in Table 4 shows that the cash-based  $q$ -factor model explains the accrual anomaly. The alpha of the high-minus-low decile is  $-0.17\%$  per month ( $t = -1.09$ ), and the model cannot be rejected by the GRS test ( $p = 0.07$ ). The investment factor loading is  $-0.48$ , which is close to  $-0.47$  in the benchmark  $q$ -factor model. More important, the cash-based ROE factor loading is only  $-0.05$  ( $t = -0.48$ ), in contrast to  $0.31$  ( $t = 4.08$ ) for the benchmark ROE factor. Intuitively, because earnings equal cash flows plus accruals, high accrual firms would appear more “profitable,” and load more heavily on the earnings-based benchmark ROE factor than low accrual firms. The cash-based ROE factor avoids this pitfall, as accruals do not enter the cash-based ROE.

#### 2.4.2 Related Literature on the Profitability Premium

The ROE factor in the  $q$ -factor model is built on the illustrious accounting literature on post-earnings-announcement drift (earnings momentum) launched by Ball and Brown (1968). In particular, the monthly sort on quarterly earnings in the ROE factor is heavily influenced by the construction of earnings momentum (Chan, Jegadeesh, and Lakonishok 1996).

**Post-earnings-announcement Drift** I do not attempt to review the literature on post-earnings-announcement drift (see Richardson, Tuna, and Wysocki 2010 for an extensive review). Rather, the goal is to describe how this literature informs and influences the investment CAPM literature, and how the investment CAPM might add to the accounting literature.

Ball and Brown (1968) show that the sign and magnitude of stock returns in the post-earnings announcement period are positively correlated with the sign and magnitude of unexpected earnings. Foster, Olsen, and Shevlin (1984) use more sophisticated models of expected earnings, and show that the magnitude of the post-earnings-announcement drift decreases with firm size. Bernard and Thomas (1989) show that a disproportionately large fraction of the drift is concentrated in the three-day period surrounding the earnings announcement dates in subsequent quarters.

The investment CAPM is consistent with the evidence that a disproportionately large fraction of earnings momentum is concentrated in a few days surrounding the subsequent earnings announcements. Intuitively, equation (7) holds ex post in realization, state by state, period by period, an

observation first made in Cochrane (1991). In the static model, stock returns should move *only* when earnings shocks hit, and all of earnings momentum should be materialized on earnings announcement dates. In the multiperiod model, investment,  $I_{it+1}$ , which correlates positively but not perfectly with earnings shocks, also appears in the numerator of equation (7). As such, only a fraction of earnings momentum should be realized on earnings announcement dates (Liu and Zhang 2014).

The investment CAPM also predicts that the magnitude of post-earnings-announcement drift is larger in small firms than in big firms. Equations (5) and (6) imply that  $r_{it+1}^S = (K_{it+1}/P_{it})X_{it+1}$ . As such, the strength of the earnings-return relation decreases with the market equity,  $P_{it}$ . The same equation also implies that the profitability effect in average returns,  $\partial E_t[r_{it+1}^S]/\partial E_t[X_{it+1}]$ , should be stronger in value firms with high  $K_{it+1}/P_{it}$  than in growth firms with low  $K_{it+1}/P_{it}$ . In addition, the value effect in average returns,  $\partial E_t[r_{it+1}^S]/\partial(K_{it+1}/P_{it})$ , should be stronger in more profitable firms than in less profitable firms (Piotroski 2000).

Bernard and Thomas (1990) present the most telling evidence suggesting that stock prices fail to reflect the extent to which earnings deviate from a naive expectation based on a seasonally adjusted random walk, in which expected earnings are the earnings from four quarters ago. Their key evidence is the negative correlation between unexpected earnings for quarter  $t$  and the abnormal returns around the earnings announcement for quarter  $t+4$ , in addition to the positive correlation between unexpected earnings for quarter  $t$  and the drift for quarters  $t+1$  to  $t+3$ . The evidence is interpreted as stock prices failing to reflect fully the mean reversion of earnings changes in four quarters.

Bernard and Thomas (1990) develop clear testable hypotheses based on underreaction to earnings news. Suppose investors form naive earnings expectations based on the detrended seasonally adjusted random walk. Let  $\Pi_{it+1}$  be the detrended earnings for stock  $i$  for quarter  $t+1$ . The market's expected earnings embedded in the stock price are given by  $E_t^M[\Pi_{it+1}] = \Pi_{it-3}$ , in which  $E_t^M[\cdot]$  is the market's expectation conditional on time- $t$  information. When  $\Pi_{it+1}$  is announced, the market perceives the unexpected earnings as  $\Pi_{it+1} - E_t^M[\Pi_{it+1}]$ . Let  $\lambda_i$  be the earnings response coefficient. The abnormal return at the earnings announcement date for quarter  $t+1$  is:

$$r_{it+1}^S - E_t^M[r_{it+1}^S] = \lambda_i(\Pi_{it+1} - E_t^M[\Pi_{it+1}]) = \lambda_i(\Pi_{it+1} - \Pi_{it-3}). \quad (15)$$

Bernard and Thomas (1990) assume the Brown-Rozeff (1979) model for the true earnings:

$$\Pi_{it+1} = \Pi_{it-3} + \phi(\Pi_{it} - \Pi_{it-4}) + \theta\epsilon_{it-3} + \epsilon_{it+1}, \quad (16)$$

in which  $\epsilon_{it+1}$  is the earnings shock at period  $t + 1$ ,  $\phi > 0$ , and  $\theta$  is negative enough that the fourth-order autocorrelation in seasonally differenced earnings is negative. The true expected earnings is:

$$E_t[\Pi_{it+1}] = \Pi_{it-3} + \phi(\Pi_{it} - \Pi_{it-4}) + \theta\epsilon_{it-3}. \quad (17)$$

However, if investors use the naive earnings expectations from the seasonally adjusted random walk:

$$\begin{aligned} r_{it+1}^S - E_t^M[r_{it+1}^S] &= \lambda_i(\Pi_{it+1} - E_t^M[\Pi_{it+1}]) = \lambda_i(\Pi_{it+1} - E_t[\Pi_{it+1}]) + \lambda_i(E_t[\Pi_{it+1}] - E_t^M[\Pi_{it+1}]) \\ &= \lambda_i\epsilon_{it+1} + \lambda_i\phi(\Pi_{it} - \Pi_{it-4}) + \lambda_i\theta\epsilon_{it-3}. \end{aligned} \quad (18)$$

Using the time series earnings model in equation (16) then yields:

$$r_{it+1}^S - E_t^M[r_{it+1}^S] = \lambda_i\epsilon_{it+1} + \lambda_i\phi\epsilon_{it} + \lambda_i\phi^2\epsilon_{it-1} + \lambda_i\phi^3\epsilon_{it-2} + \lambda_i(\theta + \phi^4)\epsilon_{it-3} + \lambda_i\nu_i, \quad (19)$$

in which  $\nu_i$  is a linear combination of earnings shocks prior to  $t - 3$ . As such, if the true earnings follow the Brown-Rozeff model but investors form naive expectations from a seasonal random walk, then the abnormal return at the earnings announcement of quarter  $t + 1$  should have positive but declining correlations with the earnings shocks from quarters  $t, t - 1$ , and  $t - 2$ , but a negative correlation with the earnings shock from quarter  $t - 3$ . The evidence confirms this prediction.

Bernard and Thomas's (1990) evidence is the "most damaging" to efficient markets (Kothari 2001, p. I-194). No rational explanation has been offered to date. However, the investment CAPM seems to make some progress. In particular, multiplying both the numerator and the denominator of equation (7) with  $K_{it+1}$ , which is known at the beginning of period  $t$ , yields:

$$r_{it+1}^S = \lambda_{it}\Pi_{it+1}, \quad (20)$$

in which  $\lambda_{it} \equiv 1/[(1 + a(I_{it}/K_{it}))K_{it+1}] = 1/P_{it}$  is the earnings response coefficient.

Continue to assume that the true earnings follow the Brown-Rozeff (1979) model. Under rational expectations, equations (17) and (20) imply:

$$\begin{aligned} E_t[r_{it+1}^S] &= \lambda_{it}\phi(\Pi_{it} - \Pi_{it-4}) + \lambda_{it}\theta\epsilon_{it-3} + \lambda_{it}\Pi_{it-3} \\ &= \lambda_{it}\phi\epsilon_{it} + \lambda_{it}\phi^2\epsilon_{it-1} + \lambda_{it}\phi^3\epsilon_{it-2} + \lambda_{it}(\theta + \phi^4)\epsilon_{it-3} + \lambda_{it}(\nu_i + \Pi_{it-3}). \end{aligned} \quad (21)$$

The abnormal return at the earnings announcement for quarter  $t + 1$  is:

$$r_{it+1}^S - E_t[r_{it+1}^S] = \lambda_{it}\epsilon_{it+1}, \quad (22)$$

which is unpredictable with time- $t$  information. However, this unpredictability requires that the expected return is measured precisely. In Bernard and Thomas (1989, 1990), abnormal return is calculated as the size-adjusted return, and size seems an imperfect control for the expected return. Suppose the last term in equation (21),  $\lambda_{it}(\nu_i + \Pi_{it-3})$ , is largely expected due to its long lag, but a portion of the remainder of the expected return is mismeasured as a part of the abnormal return. The equation then implies that the “abnormal” return at the earnings announcement of quarter  $t + 1$  would display positive and declining correlations with earnings shocks from prior quarters.

**Fundamental Analysis** In addition to post-earnings-announcement drift, the ROE factor in Hou, Xue, and Zhang (2015) is also rooted in the fundamental analysis literature in accounting. Ou and Penman (1989) combine a large set of financial statement items into one composite metric that predicts the direction of one-year-ahead earnings changes. Lev and Thiagarajan (1993) use conceptual arguments to select 12 fundamental signals to correlate with contemporaneous stock returns, and to forecast future earnings growth. In contrast, Hou et al. use a single variable (profitability) as the expected profitability proxy. This parsimony is designed to reduce estimation errors in the proxy so as to strengthen its predictive power for future profitability and returns.

Abarbanell and Bushee (1997, 1998) show that several fundamental signals, including inventory changes, changes in account receivables, gross margin, changes in selling and administrative expenses, and tax expenses-to-earnings provide information about future returns. This information is also shown to be associated with one-year-ahead earnings news and analysts’ forecast errors. Piotroski (2000) applies context-specific fundamental analysis to a broad portfolio of value stocks. Nine signals are selected to capture a firm’s profitability, leverage, liquidity, source of funds, and operating efficiency. Piotroski finds that the average return earned by a value investor can be increased by at least 7.5% per annum through selecting financially strong value stocks.

More recently, the fundamental analysis literature and the investment CAPM literature have started to show signs of convergence in perspectives. Penman and Zhu (2015) use a model similar to equation (12) to connect expected returns to expected earnings and expected earnings growth. Consistent with the model’s predictions, many accounting variables predict future earnings and

earnings growth in the data in the same direction in which these variables forecast returns. Penman and Zhu interpret the evidence as “consistent with rational pricing in the sense that the returns are those one would expect if the market were efficient in its pricing (p. 1836).”

### 3 Structural Estimation and Tests

#### 3.1 The Multiperiod Investment CAPM

Building on Cochrane (1991), Liu, Whited, and Zhang (2009) derive and test the multiperiod investment CAPM. Time is discrete and the horizon infinite. Firms use capital and costlessly adjustable inputs to produce a homogeneous output. These latter inputs are chosen each period to maximize operating profits, which are defined as revenues minus the expenditures on these inputs. Taking the operating profits as given, firms choose investment to maximize the market value of equity.

Let  $\Pi_{it} = \Pi(K_{it}, X_{it})$  be the operating profits of firm  $i$  at time  $t$ , in which  $K_{it}$  is capital, and  $X_{it}$  is a vector of exogenous aggregate and firm-specific shocks. Firm  $i$  has a Cobb-Douglas production function with constant returns to scale, meaning that  $\Pi(K_{it}, X_{it}) = K_{it} \partial \Pi(K_{it}, X_{it}) / \partial K_{it}$ . The marginal product of capital is parameterized as  $\partial \Pi(K_{it}, X_{it}) / \partial K_{it} = \kappa Y_{it} / K_{it}$ , in which  $\kappa > 0$  is a constant parameter, and  $Y_{it}$  is sales. As such, shocks to operating profits,  $X_{it}$ , are reflected in sales. Capital accumulates as  $K_{it+1} = I_{it} + (1 - \delta_{it}) K_{it}$ , in which  $I_{it}$  is investment, and  $\delta_{it}$  is an exogenous proportional rate of depreciation. Firms incur adjustment costs when investing. The adjustment costs function,  $\Phi(I_{it}, K_{it})$ , is increasing and convex in  $I_{it}$ , decreasing in  $K_{it}$ , and exhibits constant returns to scale in  $I_{it}$  and  $K_{it}$ , i.e.,  $\Phi(I_{it}, K_{it}) = I_{it} \partial \Phi(I_{it}, K_{it}) / \partial I_{it} + K_{it} \partial \Phi(I_{it}, K_{it}) / \partial K_{it}$ . A standard quadratic functional form is given by  $\Phi(I_{it}, K_{it}) = (a/2) (I_{it}/K_{it})^2 K_{it}$ , in which  $a > 0$ .

Firms can finance investment with one-period debt. At the beginning of time  $t$ , firm  $i$  can issue an amount of debt,  $B_{it+1}$ , which must be repaid at the beginning of period  $t+1$ . The gross corporate bond return on  $B_{it}$ ,  $r_{it}^B$ , can vary across firms and over time. Taxable corporate profits equal operating profits less capital depreciation, adjustment costs, and interest expenses,  $\Pi(K_{it}, X_{it}) - \delta_{it} K_{it} - \Phi(I_{it}, K_{it}) - (r_{it}^B - 1) B_{it}$ , in which adjustment costs are expensed. Let  $\tau_t$  denote the corporate tax rate at time  $t$ . The payout of firm  $i$  equals

$$D_{it} \equiv (1 - \tau_t) [\Pi(K_{it}, X_{it}) - \Phi(I_{it}, K_{it})] - I_{it} + B_{it+1} - r_{it}^B B_{it} + \tau_t \delta_{it} K_{it} + \tau_t (r_{it}^B - 1) B_{it}, \quad (23)$$

in which  $\tau_t \delta_{it} K_{it}$  is the depreciation tax shield, and  $\tau_t (r_{it}^B - 1) B_{it}$  is the interest tax shield.

