Anomalies

Lu Zhang
William E. Simon Graduate School of Business Administration
University of Rochester and NBER

March 2006

Abstract

Most asset pricing anomalies are potentially consistent with rational expectations. Under certain conditions, stock return equals investment return, which depends only on firm characteristics, corporate policies, and events. This single equation can qualitatively explain many anomalies including the associations of average stock returns with market-to-book, investment and disinvestment, equity offerings, tender offers and stock repurchases, dividend omissions and initiations, expected profitability, profitability, and earnings surprises. The equation also gives rise to a purely characteristic-based, testable expected-return model. I conclude that neoclassical economics is a good start to understand the driving forces behind the cross section of returns.

*Address correspondence to: Carol Simon Hall 3-160B, William E. Simon Graduate School of Business Administration, University of Rochester, 500 Wilson Boulevard, Rochester, NY 14627. Tel: (585)275-3491, fax: (585)273-1140, and email: zhanglu@simon.rochester.edu.

†I acknowledge helpful comments from Rui Albuquerque, Mike Barclay, Nick Barberis, Jonathan Berk (Utah discussant), Mark Bils, Robert Bloomfield, Jim Booth, Murray Carlson, V. V. Chari, Jason Chen, John Cochrane (NBER discussant), Martijn Cremers, Murray Frank, Will Goetzmann, Bob Goldstein, Joao Gomes, Jeremy Greenwood, Christopher Hennessy, Zvi Hercowitz, Patrick Kehoe, Narayana Kocherlakota, Leonid Kogan, Pete Kyle, Owen Lamont, Erica Li, John Long, Ellen McGrattan, Lionel McKenzie, Roni Michaely, Martin Schneider, Bill Schwert, Matt Spiegel, Hans Stoll, Jerry Warner, David Weinbaum (Cornell discussant), Amir Yaron, and seminar participants at W. P. Carey School of Business of Arizona State University, Johnson School of Management at Cornell University, Fisher College of Business at Ohio State University, MIT Sloan School of Management, Haas School of Business at University of California at Berkeley, Department of Economics and Carlson School of Management at University of Minnesota, Research Department at Federal Reserve Bank of Minneapolis, Kenan-Flagler Business School at University of North Carolina at Chapel Hill, Department of Economics and Simon School of Business at University of Rochester, School of Business at University of Wisconsin at Madison, Owen Graduate School of Management at Vanderbilt University, Yale School of Management, the NBER Asset Pricing meeting in Fall 2004, and the Utah Winter Finance Conference in 2005. I am solely responsible for all the remaining errors.
“Prior to the transformation, macroeconomics was largely separate from the rest of economics. Indeed, some considered the study of macroeconomics fundamentally different and thought there was no hope of integrating macroeconomics with the rest of economics, that is, with neoclassical economics.”


1 Introduction

A huge body of empirical literature in financial economics has documented relations of future stock returns with firm characteristics, corporate policies, and events, relations called anomalies because they seem impossible to explain using traditional asset pricing models (e.g., Fama (1998) and Schwert (2003)). Many experts feel that these anomalies are strong evidence against efficient markets with rational expectations (e.g., Shleifer (2000), Hirshleifer (2001), Daniel, Hirshleifer, and Teoh (2002), Barberis and Thaler (2003), and Ritter (2003)).

I construct the neoclassical microeconomic foundation for time-varying expected returns in the cross section with rational expectations. If the operating-profit and the adjustment-cost functions are homogeneous of the same degree, stock return equals investment return. And the investment return is in turn directly tied with firm characteristics, corporate policies, and events through the first principle of optimal investment.

By simply signing the partial derivatives of investment returns, I show analytically that neoclassical economics can qualitatively explain most anomalies that have often been interpreted as overreaction and underreaction in inefficient markets (e.g., Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hong and Stein (1999)). The list of anomalies covers a broad scope of the literature on empirical asset pricing, empirical corporate finance, and capital markets research in accounting.

The investment anomaly  The investment-to-asset ratio is negatively correlated, but disinvestment-to-asset is positively correlated with future stock returns. The magnitude of this anomaly is higher in firms with higher operating income-to-asset ratios.
The value anomaly The market-to-book ratio is negatively correlated with future stock returns, and the magnitude of this correlation is higher in small firms.

The seasoned-equity-offering anomaly When firms raise capital through seasoned equity, they earn lower average returns in the subsequent years than nonissuing firms with similar characteristics. The magnitude of this anomaly is higher in small firms.

The payout anomaly When firms distribute capital through tender offers, share repurchases, or cash dividends, they earn higher average returns in the subsequent years than non-distributing firms. The magnitude of this anomaly is higher in value firms.

The expected-profitability anomaly Expected profitability correlates positively with future stock returns, and this correlation is higher in small firms.

The profitability anomaly Controlling for market valuation ratios, profitability correlates positively with future stock returns, and this correlation is higher in small firms.

The post-earnings-announcement drift Firms that have experienced positive earnings surprises earn higher average returns than firms that have experienced negative earnings surprises. The magnitude of this correlation is higher in small firms.

In a nutshell, I show that, much like aggregate expected returns that vary over business cycles (e.g., Campbell and Cochrane (1999)), expected returns in the cross section vary with firm characteristics, corporate policies, and events. I achieve this task using the neoclassical framework in the spirit of Kydland and Prescott (1982) and Long and Plosser (1983).

Intuitively, the investment return from time $t$ to $t+1$ equals the ratio of the marginal profit of investment at $t+1$ divided by the marginal cost of investment at $t$. For instance, in a two-period example, expected returns can be written as (see Section 3.2 for derivations):

$$E_t[r_{t+1}] = \frac{E_t[\Pi_{t+1}] + 1 - \delta}{1 + a(I_t/K_t)}$$

where $E_t[r_{t+1}]$ is the expected return, $\frac{I_t}{K_t}$ is the investment-to-asset ratio, $E_t[\Pi_{t+1}]$ is the expected marginal product of capital, and $0 < \delta < 1$ and $a > 0$ are constant parameters.

This equation suggests two economic mechanisms that are potential drivers of anomalies. The first four anomalies can be interpreted through the optimal-investment mechanism.
Specifically, the denominator of equation (1) implies a negative relation between expected returns and investment-to-asset, controlling for the expected marginal product of capital. Intuitively, using the capital-budgeting logic in Brealey, Myers, and Allen (2006, Chapter 6), investment demand increases with the net present value of capital, and all else equal, the net present value decreases with costs of capital (i.e., expected returns). Controlling for expected cash flows, high costs of capital imply low net present value, which in turn implies low investment demand. And low costs of capital imply high net present value, which in turn implies high investment demand.

Figure 1 illustrates the downward-sloping investment-demand function. Its negative slope suggests that expected returns decrease with positive investment but increase with the magnitude of disinvestment, i.e., the investment anomaly. Further, similar to high investment-to-asset firms, growth firms, issuing firms, and low payout firms are distributed towards the right end of the curve and are associated with low expected returns. And similar to low investment-to-asset firms, value firms, nonissuing firms, and high payout firms are distributed towards the left end of the curve and are associated with high expected returns.

The intuition is simple. In the model, investment-to-asset is an increasing function of marginal $q$, the present value of future marginal profits of an additional unit capital. And the marginal $q$ is in turn proportional to the average $Q$, a close cousin of market-to-book. The negative investment-return relation then implies a negative relation between expected returns and market-to-book, i.e., the value anomaly. The payout anomaly follows because the balance-sheet constraint of firms, which equates the sources of funds with the uses of funds, implies a negative relation between payout and capital investment. If a firm distributes more capital back to shareholders, the firm is likely to invest less. The underperformance following seasoned equity offerings follows because the balance-sheet constraint implies a positive relation between equity finance and capital investment, i.e., if a firm raises more capital, the firm is likely to invest more.

With decreasing returns to scale or strictly convex adjustment costs, the relation between expected returns and market-to-book is convex. This convexity manifests itself, by the chain rule of partial derivatives, as the stronger value anomaly in small firms, the stronger seasoned-
equity-offering anomaly in small firms, and the stronger payout anomaly in value firms.

And I also use the investment-return equation to interpret the three earnings-related anomalies. Specifically, the marginal product of capital in the numerator of equation (1) is proportional to profitability. Controlling for market valuation ratios, which are essentially the denominator of the investment return, expected profitability should therefore correlate positively with expected returns. This positive correlation in turn explains the profitability anomaly. The reason is that profitability is a strong, positive predictor of future profitability, and is therefore an indispensable component of expected profitability. Moreover, firms that have experienced positive earnings surprises are likely to be more profitable than firms that have experienced negative earnings surprises. Earnings surprises should therefore correlate positively with expected returns, as in the post-earnings-announcement drift.

Intriguingly, these explanations of anomalies do not involve risk, at least directly, even though the model is entirely rational. The reason is that I derive expected returns from the optimality conditions of firms, instead of consumers. Consequently, the stochastic discount factor and its covariances with stock returns do not directly enter the expected-return
determination. *Characteristics are sufficient statistics for expected returns!* The debate on covariances versus characteristics in empirical finance (e.g., Daniel and Titman (1997, 1998) and Davis, Fama, and French (2000)) is therefore not a well-defined question.

The analytical link between stock and investment returns also suggests a new empirical model for expected stock returns as expected investment returns constructed from observable firm characteristics. Although consistent with factor models and consumption-based models in theory, this new investment-based expected-return model is likely to have advantages over traditional asset pricing models in practice. All three strands of models are subject to potential misspecification errors. But the investment-based model avoid two important hurdles specific to traditional models. First, because of difficulty in estimating time-varying loadings and risk premiums, estimated costs of equity from factor-pricing models are extremely imprecise even at the industry level (e.g., Fama and French (1997)). Second, consumption-based models must deal with potentially severe measurement errors in available aggregate consumption data (e.g., Ferson and Harvey (1992), Wilcox (1992), and Heaton (1995)).

The insight that stock and investment returns are equal appears first in Cochrane (1991). Cochrane (1991, 1996) also launches the investment-based asset pricing literature. Restoy and Rockinger (1994) formally establish this equivalence under linear homogeneity. An early version of Gomes, Yaron, and Zhang (2005) extends the result with debt financing. I extend the result under the condition that the operating-profit and adjustment-cost functions are homogeneous of the same degree. Most important, these studies largely focus on aggregate investment returns. But I aim directly at anomalies in the cross section of returns.

The neoclassical investment theory is originated by Brainard and Tobin (1968) and Tobin (1969). Hayashi (1982) establishes the equivalence between marginal $q$ and average $Q$ under linear homogeneity. Abel and Eberly (1994) extend this result into a stochastic setting with partial irreversibility and fixed costs proportional to capital. Abel and Eberly also show that marginal $q$ is proportional to average $Q$ when the operating-profits and the adjustment-cost functions are homogeneous of the same degree, a result I use extensively. This theory has been used mostly to explain the behavior of investment. But I open the door for its large-scale applications in the cross section of returns.

Section 2 sets up the model. Section 3 uses the model to explain anomalies. Section 4 interprets the results. Section 5 concludes. Appendix A contains all the proofs. And Appendix B briefly reviews the anomalies literature that motivates the theoretical work in this paper.

2 The Model

The model is heavily influenced by Abel and Eberly (1994) and Gomes, Yaron, and Zhang (2005). Section 2.1 describes the environment. Section 2.2 characterizes value-maximization.

2.1 The Environment

Consider a firm that uses capital and a vector of costlessly adjustable inputs, such as labor, to produce a perishable output. The firm chooses the levels of these inputs each period to maximize its operating profit, defined as its revenue minus the expenditures on these inputs. Taking the operating profit as given, the firm then chooses optimal investment to maximize its market value. And investment involves capital adjustment costs.

The Operating-Profit Function

Let $\Pi_t = \Pi(K_t, X_t)$ denote the maximized operating profit at time $t$, where $K_t$ is the capital stock at time $t$ and $X_t$ is a vector of random variables representing exogenous shocks to the operating profit, such as aggregate and firm-specific shocks to production technology, shocks
to the prices of costlessly adjusted inputs, and aggregate- and firm-specific shocks to the
demand of the output produced by the firm.

**Assumption 1** The operating profit function is homogeneous of degree $\alpha$ with $\alpha \leq 1$:

$$\Pi(K_t, X_t) = \Lambda(X_t)K_t^\alpha \quad \text{where} \quad \Lambda(X_t) > 0$$  \hspace{1cm} (2)

If $\alpha = 1$, the operating-profit function displays linear homogeneity in $K_t$. This applies to
a competitive firm that is a price-taker in output and factor markets. When $\alpha < 1$, the
production technology exhibits decreasing return to scale.

From Assumption 1,

$$\alpha\Pi(K_t, X_t) = \Pi_1(K_t, X_t)K_t$$  \hspace{1cm} (3)

where the marginal product of capital, $\Pi_1(K_t, X_t)$, is strictly positive, with subscript $i$
denoting the first-order partial derivative with respect to the $i^{th}$ argument. Multiple
subscripts denote partial derivatives of higher orders. Moreover, $\Pi_{11}(K_t, X_t) \leq 0$. The
inequality is strict when $\alpha < 1$. Finally, $\Pi_{111}(K_t, X_t) \geq 0$.

From equation (3), the marginal product of capital, $\Pi_1$, is proportional to the average
product of capital, $\frac{\Pi(K_t, X_t)}{K_t}$. This ratio corresponds roughly to accounting profitability, the
ratio of earnings-to-book equity, plus depreciation rate. The reason is that the operating
profit in the model corresponds roughly to earnings plus capital depreciation in the data. I
assume implicitly that accruals in practice are only used to mitigate the accounting timing
and matching problems that deviate operating cash flow from earnings (e.g., Dechow (1994)).
Accounting problems such as earnings management are abstracted away from the model.

**The Augmented Adjustment-Cost Function**

Capital accumulates according to:

$$K_{t+1} = I_t + (1 - \delta)K_t$$  \hspace{1cm} (4)

Thus the end-of-period capital equals capital investment plus the beginning-of-period capital
net of depreciation. And capital depreciates at a fixed proportional rate of $\delta$.  

---  

8
When firms invest, they incur costs because of: (i) purchase/sale costs, (ii) convex costs of physical adjustment, and (iii) weakly convex costs of raising capital when the sum of the purchase/sale and physical adjustment costs is higher than the operating profit.

(i) Purchase/sales costs are incurred when the firm buys or sells uninstalled capital. When the firm disinvests, this cost is negative. For analytical convenience, I assume that the relative purchase price and relative sale price of capital are both equal to unity. Abel and Eberly (1994) instead assume that purchase price is higher than sale price to capture costly reversibility. Consequently, the purchase/sale cost function is not differentiable at $I_t = 0$. My assumption retains this differentiability. And one can still capture costly reversibility by letting the convex costs of disinvestment be uniformly higher than the convex costs of investment with equal magnitudes as in Hall (2001) and Zhang (2004).

(ii) Convex costs of physical adjustment are nonnegative costs that are zero when $I_t = 0$. These costs are continuous, strictly convex in $I_t$, non-increasing in capital $K_t$, and differentiable with respect to $I_t$ and $K_t$ everywhere. The second-order partial derivative of the convex-cost function with respect to $K_t$ is nonnegative. The commonly-used quadratic adjustment-cost function satisfies all these assumptions.

(iii) Costs of raising capital are incurred when the financial deficit, denoted $O_t$, is strictly positive. I define $O_t$ as the higher value between zero and the sum of the purchase/sale costs and convex costs of adjustment minus the operating profit. I assume that the financing-cost function is continuous, weakly convex in $O_t$ (and hence in $I_t$) and decreasing in $K_t$. Its first-order partial derivative with respect to $O_t$ (and hence with respect to $I_t$) is zero when $O_t = 0$. The financing-cost function is differentiable with respect to $O_t$ (and hence with respect to $I_t$) and $K_t$ everywhere. And the second-order partial derivative of the function with respect to $K_t$ is nonnegative. Parametric examples including the quadratic financing-cost function, defined as $\frac{b}{2} \left( \frac{O_t}{K_t} \right)^2 K_t$ with $b > 0$, and the proportional financing-cost function used in Gomes (2001) and Hennessy and Whited (2005) satisfy these assumptions.

The flip side of financial deficit is free cash flow, denoted $C_t$. I define $C_t$ as the higher value between zero and the operating profit minus the sum of the purchase/sale costs and the convex costs of adjustment. I assume that whenever $C_t$ is strictly positive, firms pay it
back to shareholders either in the form of dividends or stock repurchases. The model is silent on the form of payout. And I do not model debt or corporate cash hoarding because these ingredients do not seem essential for understanding the anomalies listed in Introduction. Further, firms do not pay any extra costs when distributing capital, implying that firms either raise or distribute capital, but never at the same time.

The total costs of investment represent the sum of purchase/sale costs, convex costs of physical adjustment, and costs of raising capital. I denote the total costs as $\Phi(I_t, K_t)$, called the augmented adjustment-cost function.

**Assumption 2** The augmented adjustment-cost function $\Phi(I_t, K_t)$ satisfies:

$$\begin{align*}
\Phi_2(I_t, K_t) &\leq 0; \\
\Phi_{23}(I_t, K_t) &\geq 0; \\
\Phi_{11}(I_t, K_t) &> 0;
\end{align*}$$

The most important technical assumption is stated explicitly below:

**Assumption 3** The augmented adjustment-cost function is homogeneous of the same degree, $\alpha$, in $I_t$ and $K_t$, as the operating-profit function is in $K_t$, i.e.,

$$\Phi(I_t, K_t) = G \left( \frac{I_t}{K_t} \right) K_t^\alpha \quad (5)$$

Combined with Assumption 2, Assumption 3 implies that $G''(\cdot) > 0$ and that

$$\alpha \Phi(I_t, K_t) = \Phi_1(I_t, K_t) I_t + \Phi_2(I_t, K_t) K_t \quad (6)$$

I need Assumption 3 to establish the equivalence between stock and investment returns (see the proof of Proposition 2 in Appendix A). How restrictive is Assumption 3? Abel and Eberly (1994) discuss its content for the case of linear homogeneity. The linear homogeneity of $\Phi(I_t, K_t)$ means that each of its three components is linearly homogenous. (i) A doubling of $I_t$ doubles the purchase/sale costs that are linear in $I_t$, and are independent of $K_t$. (ii) The investment literature routinely assumes that physical adjustment costs are linearly homogenous (e.g., Hayashi (1982), Abel and Blanchard (1983), and Abel and Eberly (1994)). And (iii) the proportional and quadratic financing-cost functions are linearly homogeneous in $I_t$ and $K_t$.

Relative to the specification in Abel and Eberly (1994), my augmented adjustment-cost
function adds the convex costs of financing, but ignores the wedge between purchase and sale prices of capital and fixed costs of adjustment. The fixed costs of raising capital are not included either. Incorporating these features will compromise the differentiability of \( \Phi(I_t, K_t) \) with respect to \( I_t \) at the two points where \( I_t = 0 \) and \( O_t = 0 \). My results below go through almost everywhere but at these two points (with a measure of zero) where the investment return is not well-defined because \( \Phi_1 \) does not exist.

More important, including the wedge between the purchase and sale prices of capital and fixed costs of investment and raising capital leaves the crucial Assumption 3 unaltered. As argued in Abel and Eberly (1994), the purchase/sale costs are proportional to \( I_t \). And the fixed costs are linearly homogenous in \( K_t \), if these costs reflect the costs of interrupting production, and are therefore proportional to the operating profit and to capital.

Finally, it is ultimately an empirical question how restrictive Assumption 3 is. But I note that the special case of \( \alpha = 1 \) seems standard in the empirical investment literature (e.g., Whited (1992), Hubbard (1998) and Erickson and Whited (2000)). And several numerically solved models including Kogan (2004), Zhang (2005), Cooper (2006), and Gala (2006) yield qualitatively similar results on the value anomaly as my analytical results.

### 2.2 Dynamic Value Maximization

The dynamic value-maximization problem of firms is given by:

\[
V(K_t, X_t) = \max_{\{I_{t+j}, K_{t+j}\}_{j=0}^{\infty}} E_t \left[ \sum_{j=0}^{\infty} M_{t+j} \left( \Pi(K_{t+j}, X_{t+j}) - \Phi(I_{t+j}, K_{t+j}) \right) \right]
\]  

(7)

where \( V(K_t, X_t) \) is the cum-dividend market value because when \( j = 0 \), \( \Pi(K_t, X_t) - \Phi(I_t, K_t) \) is included in \( V(K_t, X_t) \). \( M_{t+j} > 0 \) denotes the stochastic discount factor from time \( t \) to \( t+j \).

**Marginal \( q \), Tobin’s (Average) \( Q \), and Market-to-Book**

**Lemma 1** Under Assumptions 1 and 3, the value function is homogenous of degree \( \alpha \), i.e.,

\[ \alpha V(K_t, X_t) = V_1(K_t, X_t) K_t. \]

Define Tobin’s \( Q \) as \( \tilde{Q}_t = \frac{V(K_t, X_t)}{K_t} \), then \( V_1(K_t, X_t) = \alpha \tilde{Q}_t. \)
Let \( q_t \) be the present-value multiplier associated with capital accumulation equation (4). The Lagrange formulation of the firm value, \( V(K_t, X_t) \), is then:

\[
\max_{\{I_t+j, K_{t+j+1}\}_{j=0}^{\infty}} E_t \left[ \sum_{j=0}^{\infty} M_{t+j} (\Pi(K_{t+j}, X_{t+j}) - \Phi(I_{t+j}, K_{t+j}) - q_{t+j}[K_{t+j+1} - (1 - \delta)K_{t+j} - I_{t+j}]) \right] 
\]

(8)

The first-order conditions with respect to \( I_t \) and \( K_{t+1} \) are, respectively,

\[
q_t = \Phi_1(I_t, K_t) 
\]

(9)

\[
q_t = E_t[M_t+1[\Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1}) + (1 - \delta)q_{t+1}]] 
\]

(10)

Solving equation (10) recursively yields an economic interpretation for marginal \( q \):

**Lemma 2** Marginal \( q \) is the expected present value of marginal profits of capital:

\[
q_t = E_t \left[ \sum_{j=1}^{\infty} M_{t+j}(1 - \delta)^{j-1}(\Pi_1(K_{t+j}, X_{t+j}) - \Phi_2(I_{t+j}, K_{t+j})) \right] 
\]

(11)

**Proposition 1** (Marginal \( q \) and Market-to-Book) Define the ex-dividend firm value as:

\[
P_t \equiv P(K_t, K_{t+1}, X_t) = V(K_t, X_t) - \Pi(K_t, X_t) + \Phi(I_t, K_t) 
\]

(12)

And define the market-to-book equity as \( Q_t \equiv \frac{P_t}{K_{t+1}} \), then

\[
q_t = \alpha Q_t 
\]

(13)

under Assumptions 1 and 3.\(^1\)

Some useful properties of \( \Phi(I_t, K_t) \) evaluated at the optimum can now be established.