Let  $M_{t+1}$  be the stochastic discount factor from  $t$  to  $t+1$ , which is correlated with the aggregate component of  $X_{it+1}$ . The cum-dividend market equity can be formulated as

$$V_{it} \equiv \max_{\{I_{it+s}, K_{it+s+1}, B_{it+s+1}\}_{s=0}^{\infty}} E_t \left[ \sum_{s=0}^{\infty} M_{t+s} D_{it+s} \right], \quad (24)$$

subject to a transversality condition that prevents firms from borrowing an infinite amount to distribute to shareholders,  $\lim_{T \rightarrow \infty} E_t [M_{t+T} B_{it+T+1}] = 0$ . The equity value-maximization implies that  $E_t [M_{t+1} r_{it+1}^I] = 1$ , in which  $r_{it+1}^I$  is the investment return, defined as

$$r_{it+1}^I \equiv \frac{(1 - \tau_{t+1}) \left[ \kappa \frac{Y_{it+1}}{K_{it+1}} + \frac{a}{2} \left( \frac{I_{it+1}}{K_{it+1}} \right)^2 \right] + \tau_{t+1} \delta_{it+1} + (1 - \delta_{it+1}) \left[ 1 + (1 - \tau_{t+1}) a \left( \frac{I_{it+1}}{K_{it+1}} \right) \right]}{1 + (1 - \tau_t) a \left( \frac{I_{it}}{K_{it}} \right)}. \quad (25)$$

The investment return is the ratio of the marginal benefits of investment at time  $t+1$  to the marginal costs of investment at  $t$ . In its numerator,  $(1 - \tau_{t+1}) \kappa Y_{it+1} / K_{it+1}$  is the marginal after-tax profits produced by an additional unit of capital,  $(1 - \tau_{t+1}) (a/2) (I_{it+1} / K_{it+1})^2$  is the marginal after-tax reduction in adjustment costs,  $\tau_{t+1} \delta_{it+1}$  is the marginal depreciation tax shield, and the last term in the numerator is the marginal continuation value of the extra unit of capital net of depreciation. Also, the first term in brackets in the numerator divided by the denominator is analogous to a dividend yield. The second term in brackets in the numerator divided by the denominator is analogous to a capital gain because this ratio is the growth of marginal  $q$ .

Let the after-tax corporate bond return be  $r_{it+1}^{Ba} \equiv r_{it+1}^B - (r_{it+1}^B - 1) \tau_{t+1}$ , then  $E_t [M_{t+1} r_{it+1}^{Ba}] = 1$ . Let  $P_{it} \equiv V_{it} - D_{it}$  be the ex-dividend equity value,  $r_{it+1}^S \equiv (P_{it+1} + D_{it+1}) / P_{it}$  be the stock return, and  $w_{it} \equiv B_{it+1} / (P_{it} + B_{it+1})$  be the market leverage. Then the investment return is the weighted average of the stock return and the after-tax corporate bond return:

$$r_{it+1}^I = w_{it} r_{it+1}^{Ba} + (1 - w_{it}) r_{it+1}^S. \quad (26)$$

Equation (26) is exactly the weighted average cost of capital in corporate finance.

Solving for the stock return  $r_{it+1}^S$  from equation (26) gives *the multiperiod investment CAPM*:

$$r_{it+1}^S = r_{it+1}^{Iw} \equiv \frac{r_{it+1}^I - w_{it} r_{it+1}^{Ba}}{1 - w_{it}}, \quad (27)$$

in which  $r_{it+1}^{Iw}$  is the levered investment return. Together, equations (25) and (26) imply that the weighted average cost of capital equals the ratio of the next-period marginal benefits of investment

divided by the current-period marginal costs of investment. This first principle of investment provides the microfoundation for the weighted average cost of capital approach to capital budgeting in corporate finance. Intuitively, firms will keep investing until the costs of doing so, which rise with investment, equal the present value of additional investment, which is the next-period marginal benefits of investment discounted by the weighted average cost of capital.

### 3.1.1 Expected Stock Returns

The multiperiod investment CAPM implies an ex-ante restriction that the expected stock return equals the expected levered investment return across all testing assets:

$$E [r_{it+1}^S - r_{it+1}^{Iw}] = 0. \quad (28)$$

The model error from the moment condition defines the investment CAPM alpha:

$$\alpha_q^i \equiv E_T [r_{it+1}^S - r_{it+1}^{Iw}], \quad (29)$$

in which  $E_T[\cdot]$  is the sample mean of the series in brackets.

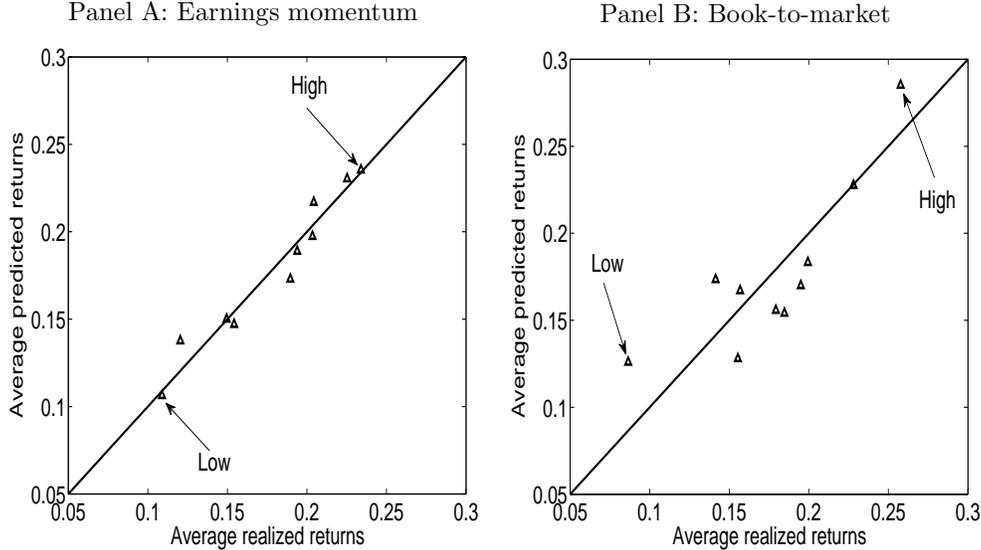
Liu, Whited, and Zhang (2009) estimate the parameters  $a$  and  $\kappa$  via one-stage generalized method of moments (GMM) at the portfolio level. Grouping stocks into portfolios enlarges the average return spread across extreme portfolios, and makes it more significant by diversifying away idiosyncratic volatilities at the stock level. Accordingly, the power of asset pricing tests increases (Black, Jensen, and Scholes 1972). The portfolio approach also has the advantage that portfolio investment data are smooth, whereas the investment data at the firm level can be lumpy.

Figure 2 shows that the scatter points of average predicted stock returns,  $\overline{r_{it+1}^{Iw}}$ , against average realized stock returns,  $\overline{r_{it+1}^S}$ , are largely aligned with the 45-degree line across the earnings surprises (Panel A) and book-to-market deciles (Panel B). The alpha of the high-minus-low earnings surprises decile is only  $-0.4\%$  per annum, and the alpha of the high-minus-low book-to-market decile  $1.2\%$ , both of which are within one standard error from zero. The mean absolute alpha is  $0.7\%$  across the earnings surprises deciles, and  $2.3\%$  across the book-to-market deciles.

However, the point estimates in Liu, Whited, and Zhang (2009) reveal a weakness. The adjustment costs parameter,  $a$ , is estimated to be  $7.7$  across the earnings surprises deciles, but  $22.3$  across the book-to-market deciles. The capital's share parameter,  $\kappa$ , is  $0.3$  across the former, but  $0.5$  across the latter deciles. Ideally, if a model is well specified, the estimates of these technological pa-

**Figure 2 : Average Predicted Stock Returns versus Average Realized Stock Returns, the Multiperiod Investment CAPM**

Source: Liu, Whited and Zhang (2009). The figure plots average returns predicted from the multiperiod investment CAPM against average realized returns across the deciles formed on earnings surprises (Panel A) and book-to-market (Panel B). “High” is the high decile, and “Low” the low decile.



rameters should be invariant across different testing assets. Clearly, the first stab at implementing the multiperiod investment CAPM via structural estimation is far from perfect.

Liu and Zhang (2014) use the model to understand price momentum. Momentum deciles are rebalanced monthly, but accounting variables in Compustat are annual. To handle this difficulty, Liu and Zhang design a more polished timing alignment procedure in that monthly levered investment returns of a momentum decile are constructed from its annual accounting variables to match with the decile’s monthly stock returns. The investment CAPM performs well across the price momentum deciles. The scatter points of average predicted versus realized returns are again largely aligned with the 45-degree line. The winner-minus-loser decile has a small alpha of 0.4% per annum, which is only about 2.65% of its average return of 15.09% (equal-weighted). Also, the mean absolute alpha is 0.83%, which is small relative to the average decile return of 12.4%.

Armed with the point estimates from matching only average momentum profits, Liu and Zhang (2014) examine the reversal of momentum in long horizons as diagnostics on the model. In the event-time window covering 36 months after the portfolio formation, average winner-minus-loser

returns in the data start at 19.98% per annum in the first month in the holding period, fall to 13.15% in month six, converge largely to zero in month ten, and turn negative afterward. The investment CAPM goes a long way toward explaining this reversal. Levered investment returns for the winner-minus-loser decile start at 18.21% per annum in the first month, fall to 10.73% in month six, and further to 2.87% in month twelve. The predicted price momentum profits converge largely to zero in month fifteen, and turn negative afterward. The expected growth plays a key role in capturing the reversal of price momentum. The investment-to-capital growth spread starts at 39.45% in month one, weakens to 23.06% in month six, converges to zero in month thirteen, and turns negative afterward. In contrast, the sales-to-capital spread is much more persistent than momentum profits.

### 3.1.2 Equity Valuation

Equity valuation is immensely important in theory and practice. In academia, a vast accounting literature has built on the dividend discounting and residual income models for equity valuation (Ohlson 1995). Widely practiced in the financial services industry, valuation is at the core of standard business school curriculum with many textbook treatments (Penman 2013; Koller, Goedhart, and Wessles 2015). Traditional valuation methods calculate the present value of future dividends, working from the perspective of investors' demand of risky equities.

Belo, Xue, and Zhang (2013) approach the valuation question from the perspective of managers' supply of assets. The idea is that managers, if behaving optimally, will adjust the supply of assets to their changes in the market value. Managers will invest until the marginal benefits of one extra unit of assets (marginal  $q$ , which is the present value of all the future dividends generated by this extra unit) equal the marginal costs of supplying this extra unit. With a specified capital adjustment technology, the marginal costs of investment (marginal  $q$ ) can be inferred from investment. With constant returns to scale, the inferred marginal  $q$  provides the value for a firm's entire capital assets.

Working within the multiperiod framework outlined in Section 3.1, Belo, Xue, and Zhang (2013) allow the marginal adjustment costs of investment to be nonlinear:

$$\Phi(I_{it}, K_{it}) = \frac{1}{\nu} \left( \eta \frac{I_{it}}{K_{it}} \right)^\nu K_{it}, \quad (30)$$

in which  $\eta > 0$  is the slope, and  $\nu > 0$  the curvature parameter. The first principle of investment

implies that the market value of the firm is given by:

$$P_{it} + B_{it+1} = \left[ 1 + (1 - \tau_t)\eta^\nu \left( \frac{I_{it}}{K_{it}} \right)^{\nu-1} \right] K_{it+1}. \quad (31)$$

In addition, the investment Euler equation is given by:

$$1 + (1 - \tau_t)\eta^\nu \left( \frac{I_{it}}{K_{it}} \right)^{\nu-1} = E_t \left[ M_{t+1} \left[ \begin{array}{l} (1 - \tau_{t+1}) \left[ \kappa \frac{Y_{it+1}}{K_{it+1}} + \frac{\nu-1}{\nu} \left( \eta \frac{I_{it+1}}{K_{it+1}} \right)^\nu \right] + \delta_{it+1}\tau_{t+1} \\ + (1 - \delta_{it+1}) \left[ 1 + (1 - \tau_{t+1})\eta^\nu \left( \frac{I_{it+1}}{K_{it+1}} \right)^{\nu-1} \right] \end{array} \right] \right]. \quad (32)$$

Belo, Xue, and Zhang (2013) measure Tobin's  $q$  in the data as  $q_{it} = (P_{it} + B_{it+1})/A_{it}$ , in which  $A_{it}$  is total assets, and test whether the average  $q$  in the data equals that implied in the model:

$$E \left[ q_{it} - \left[ 1 + (1 - \tau_t)\eta^\nu \left( \frac{I_{it}}{K_{it}} \right)^{\nu-1} \right] \frac{K_{it+1}}{A_{it}} \right] = 0. \quad (33)$$

However, this equation allows assets to be potentially misvalued, but forces managers to align investment with misvalued  $q$ . To alleviate this concern, Belo et al. estimate the valuation moment jointly with a (scaled) investment Euler equation moment, specified as:

$$E \left[ \left( \frac{\begin{array}{l} 1 + (1 - \tau_t)\eta^\nu \left( \frac{I_{it}}{K_{it}} \right)^{\nu-1} - \\ \left[ (1 - \tau_{t+1}) \left[ \kappa \frac{Y_{it+1}}{K_{it+1}} + \frac{\nu-1}{\nu} \left( \eta \frac{I_{it+1}}{K_{it+1}} \right)^\nu \right] + \delta_{it+1}\tau_{t+1} \\ + (1 - \delta_{it+1}) \left[ 1 + (1 - \tau_{t+1})\eta^\nu \left( \frac{I_{it+1}}{K_{it+1}} \right)^{\nu-1} \right] \end{array} \right)}{w_{it}r_{it+1}^{Ba} + (1-w_{it})r_{it+1}^S} \right) \frac{K_{it+1}}{A_{it}} \right] = 0, \quad (34)$$

in which  $M_{t+1}$  is specified as the inverse of the weighted average cost of capital.