**Lemma 3** Under Assumptions 1 and 2, the augmented adjustment-cost function \( \Phi(I_t, K_t) \), when evaluated at the optimum, satisfies \( \Phi_1(I_t, K_t) > 0, \Phi_{12}(I_t, K_t) \leq 0, \) and \( \Phi_{122}(I_t, K_t) \geq 0. \)

\(^1\)In continuous time, Abel and Eberly (1994) show that marginal \( q_t \) is proportional to Tobin’s average \( Q_t \), i.e., \( q_t = \alpha Q_t \). But in discrete time, \( V_t(K_t, X_t) \) is not exactly marginal \( q_t \). The time-to-build convention embedded in equation (4) implies that one unit of investment today only becomes effective next period. Consequently, \( q_t \) and \( Q_t \) are linked through:

\[
q_t = \alpha E_t[M_{t+1}Q_{t+1}] 
\]

(14)

To see this, note the derivative of equation (8) with respect to \( K_t \) is \( V_t(K_t, X_t) = \Pi_1(K_t, X_t) - \Phi_2(I_t, K_t) + q_t(1 - \delta) \). Equation (14) then follows from Lemma 1 and equation (10).
Investment and Stock Returns

Combining the first-order conditions in equations (9) and (10) yields:

$$E_t[M_{t+1}\, r_{t+1}^I] = 1$$

(15)

where $r_{t+1}^I$ denotes the investment return:

$$r_{t+1}^I = \frac{\Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1}) + (1 - \delta)\Phi_1(I_{t+1}, K_{t+1})}{\Phi_1(I_t, K_t)}$$

(16)

Equation (16) is very intuitive. $r_{t+1}^I$ is the ratio of the marginal benefit of investment at time $t+1$ divided by the marginal cost of investment at time $t$. The denominator, $\Phi_1(I_t, K_t)$, is the marginal cost of investment. By optimality, it equals the marginal $q_t$, the expected present value of marginal profits of investment. In the numerator of equation (16), $\Pi_1(K_{t+1}, X_{t+1})$ is the extra operating profit from the extra capital at $t+1$; $-\Phi_2(I_{t+1}, K_{t+1})$ captures the effect of extra capital on the augmented adjustment cost; and $(1 - \delta)\Phi_1(I_{t+1}, K_{t+1})$ is the expected present value of marginal profits evaluated at time $t+1$, net of depreciation.

**Proposition 2 (Stock Returns and Firm Characteristics)** Define stock return as:

$$r_{t+1}^S = \frac{P_{t+1} + \Pi(K_{t+1}, X_{t+1}) - \Phi(I_{t+1}, K_{t+1})}{P_t}$$

(17)

Then $E_t[M_{t+1}\, r_{t+1}^S] = 1$. Under Assumptions 1 and 3, stock return equals investment return:

$$r_{t+1}^S = r_{t+1}^I$$

(18)

Given this equivalence, I will use the common notation $r_{t+1}$ to denote both returns.

The analytical link between stock and investment returns subsists under much more general conditions than Proposition 2 allows. The two-return equivalence holds even under variable capacity utilization, flow operating costs, irreversible investment, and dividend constraints. With multiple capital goods, the stock return equals the value-weighted average of individual investment returns, each of which corresponds to a different type of capital. With debt financing, the investment return equals the leverage-weighted average of stock and bond returns. But the analytical link between the stock and investment returns breaks down...
when capital requires more than one period to build. Details are available upon requests.

A Characteristic-Based Expected-Return Model

Ex ante, Proposition 2 implies that expected stock returns equal expected investment returns. This ex-ante restriction can be tested using the following moment conditions via GMM:

\[
E \left[ \left( r^S_{t+1} - \frac{\Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1}) + (1 - \delta)\Phi_1(I_{t+1}, K_{t+1})}{\Phi_1(I_t, K_t)} \right) \otimes Z_t \right] = 0 \quad (19)
\]

where \( Z_t \) is a vector of portfolio-level and aggregate-level instrumental variables known at time \( t \). \( r^S_{t+1} \) are stock returns of portfolios sorted by the anomaly-related variables. The empirical investment literature has provided guidance on the parametrization of the operating-profit function, \( \Pi \), and the adjustment-cost functions, \( \Phi \). With estimated parameters, expected stock returns can then be constructed from observable firm characteristics through the expected investment returns.\(^2\) The investment-return equation can also be extended to incorporate more realistic ingredients such as leverage, financial constraints, and labor adjustment costs. Even empire-building can be incorporated by assuming managers derive private benefits proportional to the operating profit (e.g., Stulz (1990)).

More important, to explain anomalies quantitatively, it is sufficient to implement the test outlined in equation (19) at the portfolio level. And smooth specifications of the augmented adjustment-cost function should be sufficient because capital investment at the portfolio level is well-behaved. At the firm level, where lumpy investment is more common (e.g., Doms and Dunne (1998)), non-convex adjustment-cost functions must be introduced. But non-convexities seem less relevant for asset pricing tests that are often done at the portfolio level.

The test outlined in equation (19) is different from existing investment-based asset pricing tests in Cochrane (1996), Gomes, Yaron, and Zhang (2005), and Whited and Wu (2005). Cochrane and Gomes et al. parameterize the stochastic discount factor as a linear function of aggregate investment returns constructed from macroeconomic variables. But portfolio-level characteristics are absent. White and Wu test equation (15) by assuming \( M_{t+1} \) as a linear combination of the Fama-French (1993) factors and by constructing firm-level

\(^2\)Cochrane (1991) implicitly tests the moment condition (19) by comparing the properties of stock and investment returns at the aggregate level. Whited and Zhang (2006) carry out this test at the portfolio level.
investment returns. But stock returns are absent. By linking stock returns directly to firm characteristics, the test in equation (19) can evaluate the quantitative importance of the economic mechanisms proposed in Section 3.

3 Explaining Anomalies

Proposition 2 provides an analytical link between expected returns and firm characteristics, a link that can serve as a microeconomic foundation for asset pricing anomalies.

Developing this microfoundation is the heart of this paper. Section 3.1 discusses the general methodology of characteristic-based expected-return determination and contrasts it with the traditional risk-based determination. Section 3.2 illustrates the basic intuition using two canonical examples. Section 3.3 extends the intuition into the general neoclassical framework.

3.1 Methodology: Characteristic-Based Asset Pricing

My analytical method is simple. It basically amounts to taking and signing partial derivatives of the expected investment return given in equation (16) with respect to various anomaly-related variables. Using partial derivatives is reasonable because to establish a new anomaly, empiricists often control for other known anomalies, a practice corresponding naturally to partial derivatives. For example, Chan, Jegadeesh, and Lakonishok (1996) and Haugen and Baker (1996) control for market valuation ratios when documenting the earnings momentum and profitability anomalies, respectively. Cochrane (1991, 1996) uses similar techniques to explain the return-investment relations. Fama and French (2006) also use similar techniques to derive testable hypotheses from the Ohlson (1995) residual-income valuation model (see also Pástor and Veronesi (2003, 2005)).

A caveat is in order. Empirical controls are often done using linear regression models, but taking partial derivatives in the model controls for nonlinear effects as well. These two control methods are therefore equivalent only to a first-order approximation. Evaluating the quantitative effects of potential deviations is left for future research, however.

As a fundamental departure from the traditional risk-based approach, I argue that expected returns can be determined entirely from firm characteristics through the first principle
of optimal investment without any information about the stochastic discount factor! The reason is that the stochastic discount factor, $M_{t+1}$, and its covariances with returns do not enter the equation for expected investment returns. Characteristics are sufficient statistics for expected returns. Investment-based asset pricing can be developed independently from consumption-based asset pricing. This approach is foretold by Rubinstein (2001, p.23): “For the most part, financial economists take the stochastic process of stock prices, the value of the firms, or dividend payments as primitive. But to explain some anomalies, we may need to look deeper into the guts of corporate decision making to derive what the processes are.”

Of course, this practice only means that the effect of $M_{t+1}$ is indirect, not irrelevant. If $M_{t+1}$ is a constant, $M$, then equation (15) implies that the expected return $E_t[r_{t+1}] = \frac{1}{M}$, a constant uncorrelated with firm characteristics. And if the correlation between $M_{t+1}$ and $X_{t+1}$ is zero, i.e., the operating profits of firms are unaffected by aggregate shocks, then equation (15) implies that $E_t[r_{t+1}] = r_f$, where $r_f \equiv \frac{1}{E_t[M_{t+1}]}$ is the risk-free rate. In this case, there is no cross-sectional variation in expected returns. The analysis below in effect provides time-series correlations between the risk-free rate and firm characteristics. Because I study expected returns instead of expected excess returns, I need not specify $M_{t+1}$ that is necessary to determine the risk-free rate. Equivalently, I do not need to restrict the correlation between $M_{t+1}$ and $X_{t+1}$, which is necessary to determine expected excess returns.

More important, the characteristic-based approach is internally consistent with the traditional risk-based approach. From Proposition 2, $E_t[M_{t+1}r_{t+1}^S] = 1$. Following Cochrane (2001, p. 19), this equation can be rewritten as the beta-pricing model, i.e., $E_t[r_{t+1}^S] = r_f + \beta_t \lambda_{Mt}$, where $\beta_t \equiv \frac{-Cov_t[r_{t+1}^S,M_{t+1}]}{Var_t[M_{t+1}]}$ is the amount of risk, and $\lambda_{Mt} \equiv \frac{Var_t[M_{t+1}]}{E_t[M_{t+1}]}$ is the price of risk. But Proposition 2 also says that $E_t[r_{t+1}^S] = E_t[r_{t+1}^I]$ where the right-hand side only depends on characteristics from equation (16). And $E_t[r_{t+1}^I] = E_t[r_{t+1}^S] = r_f + \beta_t \lambda_{Mt}$, implies that $\beta_t = \frac{E_t[r_{t+1}^I]-r_f}{\lambda_{Mt}}$, which provides a one-to-one mapping between covariances and characteristics. But apart from this mechanical link, risk only plays a secondary role in the background of the characteristic-based determination of expected returns.
3.2 Intuition in Two Canonical Examples

I construct two canonical examples to illustrate the basic intuition. In the first example, the only costs of investment are linear purchase/sale costs, i.e., \( \Phi(I_t, K_t) = I_t \). And in the second example, there are also quadratic costs of physical adjustment, i.e.,

\[
\Phi(I_t, K_t) = I_t + a \left( \frac{I_t}{K_t} \right)^2 K_t \quad \text{where} \quad a > 0
\]

Using a list of examples with progressive degrees of sophistication crystalizes the model ingredients that are essential for explaining different anomalies.

The Linear-Purchase/Sale-Cost Example

This example can generate the earnings-related anomalies. Intuitively, the marginal product of capital at time \( t+1 \) in the numerator of the investment-return equation is closely related to accounting profitability. Accordingly, expected return increases with expected profitability.

Specifically, when \( \Phi(I_t, K_t) = I_t \), equation (16) implies that:

\[
E_t[r_{t+1}] = E_t[\Pi_1(K_{t+1}, X_{t+1})] + (1 - \delta)
\]  

(21)

The expected net return therefore equals the expected marginal product of capital minus the depreciation rate. Further, because earnings equals operating cash flows minus the capital depreciation that is the only accruals in the model, let \( N_t = \Pi_t - \delta K_t \).

Combining equations (3) and (21) now generates the expected-profitability anomaly:

\[
E_t[r_{t+1}] = E_t\left[ \frac{\Pi_{t+1}}{K_{t+1}} \right] + (1 - \delta) = E_t\left[ \frac{N_{t+1}}{K_{t+1}} \right] + 1
\]  

(22)

In other words, expected return is expected profitability!

The profitability anomaly then arises because profitability is persistent. High profitability implies high expected profitability, which in turn implies high expected returns.

Assumption 4 The operating profit-to-capital ratio follows:

\[
\frac{\Pi_{t+1}}{K_{t+1}} = \pi (1 - \rho_x) + \rho_x \left( \frac{\Pi_t}{K_t} \right) + \epsilon_{t+1}^\pi
\]

(23)
where $\pi > 0$ and $0 < \rho_\pi < 1$ are the long-run average and the persistence of operating profit-to-capital, respectively. And $\epsilon_{t+1}^\pi$ is a normal random variable with a zero mean.

Because operating profit-to-capital equals profitability plus the constant depreciation rate, Assumption 4 basically captures the persistence in profitability. Equation (23) becomes:

$$\frac{N_{t+1}}{K_{t+1}} = (\pi - \delta)(1 - \rho_\pi) + \rho_\pi \left( \frac{N_t}{K_t} \right) + \epsilon_{t+1}^\pi$$

(24)

where the first two terms capture expected profitability and $\epsilon_{t+1}^\pi$ captures earnings surprise. There is much evidence on the persistence in profitability (e.g., Fama and French (1995, 2000, 2006)). In particular, Fama and French (2006) report that the current-period profitability is the strongest predictor of profitability one to three years ahead.

Combining equations (22) and (24) yields:

$$E_t[r_{t+1}] = (\pi - \delta)(1 - \rho_\pi) + \rho_\pi \left( \frac{N_t}{K_t} \right) + 1$$

(25)

The expected return therefore increases with profitability. And a new testable hypothesis arises, i.e., the magnitude of the profitability anomaly should increase with the persistence of profitability. More important, the first-order autoregressive process for profitability is not important: more complex specifications will give essentially the same economic insights.

The same economic mechanism also drives the post-earnings-announcement drift in the model. Earnings surprises and profitability are both scaled earnings, and should contain similar information on future profitability. Intuitively, firms that have experienced positive earnings surprises are likely to be more profitable than firms that have experienced negative earnings surprises. Because earnings surprises are likely to be a major component of expected profitability, earnings surprises should correlate positively with expected returns.

Specifically, plugging the one-period-lagged equation (24) into equation (25) yields:

$$E_t[r_{t+1}] = (\pi - \delta)(1 - \rho_\pi)(1 + \rho_\pi) + \rho_\pi^2 \frac{N_{t-1}}{K_{t-1}} + \rho_\pi \epsilon_t + 1$$

(26)

\[3\]The empirical measure of earnings surprises is Standardized Unexpected Earnings (SUE). See, for example, Chan, Jegadeesh, and Lakonishok (1996). For stock $i$ in month $t$, $\text{SUE}_{it} \equiv \frac{e_{iq} - e_{iq-4}}{\sigma_{it}}$, where $e_{iq}$ is quarterly earnings per share most recently announced as of month $t$ for stock $i$, $e_{iq-4}$ is earnings per share four quarters ago, and $\sigma_{it}$ is the volatility of unexpected earnings, $e_{iq} - e_{iq-4}$, over the preceding eight quarters.
The expected return thus has a positive loading, $\rho_\pi$, on the current-period earnings surprise, $\varepsilon_t$. A new testable hypothesis arises, i.e., the magnitude of the post-earnings announcement drift should increase with the average persistence of firm-level profitability in the sample.

Although somewhat useful to interpret the earnings anomalies, this linear-cost example has important limitations. First, the inverse relation between the magnitude of the earnings anomalies and the market value cannot be generated. From equations (22) and (25), the loadings of expected returns on both expected profitability and profitability are constant, independent of the market value. Second, the example cannot generate the value anomaly either because $\Phi(I_t, K_t) = I_t$ implies that $Q_t = q_t = \Phi_1(I_t, K_t) = 1$: firms do not differ in market-to-book. Third, substituting $K_{t+1} = \left(\frac{K_t}{K_t} + (1 - \delta)\right) K_t$ into equation (22) and differentiating yield $\frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} = E_t[\Pi_{11}(K_{t+1}, X_{t+1})]K_t = 0$. Expected returns are independent of investment-to-asset, and therefore also independent of payout-to-asset and new equity-to-asset.

All these limitations can be overcome by introducing the quadratic adjustment costs as in equation (20) into the model. Convex capital-adjustment costs are therefore necessary to generate all the anomalies other than the first-order earnings-return relations.

**The Quadratic-Adjustment-Cost Example**

Combining equations (16) and (20) yields:

$$r_{t+1} = \frac{\Pi_1(K_{t+1}, X_{t+1}) + (a/2)(I_{t+1}/K_{t+1})^2 + (1 - \delta)[1 + a(I_{t+1}/K_{t+1})]}{1 + a(I_t/K_t)} \quad (27)$$

This investment-return equation is in essence the same equation in Cochrane (1991, 1996). Equation (1) in Introduction is the two-period version of equation (27). The two-period version assumes that firms exist in only two periods, $t$ and $t+1$. Firms invest in period $t$, produce in both periods, and liquidate at the end of $t+1$ with an exiting value of $(1 - \delta)K_{t+1}$. Because firms do not invest in $t+1$, the two terms involving $\frac{I_{t+1}}{K_{t+1}}$ drop out from the numerator of equation (27), giving rise to equation (1).

**The Expected-Profitability Anomaly**

Because $\Pi_1(K_{t+1}, X_{t+1}) = \frac{\Pi_{11} \Pi_{12}}{\Pi_{11} + \delta} = \frac{N_{t+1}}{K_{t+1}} + \delta$, taking conditional expectations and differentiating both sides of equation (27) with respect to the expected profitability yield $\frac{\partial E_t[r_{t+1}]}{\partial E_t[N_{t+1}/K_{t+1}]} =$
Controlling for market-to-book, as done in the empirical work of Haugen and Baker (1996) and Chan, Jegadeesh, and Lakonishok (1996), expected returns should therefore increase with expected profitability. More important, because marginal $q_t$ equals market-to-book from Proposition 1, $\frac{\partial E_t[N_{t+1}]}{\partial q_t} = \frac{K_{t+1}}{P_t}$, which is inversely related to the market value, $P_t$, consistent with the evidence in Fama and French (1995). Consequently, the numerator of $\frac{\partial E_t}{\partial q_t}$ remains positive. Formally, equations (11) and (23) imply that $\frac{\partial q_t}{\partial E_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{1}{1+\alpha(I_t/K_t)} > 0$. And it follows that $\frac{\partial E_t}{\partial q_t} = \frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{1}{1+\alpha(I_t/K_t)} > 0$.

The Profitability Anomaly

From equation (23), $\frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \rho_t \frac{\partial E_t[N_{t+1}]}{\partial q_t}$. The logic for the expected-profitability anomaly then implies that $\frac{\partial E_t[I_{t+1}]}{\partial q_t}$ is positive and decreasing in the market value, consistent with the evidence on the inverse relation between the magnitude of the drift and size (e.g., Bernard (1993)).

In sum, the expected-profitability, profitability, and post-earnings-announcement drift are essentially the same phenomenon all driven by the marginal product of capital. And I am not aware of other rational explanations of the earnings anomalies.

The Investment Anomaly

The logic for the profitability anomaly is exactly the same as the logic for the post-earnings-announcement drift because the one-period-lagged equation (24) implies that earnings surprises and profitability contain similar information on future profitability. Relative to the linear-cost example, the incremental contribution of the quadratic costs of adjustment is that the loading of expected returns on contemporaneous earnings surprises, $\frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \rho_t > 0$, and decreases with the market value. This result is consistent with the evidence on the inverse relation between the magnitude of the drift and size (e.g., Bernard (1993)).

This partial derivative corresponds to the case of fixing $I_{t+1}/K_{t+1}$. This practice is only for the ease of exposition. Allowing $I_{t+1}/K_{t+1}$ to vary does not affect the qualitative result because $\frac{\partial (I_{t+1}/K_{t+1})}{\partial q_t} > 0$ in the model. The pattern that more profitable firms invest more is also consistent with the evidence in Fama and French (1995). Consequently, the numerator of $\frac{\partial E_t[I_{t+1}]}{\partial q_t}$ remains positive. Formally, equations (11) and (23) imply that $\frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \rho_t > 0$. And it follows that $\frac{\partial E_t[I_{t+1}]}{\partial q_t} = \frac{\partial E_t[N_{t+1}]}{\partial q_t} = \frac{1}{1+\alpha(I_t/K_t)} > 0$.
The intuition underlying the investment anomaly is simple. In the language of capital budgeting (e.g., Brealey, Myers, and Allen (2006, Chapter 6)), investment-to-asset increases with net present value of capital, which is in turn inversely related to the cost of capital, controlling for expected cash flows. All else equal, higher cost of capital implies lower net present value, which in turn implies lower investment-to-asset. And lower cost of capital implies higher net present value, which in turn implies higher investment-to-asset.

To derive the negative relation between investment-to-asset and expected returns in the quadratic-cost example, let $U_{t+1}^q > 0$ denote the numerator of the investment return in equation (27). Taking conditional expectations and differentiating both sides with respect to $\frac{I_t}{K_t}$ yields:

$$\frac{\partial E[r_{t+1}]}{\partial (I_t/K_t)} = -\frac{aE[U_{t+1}^q]}{[1+a(I_t/K_t)]^2} + \frac{1}{1+a(I_t/K_t)} \frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)}.$$  

to show $\frac{\partial E[r_{t+1}]}{\partial (I_t/K_t)} < 0$, it suffices to show $\frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)} < 0$. But rewriting $I_{t+1}$ in $E_t[U_{t+1}^q]$ as $K_{t+2} - (1 - \delta) \left( \frac{I_t}{K_t} + (1 - \delta) \right) K_t$ and and $K_{t+1}$ as $\left( \frac{I_t}{K_t} + (1 - \delta) \right) K_t$ and differentiating, I have $\frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)} = -\frac{aK_t(E_t[K_{t+2}])^2}{K_{t+1}^2} < 0$.

The downward-sloping investment-demand function is also convex. To this end, differentiating $\frac{\partial E[r_{t+1}]}{\partial (I_t/K_t)}$ once more with respect to $\frac{I_t}{K_t}$ yields $\frac{\partial^2 E[r_{t+1}]}{\partial (I_t/K_t)^2} = -\frac{a}{[1+a(I_t/K_t)]^2} \frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)} + 2aE[U_{t+1}^q] \frac{\partial^2 E[U_{t+1}^q]}{\partial (I_t/K_t)^2} + \frac{1}{1+a(I_t/K_t)} \frac{\partial^2 E[U_{t+1}^q]}{\partial (I_t/K_t)^2} - \frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)} a \frac{\partial^2 E[U_{t+1}^q]}{\partial (I_t/K_t)^2} > 0$, where the inequality follows from $\frac{\partial E[U_{t+1}^q]}{\partial (I_t/K_t)} < 0$ and $\frac{\partial^2 E[U_{t+1}^q]}{\partial (I_t/K_t)^2} = 3aK_t^2(E_t[K_{t+2}])^2 > 0$. I use this convexity later to interpret the evidence on the second-order effects of anomalies.

Titman, Wei, and Xie (2004) document that the investment anomaly is stronger in firms with higher operating income-to-asset. This pattern can be captured within the model. Using equation (23) to express $\frac{\Pi_{t+1}}{K_{t+1}}$ in terms of $\frac{\Pi_t}{K_t}$ and differentiating $\frac{\partial E[r_{t+1}]}{\partial (I_t/K_t)}$ with respect to $\frac{\Pi_t}{K_t}$ yields $\frac{\partial}{\partial (I_t/K_t)} \left| \frac{\Pi_{t+1}}{K_{t+1}} \right| / \partial \left( \frac{\Pi_t}{K_t} \right) = -\frac{aE[U_{t+1}^q]}{[1+a(I_t/K_t)]^2} > 0$.