Empirically, the scatter points of average predicted  $q$  versus average realized  $q$  across the Tobin's  $q$  deciles are largely aligned with the 45-degree line. The high-minus-low error from the valuation moment is 0.27, which is only 6% of the high-minus-low spread in Tobin's  $q$ , 4.5. The high-minus-low Euler equation error is slightly larger at 0.32. The mean absolute valuation error is 0.08, which is only 5.13% of the average  $q$  across the deciles, 1.56. The point estimates are sensible, 4.1 for the slope parameter,  $\eta$ , and 4.09 for the curvature parameter,  $\nu$ , which is also reliably different from two, rejecting the quadratic adjustment costs function. These estimates imply an average adjustment costs-to-sales ratio of 3.95%, which is relatively low.

## 3.2 Notes

### 3.2.1 The Neoclassical $q$ -Theory of Investment

The neoclassical  $q$ -theory of investment is originally developed to explain real investment behavior. This literature is vast (see Chirinko 1993 for a review on aggregate investment, and Bond and Van Reenen 2007 on firm investment). Introduced by Brainard and Tobin (1968) and Tobin (1969), the basic idea is to relate investment to average  $q$ , which is the ratio of the market value of a firm to the replacement costs of its capital stock. Jorgenson (1963), Lucas and Prescott (1971), Mussa (1977), and Abel (1983) provide the neoclassical formulation of  $q$ -theory, and derive the relation between investment and marginal  $q$ . Hayashi (1982) shows that under constant returns to scale, marginal  $q$  equals average  $q$ . This proposition allows empirical studies to replace the unobservable marginal  $q$  with the observable average  $q$ . Abel and Eberly (1994) extend  $q$ -theory to incorporate costly reversibility and flow fixed costs of investment.

Empirically,  $q$ -theory's performance has been disappointing in describing investment behavior. In investment- $q$  regressions, for instance, the slope coefficients on cash flow are typically large and significant for firms that are financially constrained, even after controlling for  $q$  (Fazzari, Hubbard, and Petersen 1988). In investment Euler equation tests, the investment model is rejected in subsamples that consist of financially constrained firms (Whited 1992). The empirical performance deteriorates further at the micro level across firms or plants, in which fixed costs of investment and other nonconvexities play an important role (Doms and Dunne 1998; Cooper and Haltiwanger 2006).

In contrast, the investment model is more successful in cross-sectional asset pricing. The key difference is that asset pricing tests are conducted at the portfolio level. As noted, forming portfolios makes average return spreads more reliable, increasing the test power. Because of aggregation, investment data at the portfolio level are smooth relative to firm-level or plant-level investment data. Thomas (2002) and Veracierto (2002), for instance, show that aggregation largely eliminates the impact of lumpy investment on aggregate dynamics in equilibrium business cycle models. Cooper and Haltiwanger (2006) also show that the quadratic model captures most of the time series variation in aggregate investment simulated from a nonconvex adjustment costs model estimated on plant-level data. Finally, financial constraints are not a strong predictor of cross-sectional expected returns in the data (Lamont, Polk, and Saá-Requejo 2001). It remains to be seen, however, whether financial constraints affect equity valuation, especially at the firm level.

### 3.2.2 Investment-Based Asset Pricing Tests

**Aggregate Stock Market** Cochrane (1991) is the first to apply  $q$ -theory to study asset prices. Cochrane derives the investment return in equation (25) in a simplified setting without debt or taxes, and argues that the investment return should equal the return to owning capital, which is in turn the stock return. Restoy and Rockinger (1994) prove that the equality between the investment and stock returns holds under more general conditions, and that the equality is in effect a mathematical restatement of that between marginal  $q$  and average  $q$  under the Hayashi (1982) conditions. Empirically, Cochrane shows that with reasonable parameters, the mean of the aggregate investment return matches the mean of the aggregate stock market return, meaning that the equity premium is not a puzzle in  $q$ -theory. However, the investment return volatility is only about one half of the stock market volatility. The correlation between investment and stock returns is as high as 40%. Finally, investment-to-capital predicts the stock market return with a significantly negative slope.

Lettau and Ludvigson (2002) emphasize that time-varying aggregate risk premiums have implications not only for aggregate investment today, but also for future investment over long horizons. Predictive variables for market excess returns over long horizons should also forecast long horizon fluctuations in the growth of marginal  $q$  and investment. Evidence from long horizon regressions of aggregate investment growth on a variety of risk premium predictors confirms this prediction.

Merz and Yashiv (2007) introduce labor into the investment model, and fit the resulting valuation equation on the aggregate U.S. data. A convex adjustment costs function with nonlinear marginal costs is able to account for the valuation data, including the mean and volatility of the market equity-to-output ratio. The valuation ratio predicted from the estimated model also traces closely the valuation ratio in the data, including the stock price runup in the late 1990s. Merz and Yashiv attribute the model's good performance to the use of gross flows data for both investment and hiring, the joint estimation of hiring and investment including their interaction, and the nonlinear marginal costs in the adjustment costs function.

Jermann (2010) examines the determinants of the equity premium implied by firms' first-order conditions. Assuming two states of the world and two capital inputs (structures and equipment), Jermann rewrites  $E_t[M_{t+1}r_{it+1}^I] = 1$  as a system of two equations consisting of the structures and equipment investment returns. The system can then be solved for the two state prices, from which the interest rate can be calculated. With reasonable parameter values, the model fits the first and

second moments of the market return and interest rate in the data, and connects these moments with technologies. However, while this method derives state prices from the production side, it seems empirically infeasible to extend it to more than two states of the world.

**Cross-sectional Tests** Cochrane (1996) conducts cross-sectional tests on the consumption CAPM,  $E_t[M_{t+1}r_{it+1}^S] = 1$ , in which  $r_{it+1}^S$  is the stock returns across the size deciles, and  $M_{t+1}$  is specified as a linear function of two investment returns constructed per equation (25) on the gross private domestic nonresidential and residential investment data. Gomes, Yaron, and Zhang (2006) study the impact of financial constraints, which are modeled as a dividend nonnegativity constraint, on the cross section of returns. Following Cochrane, Gomes et al. specify the stochastic discount factor as a linear function of the aggregate investment return subject to financial constraints. With the 25 size and book-to-market portfolios as testing assets, financial constraints provide a common factor, but their impact seems more important in good times, when investment are high.

Instead of Cochrane's (1996) test framework, Whited and Wu (2006) estimate the investment Euler equation augmented with financial constraints at the firm level via GMM. The shadow price of external funds is specified as a linear function of several firm characteristics. Using the estimated shadow price as a financial constraints index, Whited and Wu show that there exists a financial constraints factor, but that its average return is insignificant, albeit positive.

Balvers and Huang (2007) formulate a productivity-based asset pricing model, in which the Solow residual is the only priced factor, and capital is the only conditioning variable. Their cross-sectional tests show that value and growth stocks have similar average productivity risk. However, growth stocks have their highest productivity risk, when the productivity risk premium is low or even negative in good times, giving rise to lower average returns than value stocks. Intuitively, the market value of growth firms varies more drastically with productivity shocks in booms, when capital is abundant, as positive shocks are propagated by low investment costs.

Cochrane (1993) emphasizes that standard production technologies restrict firms from adjusting their output across states of nature. Cochrane formulates a flexible state-contingent production technology to overcome this limitation, and derives a stochastic discount factor as firms' marginal rate of transformation. Belo (2010) estimates this model by identifying the marginal rate of transformation from price and output data in the durable and nondurable goods sectors. Belo's estimates suggest that firms' ability to transform output across states of nature seems high.

Vitorino (2014) quantifies the impact of advertising on expected returns and equity valuation. Interpreting advertising expenses as investment in brand capital, Vitorino extends the expected return test in equation (28) and the valuation test in equation (33) to a setting with two capital inputs, physical capital and brand capital. The model matches well the average stock returns and Tobin's  $q$  across portfolios formed on advertising expenses. Also, brand capital accounts for a substantial fraction, about 23%, of firms' market equity, and this fraction varies greatly across industries.

Cooper and Priestley (2015) apply the investment CAPM to study the expected returns and valuation of private firms, which form a substantial part of the U.S. economy. Whereas estimating costs of capital from the consumption CAPM requires historical stock returns data that are not feasible for private firms, the investment CAPM estimates costs of capital directly from economic fundamentals. As such, the investment CAPM uniquely befits private firms. Cooper and Priestley examine the cross section of ten industry portfolios formed on the fraction of the sales of public firms in a given industry to its total industry sales. The bottom two deciles, which consist of only private firms, are identified as private industries, and the top decile as public industries.

Cooper and Priestley (2015) report three key findings. First, characteristics such as investment and profitability play a similar role in driving the cross-sectional variation of investment returns across both private and public firms. This evidence lends support to the investment CAPM, but casts doubt on mispricing as an explanation for the role of these characteristics. The key identification is that private firms do not have stock prices to overreact or underreact to, and should be less affected by mispricing, if at all. Second, a four-factor model, resembling the  $q$ -factor model, constructed on both private and public firms, performs well in describing the cross section of investment returns, and provides the cost-of-capital estimates for private firms. Finally, private firms have expected returns and valuation ratios that are not far from those of public firms.

## 4 The Big Picture

After reviewing what the investment CAPM can do, I take a step back, and explain the big picture. I provide a historical perspective of the investment CAPM in Section 4.1, and clarify its relation with the consumption CAPM in Section 4.2 and with behavioral finance in Section 4.3.

## 4.1 A Historical Perspective

### 4.1.1 Origin

The origin of the investment CAPM can be traced to Böhm-Bawert (1891) and Fisher (1930). Böhm-Bawert provides three reasons why the rate of interest is positive. First, the marginal utility of income falls over time because of more resources in the future. Second, due to psychological reasons, consumers tend to underestimate future needs. Both reasons are consumption-based determinants of the equilibrium interest rate. Böhm-Bawert's third reason is "roundabout" production:

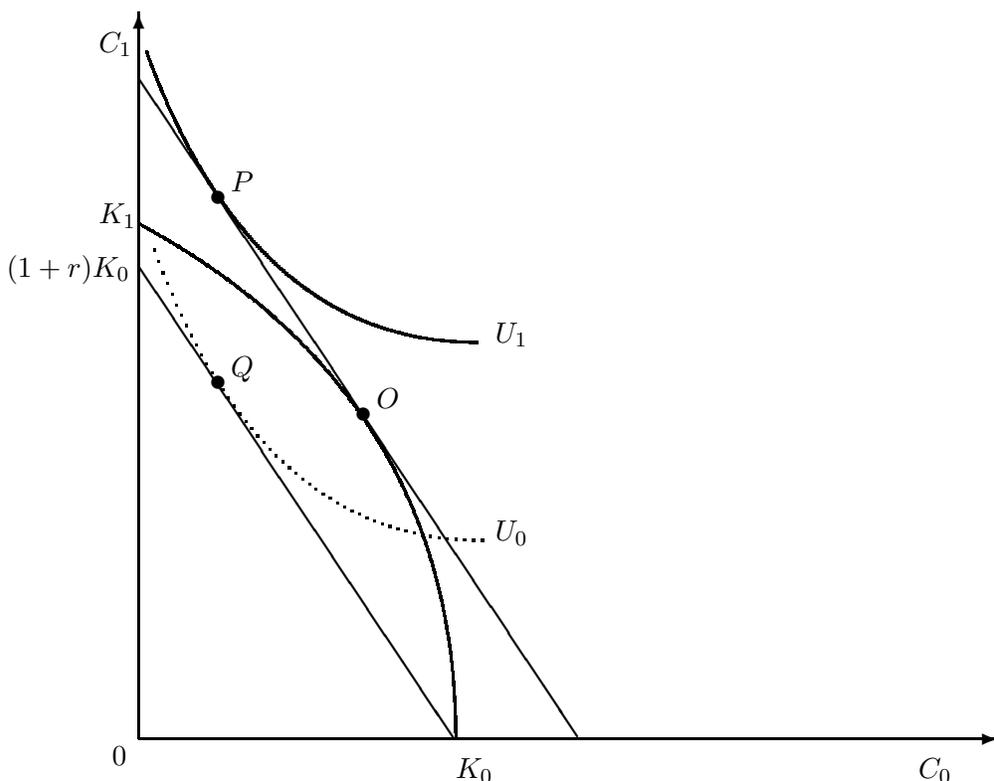
It is an elementary fact of experience that methods of production which take time are more productive. That is to say, given the same quantity of productive instruments, the lengthier the productive method employed the greater the quantity of products that can be obtained (p. 260)... Command over a sum of present consumption goods provides us with the means of subsistence during the current economic period. This leaves the means of production, which we may have at our disposal during this period (Labour, Uses of Land, Capital), free for the technically more productive service of the future, and gives us the more abundant product attainable by them in longer methods of production (p. 271).

Because production per worker rises with the length of the period of production, some counterbalancing force must exist to cause producers to choose a production period of finite length. Böhm-Bawert (1891) argues that this counterbalancing force is the positive rate of interest. This mechanism implies that the lower the interest rate, the longer will be the optimal period of production, and the more capital will be tied up in the production process. Conversely, the higher the interest rate, the shorter will be the optimal period of production, and the less capital will be tied up in the production process. This effect is exactly the negative relation between real investment and the discount rate (Figure 1), albeit without uncertainty.

Fisher (1930) studies the economic determinants of the real interest rate by constructing the first general equilibrium model with both intertemporal exchange and production. His model also shows the Fisher Separation Theorem, which justifies the maximization of the present value as the objective of the firm, without any direct dependence on shareholder preferences. Figure 3, which is adapted from Chart 38 in Fisher (p. 271), shows the key insights. The horizontal axis labeled  $C_0$  represents consumption in date 0, and the vertical axis  $C_1$  represents consumption in date 1. Endowed with an amount of resources,  $K_0$ , in date 0, the agent's problem is to choose an optimal time pattern of consumption. The agent's preferences are represented by indifference curves, such as  $U_0$  and  $U_1$ . There are two available ways to transfer resources between dates 0 and 1. The "market opportunity" represented by the two straight lines in the figure is through borrowing and lending at the interest rate,  $r$ . The "investment opportunity" represented by the curve  $K_0OK_1$

is through a real production technology. The curve is concave to the origin due to diminishing returns to scale. Without the production technology, the optimum is  $Q$ , which is the tangent point between the market line initiated at  $K_0$  and the indifference curve  $U_0$ .