Most existing investment-based models can explain the investment anomaly. I contribute by unifying their diverse modeling structures within a single, analytical framework and by using it to explain an array of other anomalies. Moreover, my explanation of the interaction between the investment-return relation and operating income-to-capital seems new.

**The Value Anomaly**

The downward-sloping and convex investment-demand function manifests itself as many anomalies other than the investment anomaly. For example, the value anomaly can be
explained using this function. From the optimality condition (9), \(1 + a \frac{\partial I_t}{\partial q_t} = q_t = Q_t\), so 
\[
\frac{\partial (I_t/K_t)}{\partial q_t} = \frac{1}{a} > 0.
\]
Growth firms with high market-to-book therefore invest more and grow faster than value firms with low market-to-book, consistent with the evidence in Fama and French (1995). The chain rule of partial derivatives then implies that 
\[
\frac{\partial E_t[r_{t+1}]}{\partial Q_t} = \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial q_t} < 0.
\]
Growth firms therefore earn lower expected returns than value firms.

The evidence that the value anomaly is stronger in small firms can be explained within the model. By the chain rule, 
\[
\frac{\partial E_t[r_{t+1}]}{\partial Q_t} / \partial P_t < 0
\]
if the investment-demand function is convex. But \(1 + a \frac{K_t}{K_{t+1}} = q_t = Q_t = \frac{P_t}{K_{t+1}}\) implies that 
\[
P_t = \left[1 + a \frac{K_t}{K_{t+1}} \right] \frac{K_t}{K_{t+1}} (1 - \delta) K_t.
\]
Differentiating both sides with respect to \(\frac{K_t}{K_{t+1}}\) yields 
\[
\frac{\partial P_t}{\partial (I_t/K_t)} = q_t K_t + a K_{t+1} > 0.5
\]


My model is built upon Zhang (2005). By making Assumptions 1–3, I obtain analytical results that are arguably more transparent to communicate. The scope of anomalies addressed here is also much wider. Zhang offers an explicitly solved model in which Assumption 3 is violated. And his simulation results on the value premium are consistent with my analytical results. Therefore, even when stock and investment returns are not exactly equal without Assumption 3, these two returns share similar time-series and cross-sectional properties.

---

\[\frac{\partial P_t}{\partial (I_t/K_t)} > 0\] and \(\frac{\partial Q_t}{\partial (I_t/K_t)} > 0\) both imply that growth firms invest more and grow faster. \(\frac{\partial P_t}{\partial (I_t/K_t)} > 0\) does not contradict the evidence that small firms invest more and grow faster than big firms (e.g., Evans (1987) and Hall (1987)). Their studies use the logarithm of employment as the measure of firm size. This measure corresponds to \(\log(K_t)\) in the model. The model is consistent with this evidence because 
\[
P_t = \left[1 + a \frac{K_t}{K_{t+1}} \right] \frac{K_t}{K_{t+1}} (1 - \delta) K_t\text{ implies that } K_t = P_t / \left[1 + a \frac{K_{t+1}}{K_t} \right] \frac{K_t}{K_{t+1}} (1 - \delta),\]
which implies \(\frac{\partial K_t}{\partial (I_t/K_t)} < 0\).
Using a putty-clay investment model, Gourio (2004) argues that imperfect capital-labor substitutability can induce more than one percent increase in operating profits given a one percent increase in sales. This effect is more important for value firms because they have low productivity. Chen (2004) explains the inverse relation between the value anomaly and size. He argues that the inverse relation arises from shorter life expectancy of small firms. This mechanism is different from mine based on convex adjustment costs and therefore applies to both short-lived and long-lived firms. Gala (2006) extends the economic mechanism driving the value anomaly in Zhang (2005) into a general equilibrium framework.

The Payout Anomaly

The payout anomaly is another manifestation of the investment-demand function. The bridge is the balance-sheet constraint that requires the sources and the uses of funds within a firm must be equal. When the free cash flow $C_t > 0$, this constraint is $\frac{O_t}{K_t} = \frac{I_t}{K_t} - \frac{q_t}{2} \left( \frac{I_t}{K_t} \right)^2$. Consequently, $\frac{\partial (C_t/K_t)}{\partial (I_t/K_t)} = -\left( 1 + a \frac{I_t}{K_t} \right) = -q_t < 0$. Intuitively, if a firm distributes more capital back to shareholders, the firm is likely to invest less. Consistent with this argument, Grullon, Michaely, and Swaminathan (2002) document that dividend-increasing firms significantly reduce their capital expenditures over the next three years, while the dividend-decreasing firms begin to increase their capital expenditure.

By the chain rule, the negative slope of the investment-demand function manifests as the positive relation between payout-to-asset and expected returns, $\frac{\partial E_t [r_{t+1}]}{\partial (C_t/K_t)} = \frac{\partial E_t [r_{t+1}]}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (C_t/K_t)} > 0$. Moreover, the convexity of the investment-demand function manifests as the stronger payout anomaly in value firms, $\frac{\partial^2 E_t [r_{t+1}]}{\partial (C_t/K_t) \partial Q_t} = \frac{\partial^2 E_t [r_{t+1}]}{\partial (I_t/K_t)^2} \frac{\partial (I_t/K_t)}{\partial (C_t/K_t)} < 0$.

I am not aware of other rational explanations for the payout anomaly.

The Seasoned-Equity-Offering Anomaly

This anomaly is yet another manifestation of the downward-sloping and convex investment-demand function. The bridge is again the balance-sheet constraint of firms. When outside equity $O_t > 0$, the constraint becomes $\frac{O_t}{K_t} = \frac{I_t}{K_t} + \frac{q_t}{2} \left( \frac{I_t}{K_t} \right)^2$. Accordingly, $\frac{\partial (O_t/K_t)}{\partial (I_t/K_t)} = q_t > 0$. Intuitively, if a firm raises more outside capital, the firm is also likely to invest more.
By the chain rule, the negative slope of the investment-demand function manifests itself as
the underperformance following seasoned equity offerings, i.e., \( \frac{\partial E_t[\tau_{t+1}]}{\partial (O_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} < 0 \).
And the convexity manifests itself as the stronger underperformance in small firms, i.e.,
\[
\frac{\partial}{\partial (O_t/K_t)} \left( \frac{\partial E_t[\tau_{t+1}]}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} \right) = \frac{\partial^2 E_t[\tau_{t+1}]}{\partial (I_t/K_t)^2} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} < 0.
\]
The last inequality follows because \( \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} = q_t = P_t = Q_t = \frac{P_t}{K_t + 1} \) implies that
\[
\frac{\partial^2 (I_t/K_t)}{\partial (O_t/K_t) \partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} < 0.
\]
Several studies provide supporting evidence of the investment mechanism driving the
underperformance. Loughran and Ritter (1997) document that issuing firms have much
higher investment-to-asset than nonissuing firms for nine years around the issuing date. Issuing firms are also disproportionately high-growth firms. Richardson and Sloan (2003) find
that the negative relation between external finance and average returns varies systematically with the use of the proceeds. When the proceeds are invested in net operating assets as opposed to being stored as cash, the negative relation is stronger, but the relation is much weaker when the proceeds are used for refinancing or retained as cash. This evidence signals the importance of capital investment in driving the underperformance. Loughran and Ritter interpret their evidence as consistent with a “window-of-opportunity” theory of capital structure under inefficient markets. Richardson and Sloan interpret their evidence as investor underreacting to the overinvestment behavior of empire-building managers, and as “strong and pervasive (p. 34)” evidence of inefficient markets. But I beg to differ.

Eckbo, Masulis, and Norli (2000) argue that issuing equity lowers leverage ratios, and
makes issuers less risky. I do not model leverage. Moreover, Lyandres, Sun, and Zhang
(2005) show that issuers have on average higher leverage ratios than matching nonissuers, inconsistent with the leverage explanation. Schultz (2003) argues that using event studies is likely to find negative abnormal performance ex post, even if there is no abnormal performance ex ante. Schultz uses his argument to explain the underperformance following initial public offerings, but the same logic also applies to seasoned equity offerings. Weighting each period equally as in calendar-time regressions solves this problem. I differ because the investment mechanism applies in both event-time and calendar-time.

Carlson, Fisher, and Giammarino (2006) argue that that prior to issuance, a firm has
both assets in place and an option to expand. This composition is a levered, risky position.
If the exercise of the option is financed by equity issuance, then risk must drop afterwards. This explanation is consistent with mine because both work through capital investment. But I do not assume growth options to be riskier than assets in place, although it is very likely to be true in good times when the option to expand is important.

More generally, the investment mechanism should also apply to other external-financing anomalies such as the underperformance following initial public offerings (e.g., Ritter (1991)), straight and convertible debt offerings (e.g., Spiess and Affleck-Graves (1999)), private placements of equity (e.g., Hertzel, Lemmon, Linck, and Rees (2002)), and loan announcements (e.g., Billett, Flannery, and Garfinkel (2005)). Intuitively, from the balance-sheet constraint, firms raising capital should invest more and therefore earn lower expected returns.

In contrast, behavioral underreaction seems an incomplete explanation for these anomalies. This hypothesis is based on the observation that the short-term announcement effects and the long-term stock price drift following equity and debt offerings are both negative, as if investors underreact to the information contained in these events. However, the announcement effects of private placements of equity, loan announcements, and seasoned equity offerings in Japanese firms studied by Kang, Kim, and Stulz (1999) are all positive, despite their negative long-term drifts. This inconsistency should suggest behavioral overreaction.

More likely, short-term and long-term stock price movements are driven by different economic forces, with short-term movements driven by asymmetric information and long-term movements driven by time-varying expected returns related to capital investment. Equity and debt offerings in the U.S. are likely to signal negative news such as insufficient internal funds and the intrinsic value based on unobservable private signals being lower than the market value. But private placements of equity and bank loans are likely to signal positive news because large block shareholders and banks, from their ongoing relationships with the firms, have information advantage and can serve as monitors. And because of the main bank system in Japan, equity shares are mostly held by banks and other large institutional investors.

In sum, the investment, value, payout, and seasoned-equity-offering anomalies are all essentially the same phenomenon driven by optimal investment. The idea that investment is a first-order determinant of expected returns seems palatable. A central tenet in corporate
finance is that investment is a first-order determinant of firm value, a pointed formalized by
the investment Euler equations (9) and (10). The investment-return relation is formalized
by equations (15) and (16), algebraically equivalent to the investment Euler equations.

3.3 The General Framework

This section establishes all the results in the general framework outlined in Section 2. This
step is extremely important because the general model allows much more flexible econo-
metric specifications than the quadratic adjustment costs in Section 3.2. Except for a few
technical details, the predictions from Section 3.2 remain largely unchanged.

THE EARNINGS ANOMALIES

Proposition 3 Under Assumptions 1 and 3, expected returns correlate positively with ex-
pected profitability, \( \frac{\partial E_t \left[ r_{t+1} \right]}{\partial E_t \left[ N_{t+1}/K_{t+1} \right]} > 0 \), and the magnitude of the correlation decreases with the
market value, \( \frac{\partial^2 E_t \left[ r_{t+1} \right]}{\partial E_t \left[ N_{t+1}/K_{t+1} \right] \partial P_t} < 0 \).

Proposition 4 Under Assumptions 1–4, expected returns correlate positively with the
current-period profitability, and the magnitude of this correlation decreases with the market
value, i.e., \( \frac{\partial E_t \left[ r_{t+1} \right]}{\partial (N_t/K_t)} > 0 \) and \( \frac{\partial^2 E_t \left[ r_{t+1} \right]}{\partial (N_t/K_t) \partial P_t} < 0 \).

Proposition 5 Under Assumptions 1–4, expected returns correlate positively with the
current-period earnings surprise, and the magnitude of this correlation decreases with the
market value, i.e., \( \frac{\partial E_t \left[ r_{t+1} \right]}{\partial \epsilon_t} > 0 \) and \( \frac{\partial^2 E_t \left[ r_{t+1} \right]}{\partial \epsilon_t \partial P_t} < 0 \).

THE INVESTMENT ANOMALY

Proposition 6 Under Assumptions 1–3, expected returns decrease with investment-to-asset:
\( \frac{\partial E_t \left[ r_{t+1} \right]}{\partial (I_t/K_t)} < 0 \). If Assumption 4 also holds, the magnitude of the investment anomaly increases
with operating profit-to-asset: \( \partial \left| \frac{\partial E_t \left[ r_{t+1} \right]}{\partial (I_t/K_t)} \right| / \partial \left( \frac{P_t}{K_t} \right) > 0 \).

THE VALUE ANOMALY

To explain the second-order effects of anomalies in the general model, I need:
Assumption 5 The augmented adjustment-cost function $\Phi(I_t, K_t)$ satisfies:

$$\Phi_{111}(I_t, K_t) \geq 0; \quad \Phi_{112}(I_t, K_t) \leq 0; \quad \text{and} \quad \Phi_{222}(I_t, K_t) \leq 0$$

For example, standard specifications such as the quadratic adjustment-cost function and the proportional and quadratic financing-cost functions satisfy this assumption.

Proposition 7 Under Assumptions 1–3, expected returns correlate negatively with market-to-book, $\frac{\partial E_t[r_{t+1}]}{\partial Q_t} < 0$. If Assumption 5 also holds, the magnitude of this correlation decreases in the market value, $\partial \left| \frac{\partial E_t[r_{t+1}]}{\partial Q_t} \right| / \partial P_t < 0$.

The Payout Anomaly

Proposition 8 Denote payout or free cash flow as:

$$C_t \equiv C(K_t, K_{t+1}, X_t) = (\Pi(K_t, X_t) - \Phi(I_t, K_t)) 1_{\{\Pi(K_t, X_t) - \Phi(I_t, K_t) > 0\}} (28)$$

where $1_{\{\cdot\}}$ is an indicator function that takes the value of one if the event described in $\{\cdot\}$ is true and zero otherwise. Under Assumptions 1–3, expected returns increase weakly with the payout rate, $C_t/K_t$, i.e., $\frac{\partial E_t[r_{t+1}]}{\partial C_t/K_t} \geq 0$, where the inequality is strict when $C_t$ is strictly positive. If Assumption 5 also holds, the payout anomaly is stronger in value firms than that in growth firms, $\frac{\partial^2 E_t[r_{t+1}]}{\partial (C_t/K_t)^2} \leq 0$, where the inequality is strict when $C_t > 0$.

The Seasoned-Equity-Offering Anomaly

Proposition 9 Denote the outside equity, $O_t$, as:

$$O_t \equiv O(K_t, K_{t+1}, X_t) = (\Phi(I_t, K_t) - \Pi(K_t, X_t)) 1_{\{\Phi(I_t, K_t) - \Pi(K_t, X_t) > 0\}} (29)$$

Under Assumptions 1 and 3, expected returns decrease weakly with the rate of external or outside equity, $O_t/K_t$, i.e., $\frac{\partial E_t[r_{t+1}]}{\partial O_t/K_t} \leq 0$, where the inequality is strict when $O_t$ is strictly positive. If Assumption 5 also holds, the magnitude of this correlation is stronger in small firms, $\partial \left| \frac{\partial E_t[r_{t+1}]}{\partial O_t/K_t} \right| / \partial P_t \leq 0$, where the inequality is strict when $O_t > 0$.

New Testable Hypotheses
Proposition 10 Under Assumptions 1–3 and 5: the investment anomaly is stronger in small firms, \( \frac{\partial E_1[r_{t+1}]}{\partial (I_t/K_t)} / \partial P_t < 0 \); and the investment anomaly is stronger in value firms, \( \frac{\partial E_1[r_{t+1}]}{\partial (I_t/K_t)} / \partial Q_t < 0 \).

4 Interpretation

I discuss implications of my work in the context of related literature and a loose end.

4.1 Rationality as the Benchmark

Two parallel interpretations coexist for stock return predictability, time-varying expected returns or predictable abnormal returns. The second interpretation suggests that forecasting errors are forecastable. But the first interpretation is fully consistent with rationality.

From this perspective, my work complements the behavioral theories of Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subramanyam (1998), and Hong and Stein (1999). In explaining anomalies, these models assume constant expected returns and focus instead on predictable abnormal returns. By constructing the neoclassical microfoundation for time-varying expected returns in the cross section, I provide a natural benchmark against which the quantitative importance of overreaction and underreaction can be gauged.

Some colleagues argue that, because of the exogenous stochastic discount factor, \( M_{t+1} \), my analysis does not have much to say about whether the underlying variation in expected returns is rational or not. If a fad or sentiment raises prices and lowers returns, firms in the model react appropriately to generate low returns after periods of high investment.

This argument is incomplete. On the one hand, my results indeed go through if \( M_{t+1} \) is from loss-averse investors studied by Barberis, Huang, and Santos (2001). But despite the loss aversion, the representative investor is rational: “While we do modify the investor’s preferences to reflect experimental evidence about the sources of utility, the investor remains rational and dynamically consistent throughout” (Barberis et al. p. 5).

On the other hand, all my results go away if \( M_{t+1} \) is from irrational investors studied by Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subramanyam (1998), and...
Hong and Stein (1999). In all these models, the aggregate discount rate is constant, suggesting that $M_{t+1}$ is also a constant, $M$. Suppose managers only maximize the fundamental value given by equation (7), then the first principle in equation (15) says $E_t[r_{t+1}] = 1/M$, i.e., all firms earn constant expected returns equal to the risk-free rate.

In contrast, the irrational-investor-rational-manager model outlined in Baker, Ruback, and Wurgler (2004) can generate the payout and external-financing anomalies. The reason is that they impose catering and market-timing incentives onto the objective function of the firms, besides the fundamental value. But why need a convoluted list of ad hoc assumptions given that the neoclassical model already, at least qualitatively, does the job?

More important, my insistence on the neoclassical model as the benchmark does not mean that psychological biases are unimportant. The point is that, without structural models as measurement tools to control for rational variations of expected returns, one cannot hope for any realistic understanding of the quantitative importance of investor sentiment.

### 4.2 Prescott: “Progress, Don’t Regress”

Factor regressions are extremely popular in empirical finance. In event studies, cumulative abnormal returns are computed as the difference between realized returns and expected returns estimated from, for example, the CAPM. Tests are performed to see if abnormal returns are on average zero (e.g., Ball and Brown (1968) and Fama, Fisher, Jensen, and Roll (1969)). In cross-sectional tests, stock returns are regressed on beta and firm characteristics (e.g., Fama and French (1992)). In time-series tests, mimicking portfolios are formed based on firm characteristics. And zero-intercept tests are performed by regressing these portfolio returns on a set of benchmark factor returns (e.g., Fama and French (1993, 1996)).

Factor models are good at describing empirical patterns, but are largely uninformative for testing economic theories. First, the null hypotheses are often the CAPM and its various ad hoc extensions, unconditional or conditional, single-index or multi-factor models. Similar to the CAPM, all these extensions say that only covariances should explain expected returns. Anomalies arise because characteristics often dominate covariances in explaining average returns. That only covariances should matter is the basic premise of Daniel and
Titman (1997), and is taken as “one general feature of the rational approach” in Barberis and Thaler (2003, p. 1091). But Section 3 shows that characteristics affect expected returns, often in the directions reported in the anomalies literature. And my model is entirely rational. The empirical debate on covariances versus characteristics is therefore not a well-defined question. And rejecting the CAPM and its close cousins is not equivalent to rejecting efficient markets (e.g., Fama (1965)) or rational expectations (e.g., Muth (1961) and Lucas (1972)).

Similarly, using multiple common factors to eliminate significant alphas in time-series regressions (e.g., Fama and French (1996), Fama (1998), and Eckbo, Masulis, and Norli (2005)) does not mean that anomalies do not exist so markets are efficient, only that different anomalies are empirically interconnected. As pointed out rightfully by Barberis and Thaler (2003, p. 1091): “[J]ust because a factor model happens to work well does not necessarily mean that we are learning anything about the economic drivers of average returns.”

Moreover, factor models face a variety of measurement issues. Decades of research have not settled on the list of priced factors, with which stock returns are supposed to covary. And all dynamic asset pricing models say that betas are time-varying. But no easy-to-implement econometric specifications have been derived. Estimates of time-varying betas often use ad hoc specifications, yielding unreliable results. And do not forget about time-varying risk premiums that are notorious to measure. Accordingly, estimates of expected returns from factor models are extremely imprecise even at the industry level (e.g., Fama and French (1997)).

The theoretical work of Berk, Green, and Naik (1999) and Gomes, Kogan, and Zhang (2003) illustrates forcefully the limitations of factor regressions. In their models, betas are sufficient statistics of expected returns, and true betas dominate characteristics in cross-sectional regressions in simulations. But when betas estimated with the Fama-French (1992) methods are used in these regressions, betas are dominated by characteristics. Factor models are therefore largely uninformative for testing economic theories.

To be clear, I define anomalies as empirical associations between firm characteristics and average returns not captured by existing asset pricing models. The CAPM can be perfectly happy with average-return dispersions if they come with similar market-beta dispersions. But precisely because of the severe inadequacies of traditional risk measures estimated from
factor regressions, in this paper I simply treat anomalies as associations between firm characteristics and average returns.

More generally, anomalies with respect to existing models can turn out to be regularities with respect to new models. This paper demonstrates this theoretical possibility from the value-maximization perspective of firms. It remains an open question to what extent the new expected-return model outlined in equation (19) can turn anomalies into regularities quantitatively. But the new model seems more useful than factor regressions because its inputs are mostly observable characteristics, much easier to measure than time-varying betas.

4.3 A Frontal Attack on Anomalies

As shown convincingly in Berk, Green, and Naik (1999), investment-based asset pricing is a much more direct approach to cracking anomalies than consumption-based asset pricing. The reason is simple. Anomalies are relations between average returns and firm characteristics, not consumer characteristics. Grilling consumers, with or without psychological biases, for answers to anomalies while ignoring firms is at best an indirect, if not a convoluted approach.

Despite repeated earlier failures, consumption-based asset pricing has had some recent success (e.g., Lettau and Ludvigson (2001), Parker and Julliard (2004), Lustig and Nieuwerburgh (2005), and Piazzesi, Schneider, and Tuzel (2005)). In all these papers, characteristics do not enter the tests directly: they are buried in portfolio returns in moment conditions. And even if the models survive the over-identification test, it is not clear what economic mechanisms drive the results. In Lettau and Ludvigson and Lustig and Nieuwerburgh, the returns of value stocks covary more with the price of risk in bad times than the returns of growth stocks. In Campbell and Vuolteenaho (2004) and Bansal, Dittmar, and Lundblad (2005), the value anomaly can be explained because value stocks have higher cash-flow betas than growth stocks. But why these patterns arise is left open because firm dynamics are not modeled.

By linking expected returns to firm characteristics, investment-based asset pricing attacks anomalies much more directly. Fama (1991, p. 1610) calls for a coherent story that “(1) relates the cross-section properties of expected returns to the variation of expected returns through time, and (2) relates the behavior of expected returns to the real economy in
a rather detailed way”. Consumption-based asset pricing is naturally fit for the first goal, but investment-based asset pricing seems better equipped for the second. And Fama’s vision of a coherent story can be provided by the conceptual framework of general equilibrium.