**Figure 3. The Fisherian Equilibrium**



With the production technology, the agent solves the optimum in two separate steps. First, the production optimum is achieved at  $O$ , which is the tangent point between its associated market line and the production frontier. Starting from  $K_0$ , the agent keeps moving up along the production curve  $K_0OK_1$ , until the marginal rate of transformation,  $\Delta K_1/\Delta K_0 - 1$ , in which  $\Delta K_0$  is the amount of input in date 0, and  $\Delta K_1$  is output in date 1, is equal to the interest rate,  $r$ . This equality is achieved at the tangent point  $O$ . Prior to this point,  $\Delta K_1/\Delta K_0 - 1 > r$ , meaning that it is optimal to invest more. Beyond this point,  $\Delta K_1/\Delta K_0 - 1 < r$ , meaning that it is suboptimal to keep investing. Second, the consumption optimum is achieved at point  $P$ , through borrowing and lending along the market line associated with  $O$ . An important implication of the Fisher Separation Theorem is that the equilibrium interest rate can be “determined” purely from the production technology, without any direct dependence on investor preferences.

In his own words, Fisher (1930) describes the interaction and equivalence between the consumption-based (impatience) and investment-based determinants of the interest rate as follows:

Our outer opportunities urge us to postpone present income—to shift it toward the future, because it will expand in the process. Impatience is impatience to spend, while opportunity is opportunity to invest. The more we invest and postpone our gratification, the lower the investment opportunity rate becomes, but the greater the impatience rate; the more we spend and hasten our gratification, the lower the impatience rate becomes but the higher the opportunity rate.

If the pendulum swings too far toward the investment extreme and away from the spending extreme, it is brought back by the strengthening of impatience and the weakening of investment opportunity. Impatience is strengthened by growing wants, and opportunity is weakened because of diminishing returns. If the pendulum swings too far toward the spending extreme and away from the investment extreme it is brought back by the weakening of impatience and the strengthening of opportunity for reasons opposite to those stated above.

Between these two extremes lies the equilibrium point which clears the market, and clears it at a rate of interest registering (in a perfect market) all impatience rates and all opportunity rates (p. 177).

Hirshleifer (1958) uses the Fisherian approach to study optimal investment decisions of firms, with and without certain capital markets imperfections. Hirshleifer (1965, 1966, 1970) extends the Fisherian equilibrium into a world with uncertainty, using the state-preference approach of Arrow (1964) and Debreu (1959). Diamond (1967) and Hirshleifer (1970) also contain an early formation of state-contingent production, a theme echoed later by Cochrane (1993) and Belo (2010).

Jorgenson (1963) formulates a model of investment demand with capital as a factor of production, and introduces the concept of the user cost of capital as the rental price of capital services. The user cost of capital contains three components, including the interest rate, the depreciation rate, and the rate of capital loss on investment goods, adjusted for the taxation of capital income. Hall and Jorgenson (1967) analyze the impact of alternative tax policies on the user cost of capital.

In a justly immortal contribution, Modigliani and Miller (1958) ask the cost of capital question: What is the “cost of capital” to a firm in a world in which funds are used to acquire assets whose yields are uncertain; and in which capital can be obtained by many different media, ranging from pure debt instruments, representing money-fixed claims, to pure equity issues, giving holders only the right to a pro-rata share in the uncertain venture (p. 261)?

Modigliani and Miller’s Proposition I says that in perfect capital markets, the value of a firm equals the market value of the total cash flows produced by its assets, and is unaffected by its capital structure. Proposition II says that the cost of equity equals the cost of unlevered equity plus a risk premium, which equals the market debt-to-equity ratio times the spread between the cost of unlevered equity and the cost of debt. Both propositions are extensively discussed in modern corporate finance textbooks, including Ross, Westerfield, and Jaffe (2008, p. 428–447), Berk and DeMarzo (2009, p. 455–466), and Brealey, Myers, and Allen (2011, p. 418–434).

However, it is the relatively overlooked Proposition III in Modigliani and Miller (1958) that contains arguably the earliest discussion of the investment CAPM:

*Proposition III.* If a firm in class  $k$  is acting in the best interest of the stockholders at the time of the decision, it will exploit an investment opportunity if and only if the rate of return on the investment, say  $\rho^*$ , is as large as or larger than  $\rho_k$ . That is, *the cut-off point for investment in the firm will in all cases be  $\rho_k$  and will be completely unaffected by the type of security used to finance the investment.* Equivalently, we may say that regardless of the financing used, the marginal cost of capital to a firm is equal to the average cost of capital, which is in turn equal to the capitalization rate for an unlevered stream in the class to which the firm belongs (p. 288, original emphasis).

Cochrane (1991) is the first in the modern era to use the investment model to study asset prices:

The logic of the production-based model is exactly analogous [to that of the consumption-based model]. It ties asset returns to marginal rates of *transformation*, which are inferred from data on investment (and potentially, output and other production variables) through a *production* function. It is derived from the *producer's* first order conditions for optimal intertemporal *investment* demand. Its testable content is a restriction on the joint stochastic process of *investment* (and/or other production variables) and asset returns. This restriction can also be interpreted in two ways. If we fix the return process, it is a version of the q theory of investment. If we fix the investment process, it is a production-based asset pricing model. For example, the production-based asset pricing model can make statements like “expected returns are high because (a function of) *investment* growth is high” (p. 210, original emphasis).

Alas, when applying the investment model to the cross section, Cochrane (1996) retreats back to the consumption CAPM, stipulates the stochastic discount factor as a linear function of the residential and nonresidential investment returns, and performs GMM tests per Hansen and Singleton (1982).

#### 4.1.2 Oblivion

Modern asset pricing is thoroughly dominated by the consumption CAPM. The term “the investment CAPM” has not even appeared in the prior literature. The glorious history of asset pricing is well known (Bernstein 1992; Rubinstein 2006). I only wish to shed light on how classic asset pricing studies have gone down the path of ignoring the supply (investment) side altogether, despite it receiving a symmetric treatment as the demand (consumption) side in Fisher (1930). The crux seems to be uncertainty. Fisher’s general equilibrium model assumes certainty and only provides a theory of the interest rate, as opposed to a theory of risky asset prices. In hindsight, thanks to Arrow (1964) and Debreu (1959), it is clear that asset pricing theory is nothing but the standard price theory in microeconomics extended to uncertainty and over time.

However, predating the Arrow-Debreu framework, Markowitz (1952, 1959) formulates the mean variance model of portfolio selection (see also Roy 1952). The model postulates that investors should maximize the expected portfolio return, while minimizing portfolio variance. Treynor (1962),

Sharpe (1964), Lintner (1965), and Mossin (1966) ask what would happen in equilibrium if investors all follow Markowitz's decision rule, leading to the derivation of the CAPM. Merton (1973) extends the one-period CAPM to the intertemporal CAPM, in which state variables related to shifts in the investment opportunity set over time are also risk factors (see also Long 1974). The classic textbooks by Fama and Miller (1972) and Fama (1976) rely almost exclusively on the mean variance model to deal with uncertainty. While acknowledging Hirshleifer's (1965, 1966) applications of the Arrow-Debreu approach, Fama and Miller view the mean variance model as much more operational empirically. This operational advantage no longer exists, thanks to the methodological contributions in Hansen (1982) and Hansen and Singleton (1982).

Recasting equilibrium asset pricing within the Arrow-Debreu framework, Rubinstein (1976), Lucas (1978), and Breeden (1979) derive the consumption CAPM from the first principle of the optimal consumption-portfolio choice problem of a representative consumer. Hansen and Singleton (1982) and Breeden, Gibbons, and Litzenberger (1989) provide early tests of the consumption CAPM. Ludvigson (2013) offers a recent review of this literature.

Since consumers and producers are two different agents integral to the Arrow-Debreu model, early asset pricing theorists justify their modeling choice as reducing a messy general equilibrium problem into a tractable consumption-based partial equilibrium problem. Merton (1973) writes:

Since movements from equilibrium to equilibrium through time involve both price and quantity adjustment, a complete analysis would require a description of both the rate of return and change in asset value dynamics. To do so would require a specification of firm behavior in determining the supply of shares, which in turn would require knowledge of the real asset structure (i.e., technology; whether capital is "putty" or "clay"; etc.). In particular, the current returns on firms with large amounts (relative to current cash flow) of non-shiftable capital with low rates of depreciation will tend to be strongly affected by shifts in capitalization rates because, in the short run, most of the adjustment to the new equilibrium will be done by prices (p. 871).

Since the present paper examines only investor behavior to derive the demands for assets and the relative yield requirements in equilibrium, only the rate of return dynamics will be examined explicitly. Hence, certain variable taken as exogenous in the model, would be endogenous to a full-equilibrium system (p. 871).

Similarly, Breeden (1979) also defends his abstraction from the production side:

[It] is not necessary to explicitly examine firms' production decisions and the supply of asset shares, provided that the assumptions made are consistent with optimal behavior of firms in a general equilibrium model. To be consistent with general equilibrium, prices must be recognized to be endogenously determined through the equilibrium of supply and demand (p. 269).

Although general equilibrium asset pricing models with production have appeared sporadically (Brock 1982; Cox, Ingersoll, and Ross 1985), the consumption CAPM has firmly dominated modern asset pricing. The standard MBA textbook by Bodie, Kane, and Marcus (2014) uses the classic

CAPM as the organizing framework. Standard Ph.D. textbooks, such as Huang and Litzenberger (1988), Ingersoll (1988), Duffie (2001), Cochrane (2005), Back (2010), and Danthine and Donaldson (2015), are all based on the consumption CAPM. In particular, Cochrane writes:

Asset pricing theory all stems from one simple concept, presented in the first page of the first chapter of this book: price equals expected discounted payoff. The rest is elaboration, special cases, and a closet full of tricks that make the central equations useful for one or another application (p. xiii).

*All* asset pricing models amount to alternative ways of connecting the stochastic discount factor to data (p. 7, original emphasis).

In hindsight, it seems clear that Cochrane meant that consumption-based asset pricing is all about the stochastic discount factor. Alas, his writings are sometimes misinterpreted as saying that only the consumption CAPM is asset pricing, but the investment CAPM is not.

### 4.1.3 Reincarnation

The consumption CAPM is an empirical failure. Starting from Ball and Brown (1968), a massive literature on asset pricing anomalies has documented pervasive evidence that rejects the consumption CAPM, casting doubt on its premise of efficient markets. Fama (1998), Barberis and Thaler (2003), and Richardson, Tuna, and Wysocki (2010) provide extensive reviews of the anomalies literature. Behavioral finance has risen to use investor sentiment to explain anomalies in inefficient markets. The field today is divided between neoclassicists, who are disturbed by anomalies, but unwilling to give up on microfoundations, and behavioralists, who view anomalies as a resounding rejection of efficient markets in particular and neoclassical economics in general. In the financial services industry, the anomalies literature has formed the scientific foundation for quantitative investment management. Practitioners are overwhelmingly sympathetic to the behavioral view, while dismissing neoclassicists as theoretical purists who have lost touch with the real world.

I view anomalies not as an indictment of efficient markets, but as the empirical difficulties of the consumption CAPM. More important, I view anomalies as suggesting that the consumption paradigm is conceptually incomplete. The genius of Breeden, Lucas, and Rubinstein is to reduce a messy multidimensional general equilibrium asset pricing problem into a tractable consumption-based partial equilibrium problem. Alas, this reduction, by its very construction, is achieved by abstracting from the investment side of general equilibrium. Unfortunately, anomalies are primarily empirical relations between firm characteristics and expected returns. Because firms are ignored, perhaps not surprisingly, the consumption CAPM fails to explain anomalies.

Inspired by Cochrane (1991), I recognize in Zhang (2005a) that the neoclassical  $q$ -theory of investment allows a different reduction of the general equilibrium asset pricing problem, a reduction that is symmetric and neatly complementary to the reduction in the form of the consumption CAPM achieved by the permanent income theory of consumption. The first principle of investment for individual firms implies that the weighted average cost of capital equals the benefits of one extra unit of capital divided by its costs (equations 25 and 26). Intuitively, firms keep investing until the costs of doing so rise to the level of the discounted investment benefits.

The big news for asset pricing is that the investment CAPM expresses the stock return purely as a function of firm characteristics. The investment CAPM does not use any information on consumption and utility functions, thereby achieving a reduction symmetric to the consumption CAPM, which does not use any information on production and investment technologies. This separation is in the spirit of the Fisher Separation Theorem. The investment CAPM implies that the evidence that firm characteristics forecast stock returns does not necessarily mean mispricing. No expectational errors are ever assumed in deriving the investment CAPM.

## 4.2 Complementarity with the Consumption CAPM

The relation between the investment CAPM and the consumption CAPM is complementary.

### 4.2.1 Marshall's Scissors

As asset pricing theory, the investment CAPM is as fundamental as the consumption CAPM. In the Arrow-Debreu economy, the first principle of consumption and the first principle of investment are two key optimality conditions. Consumption betas, expected returns, and firm characteristics are all endogenous variables, which are in turn determined by a system of simultaneous equations in general equilibrium. No causality runs across these endogenous variables. Neither consumption betas nor firm characteristics are causal forces driving expected returns. As such, the investment CAPM is no more and no less “causal” than the consumption CAPM (Lin and Zhang 2013).

Cochrane (2005) also emphasizes no causality from consumption betas to expected returns:

It is enormously tempting to slide into an interpretation that  $E(mx)$  determines  $p$ . We routinely think of betas and factor risk prices—components of  $E(mx)$ —as *determining* expected returns. For example, we routinely say things like ‘the expected return of a stock increased *because* the firm took on riskier projects, thereby increasing its beta.’ But the whole consumption process, discount factor, and factor risk premia change when the production technology changes (p. 41, original emphasis).