4.4 Koopmans (1947): “Measurement without Theory”

The neoclassical model is also related to the Ohlson (1995) residual income model and other valuation models that are popular in capital markets research in accounting (e.g., Frankel and Lee (1998), Dechow, Hutton, and Sloan (1999), and Kothari (2001)). The Ohlson model says \[ \frac{P_t}{B_t} = 1 + \sum_{j=1}^{\infty} \frac{E[Y_{t+j} - rB_{t+j-1}]}{B_t}(1+r)^j \] where \( B_t \) is book equity at time \( t \), \( Y_{t+j} \) is earnings at \( t+j \), \( Y_{t+j} - rB_{t+j-1} \) is the residual income, defined as the difference between earnings and the opportunity cost of capital, and \( r \) is the discount rate for the expected residual income or the long-term expected stock return.

As shown in Fama and French (2006), the Ohlson model has several important predictions. First, controlling for expected residual earnings and expected book equity relative to current book equity, a higher book-to-market equity implies a higher expected return. Second, given book-to-market, firms with higher expected residual income relative to current book equity have higher expected returns. Third, controlling for book-to-market and the expected growth in book equity or investment growth, more profitable firms or firms with higher expected earnings relative to current book equity have higher expected returns. These predictions are largely consistent with those from the neoclassical model, suggesting that the latter should be potentially useful in guiding empirical capital markets research.

More important, the neoclassical model contrasts with valuation models in two important ways. First, conceptually, valuation models are accounting models and are therefore silent about the behavior of underlying economic agents. In contrast, built on the first principles, the neoclassical model can be used to distinguish different hypotheses of economic behavior.

Second, the neoclassical model is also likely to have practical advantages over valuation models. Several papers use the valuation models to estimate expected returns (e.g., Claus and Thomas (2001), Gebhardt, Lee, and Swaminathan (2001), and Fama and French (2002)), and find that the estimated equity premium is only about 2–3%, much lower than the historical
average. Kothari (2001) argues that the long-term growth forecasts in these papers, especially the terminal perpetuity growth rates, seem too low. In contrast, estimating expected returns from the neoclassical model only requires inputs of the one-period-ahead profitability and investment-to-asset: applying the principle of dynamic programming effectively reformulates infinite-horizon models to two-period models. And one-period-ahead forecasts should be more reliable than their long-term counterparts.

4.5 A Loose End

This paper has an important loose end: the analysis is purely qualitative, even though as exemplified by Berk (1995), analytical results are more transparent for communicating the underlying intuition. Nevertheless, subsequent papers have tried to investigate in depth the quantitative importance of the proposed economic drivers of anomalies. Collectively, these papers show that the mechanisms seem at least quantitatively relevant, if not important.

Using the Fama-French (1993) portfolio approach, Lyandres, Sun, and Zhang (2005) implement a direct test on the investment mechanism of the underperformance following seasoned equity offerings. Lyandres et al. find that adding a return factor based on capital investment into standard factor regressions makes the underperformance largely insignificant and reduces its magnitude by around 40%. And issuers have stronger shareholder rights than matching nonissuers, inconsistent with the overinvestment-based explanation.

Using the Kydland-Prescott (1982) quantitative theory approach, Li, Livdan, and Zhang (2006) use a fully-specified neoclassical model augmented with costly external equity as a laboratory to study the quantitative relations between stock returns and equity financing. Simulations show that the model can simultaneously and in many cases quantitatively reproduce: procyclical equity issuance; the negative relation between aggregate equity share and future stock market returns; long-term underperformance following equity issuance and the positive relation of its magnitude with the volume of issuance; the mean-reverting behavior in the operating performance of issuing firms; and the positive long-term stock price drift of firms distributing cash and its positive relation with book-to-market.

Using the Hansen-Singleton (1982) structural estimation approach, Whited and Zhang
(2006) perform the test outlined in equation (19). Under various specifications, the model-implied average investment returns display similar magnitudes of dispersion as in average stock returns across portfolios sorted on investment-to-asset and on size and book-to-market. However, the model-predicted average-return dispersions across earnings-surprise portfolios are somewhat lower in magnitude than those in the data.

5 Conclusion

I develop the neoclassical microfoundation for most anomalies in efficient markets.

In the same fashion that the utility-maximization of consumers predicts that aggregate expected returns should vary with the business cycles, the value-maximization of firms predicts that expected returns in the cross-section should vary with firm characteristics, corporate policies, and events. This approach can qualitatively explain many anomalies including the associations of future stock returns with market-to-book, investment and disinvestment, seasoned equity and other forms of external finance, tender offers and stock repurchases, dividend omissions and initiations, expected profitability, profitability, and earnings surprises. I also propose a purely characteristic-based, testable expected-return model that can be used to quantify the proposed economic drivers of anomalies.

The basic point of this paper is reminiscent in spirit of the point of Kydland and Prescott (1982). The Kydland-Prescott paper says that the neoclassical framework is a good start to understand the driving forces behind the business cycles. Animal spirits and sticky prices from Keynesian economics can be important, but their effects must be quantified using the neoclassical benchmark. Similarly, this paper says that the neoclassical framework is a good start to understand the driving forces behind the cross section of returns. Overreaction and underreaction from behavioral finance can be important, but their effects must be quantified using the neoclassical benchmark.

References


Gala, Vito D., 2006, Investment and returns, working paper, University of Chicago.


Gourio, François, 2004, Operating leverage, stock market cyclicality, and the cross-section of returns, working paper, University of Chicago.


Li, Erica X. N., Dmitry Livdan, and Lu Zhang, 2006, Optimal market timing, working paper, Texas A&M University and University of Rochester.


Lyandres, Evgeny, Le Sun, and Lu Zhang, 2005, Investment-based underperformance following seasoned equity offerings, working paper, Rice University and University of Rochester.


Xing, Yuhang, 2005, Interpreting the value effect through the $Q$-theory: an empirical investigation, working paper, Rice University.


**A Proofs**

**Proof of Lemma 1** By the Principle of Optimality (e.g., Theorem 9.2 of Stokey, Lucas, and Prescott (1989)), the value function in equation (7) can be rewritten recursively as:

$$V(K_t, X_t) = \max_{\{K_{t+1}\}} \Pi(K_t, X_t) - \Psi(K_t, K_{t+1}) + E_t[M_{t+1}V(K_{t+1}, X_{t+1})]$$  \hspace{1cm} (A1)

where

$$\Psi(K_t, K_{t+1}) \equiv \Phi(K_{t+1} - (1 - \delta)K_t, K_t)$$  \hspace{1cm} (A2)

The envelope condition and the first-order condition are, respectively:

$$V_1(K_t, X_t) = \Pi_1(K_t, X_t) - \Psi_1(K_t, K_{t+1})$$  \hspace{1cm} (A3)

$$0 = -\Psi_2(K_t, K_{t+1}) + E_t[M_{t+1}V_1(K_{t+1}, X_{t+1})]$$  \hspace{1cm} (A4)

Next, from equation (A2),

$$\Psi_1(K_t, K_{t+1})K_t + \Psi_2(K_t, K_{t+1})K_{t+1} = G\left(\frac{K_{t+1}}{K_t} - (1 - \delta)\right)K_{t+1}K_t^{\alpha-1} + \alpha G\left(\frac{K_{t+1}}{K_t} - (1 - \delta)\right)K_t^\alpha$$

$$+ G\left(\frac{K_{t+1}}{K_t} - (1 - \delta)\right)K_t^{\alpha-1}K_{t+1} = \alpha \Psi(K_t, K_{t+1})$$  \hspace{1cm} (A5)

Now plugging equation (A3) into equation (A4) yields the stochastic Euler equation:

$$-\Psi_2(K_t, K_{t+1}) + E_t[M_{t+1}(\Pi_1(K_{t+1}, X_{t+1}) - \Psi_1(K_{t+1}, K_{t+2}))] = 0$$  \hspace{1cm} (A6)
Expanding the value function (A1) recursively and using equations (3) and (A5) to obtain:

\[ \alpha V(K_t, X_t) = \Pi_1(K_t, X_t)K_t - \Psi_1(K_t, K_{t+1})K_t - \Psi_2(K_t, K_{t+1})K_{t+1} + E_t[M_{t+1}[\Pi_1(K_{t+1}, X_{t+1})K_{t+1} - \Psi_1(K_{t+1}, K_{t+2})K_{t+2} + \Psi_2(K_{t+1}, K_{t+2})K_{t+2}]] = \cdots = \Pi_1(K_t, X_t)K_t - \Psi_1(K_t, K_{t+1})K_t = V_1(K_t, X_t)K_t \]

where the third equality follows from recursive substitution and from equation (A6). The last equality follows from the envelope condition (A3).

**Proof of Lemma 2** Solving equation (10) forward yields

\[
q_t = E_t[M_{t+1} \Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1})] + E_t[M_{t+1}(1 - \delta)q_{t+1}]
\]

\[
= E_t[M_{t+1} \Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1})] + E_t[M_{t+1}(1 - \delta)E_{t+1}[M_{t+2}(\Pi_1(K_{t+2}, X_{t+2}) - \Phi_2(I_{t+2}, K_{t+2}) + (1 - \delta)q_{t+2})]]
\]

\[
= E_t[M_{t+1}(\Pi_1(K_{t+1}, X_{t+1}) - \Phi_2(I_{t+1}, K_{t+1})) + M_{t+2}(1 - \delta)(\Pi_1(K_{t+2}, X_{t+2}) - \Phi_2(I_{t+2}, K_{t+2}))]
\]

\[
+ E_t[M_{t+2}(1 - \delta)^2q_{t+2}] = \cdots = E_t \left[ \sum_{j=1}^{\infty} M_{t+j}(1 - \delta)^{j-1}(\Pi_1(K_{t+j}, X_{t+j}) - \Phi_2(I_{t+j}, K_{t+j})) \right]
\]

where the last two equalities follow from recursive substitution.

**Proof of Proposition 1** From Proposition 1, \( V_1(K_t, X_t)K_t = \alpha V(K_t, X_t) \). Both sides can be rewritten as: \( \Pi_1(K_t, X_t)K_t - \Phi_2(I_t, K_t)K_t + q_t(1 - \delta)K_t = \alpha P_t + \alpha \Pi(K_t, X_t) - \alpha \Phi(I_t, K_t) \). Simplifying using homogeneity of \( \Pi(K_t, X_t) \) and \( \Phi(I_t, K_t) \) yields \( q_t(1 - \delta)K_t = \alpha P_t - \Phi_1(I_t, K_t)I_t \). Equation (13) now follows because \( q_t = \Phi_1(I_t, K_t) \) from equation (9).

**Proof of Lemma 3** To show the first inequality, from Assumptions 1 and 2, \( \Pi_1 > 0 \) and \( \Phi_2 \leq 0 \), equation (11) then implies that \( q_t > 0 \). But from equation (9), \( \Phi_1 \) equals \( q_t \) at the optimum. Therefore, although \( \Phi_1 \) in general can be positive, negative, or zero when \( I_t \leq 0 \), it is strictly positive at the optimum. Equivalently, \( G^\prime(\cdot) \) is strictly positive at the optimum.

Now from Lemma 1 and equations (9) and (14),

\[
\Phi_{12}(I_t, K_t) = \frac{\partial q_t}{\partial K_t} = E_t \left[ M_{t+1} \frac{\partial V_1(K_{t+1}, X_{t+1})}{K_t} \right] = (1 - \delta)E_t[M_{t+1}V_{11}(K_{t+1}, X_{t+1})]
\]

But differentiating both sides of \( \alpha V(K_t, X_t) = V_1(K_t, X_t)K_t \) yields:

\[
V_{11}(K_t, X_t) = \frac{(\alpha - 1)}{K_t} V_1(K_t, X_t) = \frac{(\alpha - 1)}{K_t} \frac{V(K_t, X_t)}{K_t} = \frac{(\alpha - 1)}{K_t} \alpha \hat{Q}_t \leq 0
\]

This says that the value function is weakly concave in capital. Now plugging equation (A8) into (A7) yields: \( \Phi_{12}(I_t, K_t) = (1 - \delta)(\alpha - 1) \frac{1}{K_{t+1}} E_t[M_{t+1} \alpha \hat{Q}_{t+1}] = (1 - \delta)(\alpha - 1) \frac{\alpha}{K_{t+1}} \leq 0 \).