History tends to repeat itself. As noted, asset pricing theory is just the standard price theory extended to uncertainty and over time. During the last quarter of the 19th century, economists debated the relative importance of demand and supply in price theory (Landreth and Colander 2002). While David Ricardo and John Stuart Mill had stressed costs of production, William Jevons, Carl Menger, and Léon Walras insisted that value should depend only on marginal utility. Alfred Marshall (1890) used the following analogy to resolve this controversy in his *Principles of Economics*:

We might as reasonably dispute whether it is the upper or under blade of a pair of scissors that cuts a piece of paper, as whether value is governed by utility or costs of production. It is true that when one blade is held still, and the cutting is affected by moving the other, we may say with careless brevity that the cutting is done by the second; but the statement is not strictly accurate, and is to be excused only so long as it claims to be merely a popular and not a strictly scientific account of what happens (Marshall 1890 [1961, 9th edition, p. 348]).

Even Fisher (1930) had to address an earlier criticism that he had neglected production:

Years after *The Rate of Interest* was published, I suggested the more popular term “impatience” in place of “agio” or “time preference.” This catchword has been widely adopted, and, to my surprise, has led to a widespread but false impression that I had overlooked or neglected the productivity or investment opportunity side entirely. It also led many to think that, by using the new word impatience, I meant to claim a new idea. Thus I found myself credited with being the author of “the impatience theory” which I am not, and not credited with being the author of those parts lacking any catchword. It was this misunderstanding which led me, after much search, to adopt the catchword “investment opportunity” as a substitute for the inadequate term “productivity” which had come into such general use (p. viii).

## 4.2.2 The Aggregation Critique

Given that the investment CAPM and the consumption CAPM are parallel but complementary in general equilibrium, why does the former perform better than the latter in the data?

The Roll and Hansen-Richard critiques are well known. Roll (1977) argues that the wealth portfolio is not observable, making the CAPM tests infeasible. Hansen and Richard (1987) argue that the conditioning information of investors is not observable, making the consumption CAPM tests infeasible in a dynamic world. Both critiques seem less important for the investment CAPM. Although the  $q$ -factor model includes the market factor, it is only used to anchor the time-series averages of returns, while the task of accounting for their cross-sectional variation is left to the  $q$ -factors. Also, the structural investment CAPM tests do not need to specify the conditioning information of managers.

The most damaging critique against the consumption CAPM, I believe, is aggregation. Most consumption CAPM studies sidestep the aggregation problem by examining the optimizing behavior of a representative consumer. However, the Sonnenschein-Mantel-Debreu theorem in general equilibrium theory states that the aggregate excess demand function is not restricted by the standard

rationality assumption on individual demands (Sonnenschein 1973; Debreu 1974; Mantel 1974). Individuals can be fully rational, but the aggregate behavior might appear entirely different.

Kirman (1992) raises four objections to the representative consumer. First, individual maximization does not imply collective rationality, and collective maximization does not imply individual rationality. Second, the response of the representative consumer to a parameter change might not be the same as the aggregate response of individuals. Third, it is possible for the representative to exhibit preference orderings that are opposite to all the individuals'. Finally, the aggregate behavior of rational individuals might exhibit complicated dynamics, and imposing these dynamics on one individual can lead to unnatural characteristics of the individual. Kirman concludes that:

[It] is clear that the "representative" agent deserves a decent burial, as an approach to economic analysis that is not only primitive, but fundamentally erroneous (p. 119).

In essence, aggregation in the representative consumer assumes that all 319 million Americans are identical in preferences. In case it is still not disturbing enough, try 1.4 billion Chinese.

On the critical importance of aggregation, Deaton (1992) also writes:

I have come to believe that [representative agent models] are of limited value, and that what we have learned from them is more methodological than substantive. Representative agents have two failings: they know too much, and they live too long. An aggregate of individuals with finite lives, and with limited and heterogeneous information is not likely to behave like the single individual of the textbook. We are likely to learn more about aggregate consumption by looking at microeconomic behavior, and by thinking seriously about aggregation from the bottom up (p. ix).

The main puzzle is not why these representative agent models do not account for the evidence, but why anyone ever thought that they might, given the absurdity of the aggregation assumptions that they require (p. 70).

The aggregation problem cuts both ways, however. On the one hand, without confronting this problem, imposing that asset pricing is all about the stochastic discount factor is self-defeating. On the other hand, the aggregation difficulty also means that the failure of the consumption CAPM might have nothing to say about individual rationality. As noted, the Sonnenschein-Mantel-Debreu theorem says that individual rationality imposes no restrictions on aggregate consumption. A natural corollary is that the consumption CAPM is not testable!

The investment CAPM is relatively immune to the aggregation critique. The consumption CAPM is macroeconomic in nature, and is derived from the first principle of the representative consumer, which is the aggregate stand-in for all the individuals in the economy. In contrast, the investment CAPM is microeconomic in nature, and is derived from the first principle of an individual firm. It seems circuitous, if not clumsy, to use the macroeconomic consumption CAPM to

explain micro-level anomalies, with the strong built-in aggregation assumption. By comparison, the investment CAPM is more straightforward and natural for this purpose.

### 4.2.3 Beta Measurement Errors

In addition to the aggregation critique, another relatively underappreciated problem is beta measurement errors. Fama (1991) describes the difficulties of estimating consumption betas as follows:

Consumption is measured with error, and consumption flows from durables are difficult to impute. The model calls for instantaneous consumption, but the data are monthly, quarterly, and annual aggregates. Finally, ... the elegance of the consumption model (all incentives to hedge uncertainty about consumption and investment opportunities are summarized in consumption  $\beta$ 's) likely means that consumption  $\beta$ 's are difficult to estimate because they vary through time (p. 1599).

These difficulties probably explain why the consumption CAPM often underperforms the classic CAPM in empirical tests. However, the CAPM is also subject to severe measurement errors.

Miller and Scholes (1972) simulate random returns from the CAPM, and find that test results on simulated data are consistent with those from the real data:

We have shown that much of the seeming conflict between these [empirical] results and the almost exactly contrary predictions of the underlying economic theory may simply be artifacts of the testing procedures used. The variable that measures the systematic covariance risk of a particular share is obtained from a first-pass regression of the individual company returns on a market index. Hence it can be regarded at best as an approximation to the perceived systematic risk, subject to the margin of error inevitable in any sampling process, if to nothing else. The presence of such errors of approximation will inevitably weaken the apparent association between mean returns and measured systematic risk in the critical second-pass tests.

More recently, Gomes, Kogan, and Zhang (2003) and Carlson, Fisher, and Giammarino (2004) show how size and book-to-market can dominate empirically estimated betas in cross-sectional regressions on data simulated from economies with the conditional CAPM structure. Li, Livdan, and Zhang (2009) report similar results on real investment and equity issues in cross-sectional regressions. Bai, Hou, Kung, Li, and Zhang (2017) show how the CAPM can fail to explain the value premium in an economy with rare disasters, and how measurement errors in estimated betas can completely flatten the risk-return relation in an economy without any mispricing. These simulation results do not mean that the CAPM is alive and well, but do suggest that beta measurement errors might have a lot to do with the empirical difficulties of the CAPM.

Using simulated data from a dynamic investment model as in Zhang (2005b), in which the conditional CAPM holds by construction, Lin and Zhang (2013) show how characteristics can easily dominate covariances in predicting returns in Daniel and Titman's (1997) two-way sorts on covariances and characteristics. The beta estimates from 36-month rolling-window regressions are

averaged betas in the past three years, whereas the true betas in the model are time-varying. The time-lag between the estimated betas and portfolio formation reduces the power of covariances to predict returns. In addition, using latest instruments in estimating conditional betas hardly alleviates the measurement errors. Because the theoretical relations between instruments and betas are typically nonlinear, linear beta models contain large specification errors (Harvey 2001).

Some empiricists tend to downplay theoretical results based on simulated data. It is true that simulated data are not real data. That said, if the estimated betas from standard empirical procedures differ drastically from the true betas in a controlled laboratory, what confidence should one have in beta estimates in the real world that is vastly more complex than our laboratories?

### **4.3 An Efficient Markets Counterrevolution to Behavioral Finance**

Asset pricing anomalies obey standard economic principles.

#### **4.3.1 Mispricing**

It is standard practice to equate anomalies to mispricing in behavioral finance.

Research in experimental psychology suggests that, in violation of Bayes' rule, most people tend to "overreact" to unexpected and dramatic news events. This study of market efficiency investigates whether such behavior affects stock prices. The empirical evidence, based on CRSP monthly return data, is consistent with the overreaction hypothesis. Substantial weak form market inefficiencies are discovered (De Bondt and Thaler 1985, p. 793).

Evidence presented here is consistent with a failure of stock prices to reflect fully the implications of current earnings for future earnings.... Even more surprisingly, the signs and magnitudes of the three-day reactions are related to the autocorrelation structure of earnings, as if stock prices fail to reflect the extent to which each firm's earnings series differs from a seasonal random walk (Bernard and Thomas 1990, p. 305).

The pattern that emerges is that the underperformance is concentrated among relatively young growth companies, especially those going public in the high-volume years of the 1980s. While this pattern does not rule out bad luck being the cause of the underperformance, it is consistent with a scenario of firms going public when investors are irrationally over optimistic about the future potential of certain industries which, following Shiller (1990), I will refer to as the "fad" explanation (Ritter 1991, p. 4).

[It] is possible that the market underreacts to information about their short-term prospects of firms but overreacts to information about their long-term prospects. This is plausible given that the nature of the information available about a firm's short-term prospects, such as earnings forecasts, is different from the nature of the more ambiguous information that is used by investors to assess a firm's longer-term prospects (Jegadeesh and Titman 1993, p. 90).

Investor expectations of future growth appear to have been excessively tied to past growth despite the fact that future growth rates are highly mean reverting. In particular, investors were systematically disappointed (Lakonishok, Shleifer, and Vishny 1994, p. 1575).

The results indicate that earnings performance attributable to the accrual component of earnings exhibits lower persistence than earnings performance attributable to the cash flow component of earnings. The results also indicate that stock prices act as if investors "fixate" on earnings, failing to distinguish fully

between the different properties of the accrual and cash flow components of earnings. Consequently, firms with relatively high (low) levels of accruals experience negative (positive) future abnormal stock returns that are concentrated around future earnings announcements (Sloan 1996, p. 290).

[Managers] have an incentive to put the best possible spin on both their new opportunities as well their overall business when their investment expenditures are especially high because of their need to raise capital as well as to justify their expenditures. If investors fail to appreciate managements' incentives to oversell their firms in these situations, stock returns subsequent to an increase in investment expenditures are likely to be negative. This effect is likely to be especially important for managers who are empire builders, and invest for their own benefits rather than the benefits of the firm's shareholders (Titman, Wei, and Xie 2004, p. 678).

By the early 2000s, Barberis and Thaler (2003) describe the state of affairs as follows:

While the behavior of the aggregate stock market is not easy to understand from the rational point of view, promising rational models have nonetheless been developed and can be tested against behavioral alternatives. Empirical studies of the behavior of *individual* stocks have unearthed a set of facts which is altogether more frustrating for the rational paradigm. Many of these facts are about the *cross-section* of average returns: they document that one group of stocks earn higher average returns than another. These facts have come to be known as "anomalies" because they cannot be explained by the simplest and most intuitive model of risk and return in the financial economist's toolkit, the Capital Asset Pricing Model, or CAPM (p. 1087, original emphasis).

It is my professional view that the argument for inefficient markets based on the failure of the consumption CAPM in explaining asset pricing anomalies represents, to paraphrase Shiller (1984, p. 459), "one of the most remarkable errors in the history of economic thought."

#### **4.3.2 Do Psychological Biases Explain Asset Pricing Anomalies?**

Shleifer (2000) and Barberis and Thaler (2003) describe behavioral finance as consisting of two building blocks, psychological biases and limits to arbitrage. While psychological biases give rise to anomalies, limits to arbitrage prevent arbitrageurs from eliminating the mispricing, at least in the short run. However, recent empirical work has cast doubt on both building blocks.

**Are Investors More Biased Than Managers?** Many psychological biases have been proposed. Two prominent studies are Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998). In Barberis et al., conservatism means that individuals' beliefs are excessively slow to change, and representative heuristics means that after a consistent history of earnings news over several years, investors erroneously believe that the past history is representative of future earnings news. Conservatism is used to explain earnings momentum, as investors disregard the full information content of earnings announcements, and assign disproportionately high weights to their priors. Representative heuristics is used to explain the value premium. After

a consistent history of, say, positive earnings news, investors become overly optimistic about future earnings news, and overreact, pushing the stock price to unduly high levels.

In Daniel, Hirshleifer, and Subrahmanyam (1998), overconfidence means that investors overestimate the precision of, and tend to overreact to, their private information signals. Self-attribution means that investors too strongly attribute public signals that confirm the validity of their actions to high ability, and those that disconfirm their actions to bad luck or even sabotage. As a result, starting with unbiased beliefs about their ability, these investors tend to view public signals as on average confirming their prior private signals, at least in the short run. Overconfidence and self-attribution combine to produce continuous overreaction in the short run, giving rise to momentum, and the correction of the mispricing in the long run, giving rise to the value premium.

On the other hand, the body of evidence reviewed in this essay suggests that managers of individual firms seem to do a good job in aligning investment policies with their costs of capital, and this assignment drives many cross-sectional patterns that are anomalous in the consumption CAPM. If investors are psychologically biased, why would the managers of publicly traded companies be less biased? After all, investors and managers are from the same population. The recent intermediary asset pricing literature suggests that marginal investors tend to be sophisticated institutional investors in the financial services industry. Why would the managers of financial firms be more biased than the managers of nonfinancial firms? Other than institutional investors, why would individual investors exhibit a variety of biases while at home making portfolio choice decisions, but switch them off readily while at work making real investment decisions?

A more plausible explanation for the simultaneous failure of the consumption CAPM and the relative success of the investment CAPM, I believe, is aggregation (Section 4.2.2). The investment CAPM goes straight to the individual behavior, but the consumption CAPM works via a representative investor, which, according to the Sonnenschein-Mantel-Debreu theorem, does not really exist, and even if it does, is almost entirely detached from individual rationality. The recent U.S. presidential election offers a case in point. Given its extremely contentious and divisive nature, it would be exceedingly challenging, albeit not inconceivable, to assign a rational set of preferences to “the representative voter” who elected Donald Trump right after electing Barack Obama. By the same token, without addressing the aggregate problem, at least to some extent, imposing irrational preferences directly on the fictitious representative investor as in Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subramanyam (1998) is unlikely to be particularly illuminating.

**Are U.S. Investors More Biased Than Chinese Investors?** As noted, Titman, Wei, and Xie (2013) and Watanabe, Xu, Yao, and Yu (2013) document that the investment effect is stronger in developed countries than in developing countries. In particular, Titman et al.'s Table 3 reports that the equal-weighted size-adjusted return of the high-minus-low investment quintile is  $-0.35\%$  per month ( $t = -6.04$ ) averaged across developed countries, in contrast to  $-0.17\%$  ( $t = -1.06$ ) averaged across developing countries. Similar evidence has been reported for price momentum. Griffin, Ji, and Martin (2003, Table 1) report that momentum profits are on average  $0.51\%$  per month ( $t = 3.09$ ) in developed markets, in contrast to  $0.27\%$  ( $t = 1.21$ ) in emerging markets. Table 5, adapted from Chui, Titman, and Wei (2010), shows that momentum profits are on average  $0.86\%$  in developed markets, in contrast to  $0.49\%$  in emerging markets. Out of 21 developed markets, 19 show significant momentum at the 5% level, whereas only six out of 20 emerging markets show significant momentum. In particular, momentum profits are  $0.79\%$  ( $t = 3.44$ ) in the U.S., but  $0.26\%$  ( $t = 0.92$ ) in China. Momentum is a stronghold of behavioral finance. Unless investors in developed markets such as U.S. can be argued as more psychologically biased than investors in emerging markets such as China, the evidence seems an emphatic rejection of behavioral finance.

The evidence is more consistent with the investment CAPM. The crux is that whereas behavioral finance rides on dysfunctional, inefficient markets, the investment CAPM relies on well functioning, efficient markets for its mechanisms to work. As such, behavioral finance predicts that anomalies should be stronger in emerging markets, in which investors are less sophisticated, but weaker in developed markets, in which investors are more sophisticated. In contrast, the investment CAPM predicts that anomalies should be stronger in developed markets, such as the U.S., in which the incentives of managers and shareholders are more aligned, and managers are more likely to maximize the market equity in making investment decisions. Also, anomalies should be weaker in emerging markets, such as China, in which managers are often appointed by the government, and more likely to pursue social objectives such as maximizing employment, rather than the market value of equity.

**Limits to Limits to Arbitrage?** As noted, Shleifer and Vishny (1997) argue that trading frictions make arbitrage activities costly and incomplete, allowing mispricing to persist. Because limits to arbitrage are more severe in emerging markets than in developed markets, anomalies should be stronger in emerging markets. This prediction contradicts Table 5 and the related evidence.

**Table 5 : Momentum Profits by Country**

Source: Chui, Titman, and Wei (2010, Table III). The definition of developed markets and emerging markets is from Hou, Karolyi, and Kho (2011, Table 1). WML is momentum profits in percent per month, and  $t$  is the corresponding  $t$ -statistics. See Chui et al. for the construction of momentum profits and data source.

Panel A: Developed markets			Panel B: Emerging markets		
	WML	$t$		WML	$t$
Australia	1.08	4.76	Argentina	0.08	0.12
Austria	0.63	2.70	Bangladesh	1.68	2.75
Belgium	0.89	5.50	Brazil	0.46	0.96
Canada	1.35	6.29	Chile	0.99	3.60
Denmark	0.96	4.29	China	0.26	0.92
Finland	0.98	2.62	Greece	0.59	1.49
France	0.94	4.68	India	1.14	2.91
Germany	0.99	4.41	Indonesia	0.14	0.30
Hong Kong	0.77	3.18	Israel	0.32	1.19
Ireland	0.88	3.06	Korea	-0.34	-0.81
Italy	0.90	4.47	Malaysia	0.10	0.26
Japan	-0.04	-0.18	Mexico	0.69	2.00
Netherlands	0.83	4.40	Pakistan	0.46	1.05
New Zealand	1.58	5.01	Philippines	0.37	0.68
Norway	1.05	3.77	Poland	1.76	3.33
Singapore	0.14	0.47	Portugal	0.31	0.93
Spain	0.63	2.24	South Africa	0.94	3.29
Sweden	0.71	2.27	Taiwan	-0.20	-0.48
Switzerland	0.82	4.39	Thailand	0.48	1.10
United Kingdom	1.13	7.08	Turkey	-0.41	-0.96
United States	0.79	3.44			
Average	0.86		Average	0.49	

**Measurement Ahead of Theory?** Koopmans (1947) emphasizes an artful combination of theory, econometrics, and measurement in scientific work:

[In] research in economic dynamics the Kepler stage and the Newton stage of inquiry need to be more intimately combined and to be pursued simultaneously. Fuller utilization of the concepts and hypotheses of economic theory ... *as a part of the processes of observation and measurement* promises to be a shorter road, perhaps even the only possible road, to the understanding of cyclical fluctuations (p. 162, original emphasis).

The historical rise of behavioral finance has been fueled by the discoveries of anomalies. Despite enormous empirical contributions in the anomalies literature, what is missing, it seems, is a coherent behavioral asset pricing framework. Also, behavioral models have not been subject to the rigorous structural econometric tests similar to those imposed on the consumption CAPM and the investment CAPM. While these tests reveal unreasonable parameter estimates and parameter instability, not going through these tests might have obscured similar problems in behavioral models.

### 4.3.3 The Joint-Hypothesis Problem Revisited

Fama (1970, 1991, 1998) and Fama and French (1993, 1996) defend efficient markets. With the investment CAPM, I drastically modify, but ultimately strengthen their arguments.

Fama’s (1970) definition of efficient markets needs no modification almost half a century later:

The assumptions that the conditions of market equilibrium can be stated in terms of expected returns and that equilibrium expected returns are formed on the basis of (and thus “fully reflect”) the information set  $\Phi_t$  have a major empirical implication—they rule out the possibility of trading systems based only on information in  $\Phi_t$  that have expected profits or returns in excess of equilibrium expected profits or returns. Thus, let

$$x_{jt+1} = p_{jt+1} - E[p_{jt+1}|\Phi_t]. \quad (35)$$

Then

$$E[x_{jt+1}|\Phi_t] = 0 \quad (36)$$

which, *by definition*, says that the sequence  $\{x_{jt}\}$  is a “fair game” with respect to the information sequence  $\{\Phi_t\}$ . Or, equivalently, let

$$z_{jt+1} = r_{jt+1} - E[r_{jt+1}|\Phi_t], \quad (37)$$

then

$$E[z_{jt+1}|\Phi_t] = 0, \quad (38)$$

so that the sequence  $\{z_{jt}\}$  is also a “fair game” with respect to the information sequence  $\{\Phi_t\}$  (p. 384–385, original emphasis).

Twenty years after Fama (1970), with anomalies all the rage, Fama (1991) brings the joint-hypothesis problem to the front and center of asset pricing research:

Ambiguity about information and trading costs is not, however, the main obstacle to inferences about market efficiency. The joint-hypothesis problem is more serious. Thus, market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model. This point ... says that we can only test whether information is properly reflected in prices in the context of a pricing model that defines the meaning of “properly.” As a result, when we find anomalous evidence on the behavior of returns, the way it should be split between market inefficiency or a bad model of market equilibrium is ambiguous (p. 1575–1576).

Fama (1998) reiterates the joint-hypothesis problem, and shows that some anomalies disappear in the three-factor model. However, the three-factor model leaves many anomalies unexplained. The  $q$ -factor model is a direct attempt at alleviating the joint-hypothesis problem. It represents a major step forward from the three-factor model, as evidenced in the response in Fama and French (2015).

**A Tribute to Fama and French (1993)** The Fama-French three-factor model has served its historical purpose, rather admirably. After the CAPM is firmly rejected and abandoned in Fama and French (1992), the three-factor model fills the vacuum left by the CAPM as the workhorse model, and takes on the burden of defending efficient markets. The mainstream asset pricing theory is of little help, as the consumption CAPM often underperforms the CAPM.

Alas, the largely empirical nature of the three-factor model leaves it vulnerable to the data mining critique. The conjecture of the book-to-market factor as a relative distress factor in Fama and French (1996) is also rejected by the distress anomaly, which shows that more distressed firms earn lower average returns than less distressed firms (Dichev 1998). More generally, the interpretation of characteristics-based factors as sources of risk in the intertemporal CAPM or arbitrage pricing theory has fallen on deaf ears. Both theories are silent about the identities of state variables, and neither make the conceptual leap of linking expected returns to firm characteristics.

**Interpreting Factors: The Investment CAPM Perspective** Fama and French (1996) interpret their three-factor model as a risk factor model:

[If] assets are priced rationally, variables that are related to average returns, such as size and book-to-market equity, must proxy for sensitivity to common (shared and thus undiversifiable) risk factors in returns. The time-series regressions give direct evidence on this issue. In particular, the slopes and  $R^2$  values show whether mimicking portfolios for risk factors related to size and [book-to-market] capture shared variation in stock and bond returns not explained by other factors (p. 4–5).

[The] empirical successes of [the three-factor model] suggest that it is an equilibrium pricing model, a three-factor version of Merton's (1973) intertemporal CAPM (ICAPM) or Ross's (1976) arbitrage pricing theory (APT). In this view, SMB and HML mimic combinations of two underlying risk factors or state variables of special hedging concern to investors (p. 57).

This interpretation has been controversial. Daniel and Titman (1997) show that expected returns do not correlate positively with factor loadings after controlling for characteristics:

In equilibrium asset pricing models the covariance structure of returns determines expected returns. Yet we find that variables that reliably predict the future covariance structure do not predict future returns. Our results indicate that high book-to-market stocks and stocks with low capitalizations have high average returns whether or not they have the return patterns (i.e., covariances) of other small and high book-to-market stocks (p. 4).

This evidence has often been interpreted as suggesting mispricing in behavioral finance:

One general feature of the rational approach is that it is loadings or betas, and not firm characteristics, that determine average returns.... [Daniel and Titman's (1997)] results appear quite damaging to the rational approach (Barberis and Thaler 2003, p. 1091).

[It] is important to empirically distinguish (1) the covariance between stock returns and a given attribute from (2) the returns attributable to the characteristic. Finding evidence in support of (1) is consistent with a risk based explanation for the return relation, whereas finding (2) would suggest mispricing (Richardson, Tuna, and Wysocki 2010, p. 430).

From the investment CAPM perspective, I interpret characteristics-based factor models as linear approximations of the investment CAPM in equation (7). As a more fundamental departure from Fama and French (1993, 1996), I interpret the  $q$ -factor model as just a parsimonious description of

the cross section, not necessarily a risk factor model, and the  $q$ -factor loadings as just regression slopes, not necessarily measures of some inexplicable sources of risk. Fama and French go for the risk interpretation, probably thinking that it is the only way to defend efficient markets, but this step is not necessary either. The investment CAPM predicts all kinds of relations between characteristics and expected returns, without any trace of mispricing. And as noted, the evidence of characteristics dominating covariances in Daniel and Titman (1997) can be largely attributed to measurement errors in covariances (Lin and Zhang 2013). Finally, the failure of the consumption CAPM can be largely attributed to the aggregation problem (as well as measurement errors in consumption data).

Time series and cross-sectional regressions are just two different ways of summarizing empirical correlations, and are largely equivalent in economic terms. If a characteristic shows up significant in cross-sectional regressions, its factor mimicking portfolio is likely to show “explanatory” power in factor regressions. If a factor earns a significant average return in the time series, the slope of its underlying characteristic is likely to be significant in cross-sectional regressions. Factor loadings are no more primitive than characteristics, and vice versa, in “explaining” expected returns.

**Beyond the “Risk Doctrine”** More generally, the investment CAPM broadens efficient markets beyond the “risk doctrine.” An early statement of the doctrine is in Fama (1970):

Most of the available work is based only on the assumption that the conditions of market equilibrium can (somehow) be stated in terms of expected returns. In general terms, like the two parameter model such theories would posit that conditional on some relevant information set, the equilibrium expected return on a security is a function of its “risk.” And different theories would differ primarily in how “risk” is defined (p. 384).

The anomalies literature puts efficient markets in peril because of the failure of traditional risk models. The investment CAPM shows that the peril is more apparent than real.

The risk doctrine asks what risks “explain” asset pricing anomalies. This question has been at the center of modern asset pricing research. The investment CAPM questions the question (Lin and Zhang 2013). As shown in Section 1, the consumption CAPM and the investment CAPM are the two sides of the same coin in general equilibrium, delivering identical expected returns. However, the risk doctrine only describes the consumption CAPM, which predicts that consumption risks are sufficient statistics for expected returns, and after risks are controlled for, characteristics should not matter. It is clear that the risk doctrine is conceptually incomplete. It does not apply to the investment CAPM, which predicts that characteristics are sufficient statistics for expected returns, and after characteristics are controlled for, risks should not matter.

In general equilibrium, risks, expected returns, and characteristics are all endogenously determined simultaneously. Neither risks nor characteristics “determine” expected returns. The investment CAPM is as fundamental as the consumption CAPM in economy theory. The concept that risks determine expected returns is a relic and an illusion from the CAPM, because its inventors happen to take stock returns and their covariances as exogenous. Risks are no more exogenous than investment and profitability in theory, and should not be put on a pedestal in applied work.

The risk doctrine is also impractical. Without addressing the aggregation critique, the consumption CAPM is not testable. After almost 50 years of the anomalies literature since Ball and Brown (1968), it is perhaps time to heed the advice from Kuhn (1962):

[The] really pressing problems, e.g., a cure for cancer and the design of a lasting peace, are often not puzzles at all, largely because they may not have any solution. Consider the jigsaw puzzle whose pieces are selected at random from each of two different puzzle boxes. Since that problem is likely to defy (though it might not) even the most ingenious of men, it cannot serve as a test of skill. In solution in any usual sense, it is not a puzzle at all. Though intrinsic value is no criterion for a puzzle, the assured existence of a solution is (p. 36–37).

## 5 Challenges for Future Work

In this article I have attempted to articulate the broad, and hopefully internally consistent, perspective of the investment CAPM for asset pricing. This perspective is quite different from the perspective of the consumption CAPM and that of behavioral finance. My message is a simple one: Like any other prices in microeconomic theory, asset prices are determined jointly by supply and demand in general equilibrium. As such, the consumption CAPM and behavioral finance, both of which are demand-based, cannot possibly be the whole story. At the core, the investment CAPM is tantamount to the net present value rule in corporate finance, which should not be controversial. Explaining asset pricing anomalies requires a fundamental change in perspective.

I have reviewed two workhorse models in this rapidly expanding literature, the  $q$ -factor model for estimating expected stock returns and the multiperiod investment CAPM for structural estimation and tests. The  $q$ -factor model of Hou, Xue, and Zhang (2015) achieves the important goal of dimension reduction in the anomalies literature in the spirit of what Fama and French (1996) did 20 years ago. The structural investment CAPM of Liu, Whited, and Zhang (2009) takes the first principle of real investment to cross-sectional returns data in the spirit of what Hansen and Singleton (1982) did with the consumption CAPM 35 years ago.

I have also attempted to put the investment CAPM into the historical perspective. I have traced

the origin of the investment CAPM to at least Fisher (1930), who provided a symmetric treatment of the impatience and investment opportunity theories of the equilibrium interest rate. Despite the important writings in Hirshleifer (1958, 1970), Modigliani and Miller (1958), and Cochrane (1991), modern asset pricing is thoroughly dominated by the consumption CAPM. A long list of foundational contributions, including Markowitz (1952), Sharpe (1964), Lintner (1965), Merton (1973), Rubinstein (1976), Lucas (1978), and Breeden (1979), has put investors at the center of inquiry, while abstracting from managers altogether. This abstraction is unjustifiable, ultimately, because asset prices are equilibrated by both demand and supply. Fixing the supply of assets in one's model to focus on the demand does not mean that the supply is not important in real life. On the contrary, the abstraction is disastrous, empirically, because the aggregation critique means that the consumption CAPM is ill-equipped for anomalies. The investment CAPM, which is derived from the first principle of investment for individual firms, is largely immune to the aggregation problem. In the data, anomalies in the consumption CAPM turn out to be regularities in the investment CAPM.

The investment CAPM literature is young, with many exciting new opportunities for future work. Are the investment and profitability premiums reliable? While the ubiquitous data mining concern is hard to address completely, an effective antidote is to study global financial data. Do these premiums exist in countries outside the U.S.? Does the  $q$ -factor model absorb the explanatory power of the Carhart (1997) four-factor model in global data as in the U.S. data?

The investment CAPM has provided economic explanations to most anomalies that have been attributed to behavioral finance. It has become important to disentangle the investment-based from behavioral explanations. The key identification is that behavioral finance relies on dysfunctional, inefficient markets for its mechanisms to work, but the investment CAPM relies on well functioning, efficient markets. As noted, in global data, both the investment and momentum effects are stronger in developed markets than in emerging markets. Do similar cross-country patterns also hold for other important cross-sectional patterns, such as the value and profitability premiums?

Asset management is a vast and important field (see Fischer and Wermers 2012 and Ang 2014 for reviews). Can the  $q$ -factor model help improve the measurement of mutual fund performance? The Carhart (1997) four-factor model is the current workhorse model in this area. In view of the superior performance of the  $q$ -factor model over the Carhart model, it seems natural to apply the  $q$ -factor model in this area. Xue (2014) takes a first step in this direction.

A significant portion of the return spreads from anomaly-based strategies occurs over a few

days surrounding earnings announcement dates (see Jegadeesh and Titman 1993 for momentum and La Porta, Lakonishok, Shleifer, and Vishy 1997 for the value premium). Because the consumption CAPM has nothing to say about ex-post returns, this evidence has often been cited as indicating investors' expectational errors. However, the investment CAPM applies ex-ante as well as ex-post, and consequently, does predict that a portion of anomalies should occur around earnings announcements. Is there any way to disentangle these two competing explanations empirically?

Asness, Moskowitz, and Pedersen (2013) show that value and momentum are pervasive internationally across different asset classes, including individual stocks, country equity index futures, government bonds, currencies, and commodity futures. To what extent can the  $q$ -factor model be applied to these alternative assets? While it seems straightforward to apply the  $q$ -factor model to global stocks, country equity indices, corporate bonds, and even real estate assets, other asset classes such as currencies, government bonds, and commodities require additional theorizing. In the context of real estate investment trusts, Bond and Xue (2015) measure investment as property acquisition and profitability as property income and price appreciation, and show that the two economic fundamentals have substantial predictive power for real estate returns.

While the investment CAPM is immune to the aggregation critique in theory, its current structural estimation, which is done at the portfolio level, is not. To be free of any aggregation bias, the structural estimation should be done at the firm level. This step requires additional methodological innovations on handling unbalanced panel data. Can the firm-level estimation address the parameter instability issue across different testing portfolios? Or equivalently, can the firm-level estimation account for value and momentum simultaneously in a single structural investment CAPM specification? What about the investment and profitability premiums simultaneously?

How can the supply approach to valuation in Belo, Xue, and Zhang (2013) be applied to individual firms? Their valuation framework can be extended to incorporate additional productive inputs such as labor and intangible assets. Their econometric framework can also be used to quantify the valuation impact of corporate finance frictions, such as the agency conflicts between shareholders and managers. More generally, what is the impact of time-varying and cross-sectionally varying expected returns on corporate investment, financing, and payout policies, and vice versa? The dynamic corporate finance literature has mostly worked with risk neutrality (Hennessy and Whited 2005). Integrating this prominent literature with the investment CAPM literature seems fruitful.

The investment CAPM provides a neoclassical foundation for Graham and Dodd's (1934) se-

curity analysis. Ou and Penman (1989) describe the traditional view as follows:

Firms' ('fundamental') values are indicated by information in financial statements. Stock prices deviate at times from these values and only slowly gravitate towards the fundamental values. Thus, analysis of published financial statements can discover values that are not reflected in stock prices. Rather than taking prices as value benchmarks, 'intrinsic values' discovered from financial statements serve as benchmarks with which prices are compared to identify overpriced and underpriced stocks. Because deviant prices ultimately gravitate to the fundamentals, investment strategies which produce 'abnormal returns' can be discovered by the comparison of prices to these fundamental values (p. 296).

When discussing "abnormal returns," the traditional view seems to use a cross-sectionally constant discount rate as the null in efficient markets. The investment CAPM changes the big picture, by showing cross-sectionally varying expected returns from the first principle of real investment. Much remains to be done to integrate the investment CAPM with capital markets research in accounting.

## References

- Abarbanell, Jeffery S., and Brian J. Bushee, 1997, Fundamental analysis, future earnings, and stock prices, *Journal of Accounting Research* 35, 1–24.
- Abarbanell, Jeffery S., and Brian J. Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *The Accounting Review* 73, 19–45.
- Abel, Andrew B., 1983, Optimal investment under uncertainty, *American Economic Review* 73, 228–233.
- Abel, Andrew B., and Janice C. Eberly, 1994, A unified model of investment under uncertainty, *American Economic Review* 84, 1369–1384.
- Aharoni, Gil, Bruce Grundy, and Qi Zeng, 2013, Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis, *Journal of Financial Economics* 110, 347–357.
- Anderson, Christopher W., and Luis Garcia-Feijóo, 2006, Empirical evidence on capital investment, growth options, and security returns, *Journal of Finance* 61, 171–194.
- Ang, Andrew, 2014, *Asset Management: A Systematic Approach to Factor Investing*, Oxford University Press.
- Arrow, Kenneth J., 1964, The role of securities in the optimal allocation of risk-bearing, *Review of Economic Studies* 31, 91–96.
- Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.
- Back, Kerry E., 2010, *Asset Pricing and Portfolio Choice Theory*, Oxford University Press.
- Bai, Hang, Kewei Hou, Howard Kung, Erica X. N., and Lu Zhang, 2017, The CAPM strikes back? An investment model with disasters, working paper, The Ohio State University.

- Ball, Ray, and Philip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research* 6, 159-178.
- Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28-45.
- Balvers, Ronald J., and Dayong Huang, 2007, Productivity-based asset pricing: Theory and evidence, *Journal of Financial Economics* 86, 405-445.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307-343.
- Barberis, Nicholas, and Richard Thaler, 2003, A survey of behavioral finance, in George M. Constantinides, Milton Harris, and René M. Stulz eds., *Handbook of the Economics of Finance*, p. 1053-1123, Elsevier, North Holland.
- Barillas, Francisco, and Jay Shanken, 2015, Comparing asset pricing models, working paper, Emory University.
- Belo, Frederico, 2010, Production-based measures of risk for asset pricing, *Journal of Monetary Economics* 57, 146-163.
- Belo, Frederico, Chen Xue, and Lu Zhang, 2013, A supply approach to valuation, *Review of Financial Studies* 26, 3029-3067.
- Berk, Jonathan B., and Peter M. DeMarzo, 2009, *Corporate Finance*, 2nd ed., Prentice Hall.
- Bernard, Victor L., and Jacob K. Thomas, 1989, Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* Supplement 27, 1-48.
- Bernard, Victor L., and Jacob K. Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305-340.
- Bernstein, Peter L., 1992, *Capital Ideas: The Improbably Origins of Modern Wall Street*, New Jersey: John Wiley & Sons.
- Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The Capital Asset Pricing Model: Some empirical tests, in Michael C. Jensen, ed., *Studies in the Theory of Capital Markets* 79-121, New York: Praeger.
- Bodie, Zvi, Alex Kane, and Alan J. Marcus, 2014, *Investments*, McGraw-Hill Education.
- Böhm-Bawerk, Eugen V., 1891, *The Positive Theory of Capital*, London: Macmillan. Translated by William Smart.
- Bond, Shaun, and Chen Xue, 2015, The cross section of expected real estate returns: Insights from investment-based asset pricing, forthcoming, *Journal of Real Estate Finance and Economics*.
- Bond, Stephen, and John Van Reenen, 2007, Microeconomic models of investment and employment, in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics* 4417-4498, Amsterdam: North-Holland.

- Brainard, William C., and James Tobin, 1968, Pitfalls in financial model building, *American Economic Review* 58, 99–122.
- Brealey, Richard A., Stewart C. Myers, and Franklin Allen, 2011, *Principles of Corporate Finance*, 10th ed., McGraw-Hill Irwin.
- Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.
- Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.
- Brock, William A., 1982, Asset prices in a production economy, in: John L. McCall, ed., *Economics of Information and Uncertainty*, Illinois: University of Chicago Press.
- Brown, Lawrence D., and Michael S. Rozeff, 1979, Univariate time-series models of quarterly accounting earnings per share: A proposal model, *Journal of Accounting Research* 17, 179–189.
- Butler, Alexander W., Jess Cornaggia, Gustavo Grullon, and James P. Weston, 2011, Corporate financing decisions, managerial market timing, and real investment, *Journal of Financial Economics* 101, 666–683.
- Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.
- Carhart, Mark M. 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2004, Corporate investment and asset price dynamics: Implications for the cross section of returns, *Journal of Finance* 59, 2577–2603.
- Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1713.
- Chen, Shuping, Mark L. DeFond, and Chul W. Park, 2002, Voluntary disclosure of balance sheet information in quarterly earnings announcements, *Journal of Accounting and Economics* 33, 229–251.
- Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.
- Chirinko, Robert S., 1993, Business fixed investment spending: Modeling strategies, empirical results, and policy implications, *Journal of Economic Literature* 31, 1875–1911.
- Chui, Andy C. W., Sheridan Titman, and K. C. John Wei, 2010, Individualism and momentum around the world, *Journal of Finance* 65, 361–392.
- Cochrane, John H., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* 46, 209–237.
- Cochrane, John H., 1993, Rethinking production under uncertainty, working paper, University of Chicago.

- Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.
- Cochrane, John H., 2005, *Asset Pricing*, New Jersey: Princeton University Press.
- Cooper, Ilan, and Richard Priestley, 2011, Real investment and risk dynamics, *Journal of Financial Economics* 101, 182–205.
- Cooper, Ilan, and Richard Priestley, 2015, The expected returns and valuations of private and public firms, forthcoming, *Journal of Financial Economics*.
- Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1652.
- Cooper, Russell W., and John C. Haltiwanger, 2006, On the nature of capital adjustment costs, *Review of Economic Studies* 73, 611–633.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363–384.
- Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreaction, *Journal of Finance* 53, 1839–1885.
- Daniel, Kent D., and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–33.
- Daniel, Kent D., and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.
- Danthine, Jean-Pierre, and John B. Donaldson, 2015, *Intermediate Financial Theory*, 3rd edition, Academic Press.
- Deaton, Angus, 1992, *Understanding Consumption*, Oxford: Clarendon Press.
- De Bondt, Werner F. M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.
- Debreu, Gerard, 1959, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New Haven and London: Yale University Press.
- Debreu, Gerard, 1974, Excess-demand functions, *Journal of Mathematical Economics* 1, 15–21.
- Diamond, Peter A., 1967, The role of a stock market in a general equilibrium model with technological uncertainty, *American Economic Review* 57, 759–776.
- Dichev, Iliia, 1998, Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53, 1141–1148.
- Dixit, Avinash K., and Robert S. Pindyck, 1994, *Investment Under Uncertainty*, New Jersey: Princeton University Press.
- Doms, Mark, and Timothy Dunne, 1998, Capital adjustment patterns in manufacturing plants, *Review of Economic Dynamics* 1, 409–429.

- Duffie, Darrell, 2001, *Dynamic Asset Pricing Theory*, 3rd edition, New Jersey: Princeton University Press.
- Easton, Peter D., and Mark E. Zmijewski, 1993, SEC form 10K/10Q reports and annual reports to shareholders: Reporting lags and squared market model prediction errors, *Journal of Accounting Research* 31, 113–129.
- Fairfield, Patricia M., Scott Whisenant, and Terry Lombardi Yohn, 2003, Accrued earnings and growth: Implications for future profitability and market mispricing, *The Accounting Review* 78, 353–371.
- Fama, Eugene F., 1965, The behavior of stock-market prices, *Journal of Business* 38, 34–105.
- Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.
- Fama, Eugene F., 1976, *Foundations of Finance*, New York: Basic Books, Publishers.
- Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.
- Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283–306.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanation of asset pricing anomalies, *Journal of Finance* 51, 55–84.
- Fama, Eugene F., and Kenneth R. French, 2006, Profitability, investment, and average returns, *Journal of Financial Economics* 82, 491–518.
- Fama, Eugene F., and Kenneth R. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.
- Fama, Eugene F., and Kenneth R. French, 2013, A four-factor model for the size, value, and profitability patterns in stock returns, Fama-Miller working paper, University of Chicago.
- Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
- Fama, Eugene F., and Kenneth R. French, 2016, Dissecting anomalies with a five-factor model, *Review of Financial Studies* 29, 69–103.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Fama, Eugene F., and Merton H. Miller, 1972, *The Theory of Finance*, Illinois: Dryden Press.
- Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, 1988, Financing constraints and corporate investment, *Brookings Papers on Economic Activities* 1, 141–195.

- Fischer, Bernd R., and Russ Wermers, 2012, *Performance Evaluation and Attribution of Security Portfolios*, Academic Press.
- Fisher, Irving, 1930, *The Theory of Interest: As Determined by Impatience to Spend Income and Opportunity to Invest It*, 1st ed., New York: Macmillan.
- Foster, George, Chris Olsen, and Terry Shevlin, 1984, Earnings releases, anomalies, and the behavior of security returns, *The Accounting Review* 59, 574–603.
- Gebhardt, William R., Charles M. C. Lee, and Bhaskaram Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135–176.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.
- Gomes, Joao F., Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross section of returns, *Journal of Political Economy* 111, 693–732.
- Gomes, Joao F., Amir Yaron, and Lu Zhang, 2006, Asset pricing implications of firms' financing constraints, *Review of Financial Studies* 19, 1321–1356.
- Graham, Benjamin, and David Dodd, 1934, *Security Analysis*, New York: Whittlesey House, McGraw-Hill.
- Griffin, John M., Xiuqing Ji, and J. Spencer Martin, 2003, Momentum investing and business cycle risk: Evidence from pole to pole, *Journal of Finance* 58, 2515–2547.
- Hall, Robert E., and Dale W. Jorgenson, 1967, Tax policy and investment behavior, *American Economic Review* 57, 391–414.
- Hansen, Lars Peter, 1982, Large sample properties of generalized methods of moments estimators, *Econometrica* 40, 1029–1054.
- Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–614.
- Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1288.
- Harvey, Campbell R., 2001, The specification of conditional expectations, *Journal of Empirical Finance* 8, 573–638.
- Hayashi, Fumio, 1982, Tobin's marginal  $q$  and average  $q$ : A neoclassical interpretation, *Econometrica* 50, 213–224.
- Hennessy, Christopher A., and Toni M. Whited, 2005, Debt dynamics, *Journal of Finance* 60, 1129–1165.
- Hirshleifer, Jack, 1958, On the theory of optimal investment decision, *Journal of Political Economy* 66, 329–352.
- Hirshleifer, Jack, 1965, Investment decision under uncertainty: Choice-theoretic approaches, *Quarterly Journal of Economics* 79, 509–536.

- Hirshleifer, Jack, 1966, Investment decision under uncertainty: Applications of the state-preference approach, *Quarterly Journal of Economics* 80, 252–277.
- Hirshleifer, Jack, 1970, *Investment, Interest, and Capital*, New Jersey: Prentice-Hall.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2016, A comparison of new factor models, The Ohio State University and University of Cincinnati.
- Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2017, Motivating factors, Louisiana State University, The Ohio State University, and University of Cincinnati.
- Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho, 2011, What factors drive global stock returns? *Review of Financial Studies* 24, 2527–2574.
- Huang, Chi-fu, and Robert H. Litzenberger, 1988, *Foundations for Financial Economics*, North-Holland.
- Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield Publishers.
- Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.
- Jensen, Michael C., 1986, Agency costs of free cash flow, corporate finance, and takeovers, *American Economic Review* 76, 323–329.
- Jermann, Urban J., 2010, The equity premium implied by production, *Journal of Financial Economics* 98, 279–296.
- Jorgenson, Dale W., 1963, Capital theory and investment behavior, *American Economic Review* 53, 247–259.
- Kirman, Alan P., 1992, Whom or what does the representative individual represent? *Journal of Economic Perspectives* 6, 117–136.
- Koller, Tim, Marc Goedhart, and David Wessels, 2015, *Valuation: Measuring and Managing the Value of Companies*, 6th ed., New Jersey: John Wiley and Sons.
- Koopmans, Tjalling C., 1947, Measurement without theory, *Review of Economics and Statistics* 29, 161–172.
- Kothari, S. P., 2001, Capital markets research in accounting, *Journal of Accounting and Economics* 31, 105–231.
- Kuhn, Thomas S., 1962, *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.
- Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

- Lam, F. Y. Eric C., and K. C. John Wei, 2011, Limits-to-arbitrage, investment frictions, and the asset growth anomaly, *Journal of Financial Economics* 102, 127–149.
- Lamont, Owen, Christopher Polk and Jesús Saá-Requejo, 2001, Financial constraints and stock returns, *Review of Financial Studies* 14, 529–554.
- Landreth, Harry, and David C. Colander, 2002, *History of Economic Thought*, 4th edition, Houghton Mifflin Company.
- La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859–874.
- Lettau, Martin, and Sydney Ludvigson, 2002, Time-varying risk premia and the cost of capital: An alternative implication of the  $Q$  theory of investment, *Journal of Monetary Economics* 49, 31–66.
- Lev, Baruch, and S. Ramu Thiagarajan, 1993, Fundamental information analysis, *Journal of Accounting Research* 31, 190–215.
- Li, Dongmei, and Lu Zhang, 2010, Does  $q$ -theory with investment frictions explain anomalies in the cross section of returns? *Journal of Financial Economics* 98, 297–314.
- Li, Erica X. N., Dmitry Livdan, and Lu Zhang, 2009, Anomalies, *Review of Financial Studies* 22, 4301–4334.
- Lin, Xiaoji, and Lu Zhang, 2013, The investment manifesto, *Journal of Monetary Economics* 60, 351–366.
- Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- Lipson, Marc L., Sandra Mortal, and Michael J. Schill, 2011, On the scope and drivers of the asset growth effect, *Journal of Financial and Quantitative Analysis* 46, 1651–1682.
- Liu, Laura Xiaolei, Toni M. Whited, and Lu Zhang, 2009, Investment-based expected stock returns, *Journal of Political Economy* 117, 1105–1139.
- Liu, Laura Xiaolei, and Lu Zhang, 2014, A neoclassical interpretation of momentum, *Journal of Monetary Economics* 67, 109–128.
- Long, John B., Jr., 1974, Stock prices, inflation, and the term structure of interest rates, *Journal of Financial Economics* 1, 131–170.
- Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.
- Lucas, Robert E., Jr., 1972, Expectations and the neutrality of money, *Journal of Economic Theory* 4, 103–124.
- Lucas, Robert E., Jr., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1445.
- Lucas, Robert E., Jr., and Edward C. Prescott, 1971, Investment under uncertainty, *Econometrica* 39, 659–681.

- Ludvigson, Sydney C., 2013, Advances in consumption-based asset pricing: Empirical tests, in George M. Constantinides, Milton Harris, and René M. Stulz eds., *Handbook of the Economics of Finance* volume 2B, 799–906.
- Lyandres, Evgeny, Le Sun, and Lu Zhang, 2008, The new issues puzzle: Testing the investment-based explanation, *Review of Financial Studies* 21, 2825–2855.
- Mantel, Rolf R., 1974, On the characterization of aggregate excess-demand, *Journal of Economic Theory* 7, 348–353.
- Markowitz, Harry M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.
- Markowitz, Harry M., 1959, *Portfolio Selection: Efficient Diversification of Investments*, New York: John Wiley & Sons.
- Marshall, Alfred, 1961, *Principles of Economics*, 9th edition, London: Macmillan, first published in 1890.
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Merz, Monika, and Eran Yashiv, 2007, Labor and the market value of the firm, *American Economic Review* 97, 1419–1431.
- Miller, Merton H., and Franco Modigliani, 1961, Dividend policy, growth, and the valuation of shares, *Journal of Business* 34, 411–433.
- Miller, Merton H., and Myron S. Scholes, 1972, Rate of return in relation to risk: A re-examination of some recent findings. In Michael C. Jensen, ed. *Studies in the Theory of Capital Markets*. New York: Praeger.
- Modigliani, Franco, and Merton H. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review* 48, 261–297.
- Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768–783.
- Mussa, Michael L., 1977, External and internal adjustment costs and the theory of aggregate and firm investment, *Economica* 44, 163–178.
- Muth, John F., 1961, Rational expectations and the theory of price movements, *Econometrica* 29, 315–335.
- Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.
- Ohlson, James A., 1995, Earnings, book values, and dividends in equity valuation, *Contemporary Accounting Research* 11, 661–687.
- Ou, Jane A., and Stephen H. Penman, 1989, Financial statement analysis and the prediction of stock returns, *Journal of Accounting and Economics* 11, 295–329.
- Penman, Stephen H., 2013, *Financial Statement Analysis and Security Returns*, 5th ed., New York: McGraw-Hill Irwin.

- Penman, Stephen H., and Julie Lei Zhu, 2015, Accounting anomalies, risk, and return, *The Accounting Review* 89, 1835–1866.
- Penman, Stephen H., Scott A. Richardson, and Irem Tuna, 2007, The book-to-market effect in stock returns: Accounting for leverage, *Journal of Accounting Research* 45, 427–468.
- Piotroski, Joseph D., 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* 38, Supplement: Studies on accounting information and the economics of the firm, 1–41.
- Restoy, Fernando, and G. Michael Rockinger, 1994, On stock market returns and returns on investment, *Journal of Finance* 49, 543–556.
- Richardson, Scott A., Richard G. Sloan, Mark T. Soliman, and Irem Tuna, 2005, Accrual reliability, earnings persistence and stock prices, *Journal of Accounting and Economics* 39, 437–485.
- Richardson, Scott A., Irem Tuna, and Peter Wysocki, 2010, Accounting anomalies and fundamental analysis: A review of recent research advances, *Journal of Accounting and Economics* 50, 410–454.
- Ritter, Jay R., 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.
- Roll, Richard, 1977, A critique of the asset pricing theory's tests: Part I, *Journal of Financial Economics* 4, 129–176.
- Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–16.
- Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Ross, Stephen A., Randolph W. Westerfield, and Jeffrey Jaffe, 2008, *Corporate Finance*, 8th ed., McGraw-Hill Irwin.
- Roy, Andrew D., 1952, Safety first and the holding of assets, *Econometrica* 20, 431–449.
- Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Management Science* 7, 407–425.
- Rubinstein, Mark, 2006, *A History of the Theory of Investments*, New Jersey: John Wiley & Sons.
- Shanken, Jay, 2015, Discussion on “HXZ: A comparison of new factor models,” ASU Sonoran Winter Finance Conference.
- Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.
- Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457–498.

- Shiller, Robert J., 1990, Speculative prices and popular models, *Journal of Economic Perspectives* 4, 55–65.
- Shleifer, Andrei, 2000, *Inefficient Markets: An Introduction to Behavioral Finance*, Oxford University Press.
- Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.
- Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289–315.
- Sonnenschein, Hugo, 1973, Do Walras' identity and continuity characterize the class of community excess-demand function? *Journal of Economic Theory* 6, 345–354.
- Stambaugh, Robert F., and Yu Yuan, 2016, Mispricing factors, forthcoming, *Review of Financial Studies*.
- Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.
- Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2013, Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48, 1405–1432.
- Tobin, James, 1969, A general equilibrium approach to monetary theory, *Journal of Money, Credit, and Banking* 1, 15–29.
- Treynor, Jack L., 1962, Toward a theory of market value of risky assets, published in 1999 in Robert A. Korajczyk, ed. *Asset Pricing and Portfolio Performance*, 15–22, London: Risk Publications.
- Veracierto, Marcelo L., 2002, Plant-level irreversible investment and equilibrium business cycles, *American Economic Review* 92, 181–197.
- Vitorino, Maria Ana, 2014, Understanding the effect of advertising on stock returns and firm value: Theory and evidence from a structural model, *Management Science* 60, 227–245.
- Watanabe, Akiko, Yan Xu, Tong Yao, and Tong Yu, 2013, The asset growth effect: Insights from international equity markets, *Journal of Financial Economics* 108, 529–563.
- Whited, Toni M., 1992, Debt, liquidity constraints, and corporate investment: Evidence from panel data, *Journal of Finance* 47, 1425–1460.
- Whited, Toni M., 1998, Why do investment Euler equations fail? *Journal of Business and Economic Statistics* 16, 479–488.
- Whited, Toni M., and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19, 531–559.
- Wu, Jin G., Lu Zhang, and X. Frank Zhang, 2010, The  $q$ -theory approach to understanding the accrual anomaly, *Journal of Accounting Research* 48, 177–223.

- Xing, Yuhang, 2008, Interpreting the value effect through the  $Q$ -theory: An empirical investigation, *Review of Financial Studies* 21, 1767–1795.
- Xue, Chen, 2014, An investment-based investigation of mutual fund performance, working paper, University of Cincinnati.
- Zhang, Lu, 2005a, Anomalies, NBER working paper 11322.
- Zhang, Lu, 2005b, The value premium, *Journal of Finance* 60, 67–103.
- Zhang, Lu, 2017, Quantitative investment CAPM, working paper, The Ohio State University.
- Zhang, X. Frank, 2007, Accruals, investment, and the accrual anomaly, *The Accounting Review* 82, 1333–1363.