Differentiating both sides with respect to \( K_t \) yields: \( \Phi_{122}(I_t, K_t) = (\alpha - 1)(1 - \delta) \frac{\alpha}{K_{t+1}} \frac{\partial q_t}{\partial K_t} + \frac{(\alpha - 1)}{K_t} \alpha \hat{Q}_t \leq 0 \)
\((\alpha - 1)(1 - \delta)^2 q_t \left(-\frac{1}{K_{t+1}^2}\right) \geq 0. \quad \blacksquare\)

**Proof of Proposition 2** First express stock return in equation (17) in terms of cum-dividend firm value as \(r_{t+1}^S = \frac{V_t(K_{t+1}, X_t) - q_t}{V_t(K_{t+1}, X_t) + \psi(K_{t+1})}\). The recursive value function (A1) evaluated at the optimum then yields \(E_t[M_t r_{t+1}^S] = 1\).

Combining equations (A3) and (A4) yields an alternative investment return, \(r_{t+1}^I\):

\[
r_{t+1}^I = \frac{V_t(K_{t+1}, X_{t+1})}{\Psi_2(K_t, K_{t+1})} = \frac{V_t(K_{t+1}, X_t)K_{t+1}}{\alpha \Psi(K_t, K_{t+1}) - \Psi_1(K_t, K_{t+1})K_t}
\]

which is equal to equation (16) since (A2) implies \(\Psi_2 = \Phi_1\) and \(\Psi_1 = \Phi_2 - \Phi_1(1 - \delta)\). Now,

\[
r_{t+1}^I = \frac{V_t(K_{t+1}, X_{t+1})}{\Psi_2(K_t, K_{t+1})} = \frac{V_t(K_{t+1}, X_{t+1})K_{t+1}}{\alpha \Psi(K_t, K_{t+1}) - \Psi_1(K_t, K_{t+1})K_t}
\]

where the first equality follows from equation (A9), the second follows from equation (A5), the third equality follows from the envelope condition (A3), and the fourth equality follows from Lemma 1 and equation (3). \(\blacksquare\)

**Lemma 4** Define the numerator of investment return:

\[
U_{t+1} \equiv \Pi_1(K_{t+1}, X_t) - \Phi_2(I_{t+1}, K_{t+1}) + (1 - \delta)\Phi_1(I_{t+1}, K_{t+1}) \quad \text{(A10)}
\]

Under Assumptions 1–3, \(U_{t+1} > 0\), and the (gross) returns are positive, \(r_{t+1} = \frac{U_{t+1}}{\Phi_1(I_t, K_{t+1})} > 0\).

**Proof.** \(\Pi_1 > 0\) and \(\Phi_2 \leq 0\) follow from Assumptions 1 and 2, respectively. And \(\Phi_1 = q_t > 0\) follows from Lemma 3. \(\blacksquare\)

**Proof of Proposition 3** From equation (3), \(\Pi_1(K_t, X_t) = \alpha \left(\frac{K_t}{q_t}\right) = \alpha \left(\frac{N_t}{q_t} + \delta\right)\). Equation (16) then implies that

\[
\frac{\partial E_t[r_{t+1}]}{\partial E_t[N_{t+1}/K_{t+1}]} = -\frac{K_t^{-1}}{q_t} E_t \left[ \frac{1}{q_t} G'(\frac{I_{t+1}}{K_{t+1}}) + G''(\frac{I_{t+1}}{K_{t+1}}) \frac{K_{t+1}}{K_t} - \frac{\partial(I_{t+1}/K_{t+1})}{\partial q_{t+1}} \frac{\partial q_{t+1}}{\partial E_t[N_{t+1}/K_{t+1}]} \right]
\]

But equation (9) implies that \(\frac{\partial(I_{t+1}/K_{t+1})}{\partial q_{t+1}} = \frac{K_t^{-1}}{G'(I_{t+1}/K_{t+1})} \text{ and equation (11) implies that} \)

\[
\frac{\partial q_{t+1}}{\partial E_t[N_{t+1}/K_{t+1}]} = \rho_q \frac{r_{t+1}}{r_{t+1}}. \text{ Using these results and noting that} q_t = \alpha Q_t = \alpha \frac{K_t}{K_{t+1}} \text{ yields:}
\]

\[
\frac{\partial E_t[r_{t+1}]}{\partial E_t[N_{t+1}/K_{t+1}]} = \frac{K_{t+1}}{P_t} \left( 1 + \frac{1}{\alpha} E_t \left[ \frac{(1 - \alpha) G'(I_{t+1}/K_{t+1})}{G''(I_{t+1}/K_{t+1})} + \frac{K_{t+2}}{K_{t+1}} \right] \right)
\]
which is positive and decreasing in the market value, $P_t$. ■

**Proof of Proposition 4** Proposition 4 follows directly from the proof of Proposition 3 by noting that
\[
\frac{\partial E_t[r_{t+1}]}{\partial (N_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (N_{t+1}/K_{t+1})} \frac{\partial E_t[N_{t+1}/K_{t+1}]}{\partial (N_t/K_t)} = \rho \frac{\partial E_t[r_{t+1}]}{\partial (N_t/K_t)}.
\]

**Proof of Proposition 5** Proposition 5 also follows directly from the proof of Proposition 3 by noting that
\[
\frac{\partial E_t[r_{t+1}]}{\partial (N_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (N_{t+1}/K_{t+1})} \frac{\partial E_t[N_{t+1}/K_{t+1}]}{\partial (N_t/K_t)} = \rho \frac{\partial E_t[r_{t+1}]}{\partial (N_t/K_t)}.
\]

**Proof of Proposition 6** From equations (5) and (16),
\[

r_{t+1} = \frac{U_{t+1}}{\Phi_1(I_t, K_t)} = \frac{U_{t+1}}{G'(I_t/K_t)K_t^{\alpha-1}}
\]
\[
\frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} = -\frac{E_t[U_{t+1}]}{K_t^{\alpha-1}[G'(I_t/K_t)]^2}G''(I_t/K_t) + \frac{1}{\Phi_1(I_t, K_t)} \frac{\partial E_t[U_{t+1}]}{\partial (I_t/K_t)} (A12)
\]

where the first term in equation (A12) is less than zero because $G''(\cdot) > 0$ from Assumption 3. It then suffices to show that $\frac{\partial E_t[U_{t+1}]}{\partial (I_t/K_t)} < 0$. But plugging equation (4) into (A10) yields:
\[
E_t[U_{t+1}] = E_t \left[ \Pi_1 \left( \left[ \frac{I_t}{K_t} + (1 - \delta) \right] K_t, X_{t+1} \right) \right. \\
- \Phi_2 \left( K_{t+2} - (1 - \delta) \left[ \frac{I_t}{K_t} + (1 - \delta) \right] K_t, \left[ \frac{I_t}{K_t} + (1 - \delta) \right] K_t \right) \\
+ (1 - \delta) \Phi_1 \left( K_{t+2} - (1 - \delta) \left[ \frac{I_t}{K_t} + (1 - \delta) \right] K_t, \left[ \frac{I_t}{K_t} + (1 - \delta) \right] K_t \right) \right] (A13)
\]

Differentiating both sides with respect to $(I_t/K_t)$ yields:
\[
\frac{\partial E_t[U_{t+1}]}{\partial (I_t/K_t)} = E_t[\Pi_1(K_{t+1}, X_{t+1})K_t + 2(1 - \delta)\Phi_2(I_{t+1}, K_{t+1})K_t - \Phi_2(I_{t+1}, K_{t+1})K_t \\
- (1 - \delta)^2 \Phi_1(I_{t+1}, K_{t+1})K_t] < 0 (A14)
\]

where the inequality follows from Assumptions 1 and 2 and Lemma 3. Finally, plugging equations (3), (A10), (A14), and (23) into (A12) and using $\Pi_1(K_{t+1}, X_{t+1}) = \alpha(\alpha - 1)\frac{\Pi_{t+1}}{K_{t+1}}$:
\[
\frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \bigg| \frac{\partial E_t}{\partial (I_t/K_t)} \left( \frac{\Pi_t}{K_t} \right) = \frac{\alpha \rho \pi G''(I_t/K_t)}{\Phi_1(I_t, K_t)G'(I_t/K_t)} + \alpha(1 - \alpha)\rho \frac{K_t}{K_{t+1}} > 0 (A15)
\]

■

**Proof of Proposition 7** From the investment-return equation (16),
\[
\frac{\partial E_t[r_{t+1}]}{\partial Q_t} = -\frac{E_t[U_{t+1}]}{\alpha Q_t^2} + \frac{1}{\alpha Q_t} \frac{\partial E_t[U_{t+1}]}{\partial Q_t} (A16)
\]
By Lemma 4, to show \( \frac{\partial E_t[U_{t+1}]}{\partial Q_t} < 0 \), it suffices to show that \( \frac{\partial E_t[U_{t+1}]}{\partial Q_t} < 0 \). But \( I_t/K_t = G^{-1}(q_t, K_t^{-1}) \), from equation (9). Writing \( q_t \) further as \( \alpha Q_t \), plugging \( I_t/K_t \) into equation (A13), and using the Inverse Function Theorem yield \( \frac{\partial (I_t/K_t)}{\partial Q_t} = \alpha K_t^{-1} / G'(C^{-1}(\alpha Q_t, K_t^{-1})) \), where \( C^{-1}(\cdot) \) is the inverse function of \( C' \). Now by the chain rule and equation (A14),

\[
\frac{\partial E_t[U_{t+1}]}{\partial Q_t} = \frac{\alpha K_t^{1-\alpha} K_t}{G'(C^{-1}(\alpha Q_t, K_t^{-1}))} E_t[\Pi_{t1}(K_{t+1}, X_{t+1}) + 2(1-\delta)\Phi_{12}(I_{t+1}, K_{t+1}) - \Phi_{22}(I_{t+1}, K_{t+1}) - (1-\delta)^2\Phi_{11}(I_{t+1}, K_{t+1})] < 0 \quad (A17)
\]

where the inequality follows because \( \Pi_{t1} \leq 0, \Phi_{12} \leq 0, \Phi_{22} \geq 0 \), and \( \Phi_{11} > 0 \).

To establish the second inequality in the proposition, it suffices to show:

\[
\frac{\partial^2 E_t[U_{t+1}]}{\partial Q_t^2} > 0 \quad (A18)
\]

because the chain rule of partial derivatives implies that \( \frac{\partial}{\partial P_t} \left[ \frac{\partial E_t[U_{t+1}]}{\partial Q_t} \right] = -\frac{\partial^2 E_t[U_{t+1}]}{\partial Q_t^2} \). From equation (A16), \( \frac{\partial E_t[U_{t+1}]}{\partial Q_t} = \frac{2E_t[U_{t+1}]}{\partial Q_t} - \frac{2E_t[U_{t+1}]\partial Q_t}{\partial Q_t} + \frac{1}{\partial Q_t} \frac{\partial^2 E_t[U_{t+1}]}{\partial Q_t^2} \). To show equation (A18), it suffices to show \( \frac{\partial^2 E_t[U_{t+1}]}{\partial Q_t^2} \geq 0 \) because \( \frac{\partial E_t[U_{t+1}]}{\partial Q_t} < 0 \).

For notational convenience, denote the term in the conditional expectation in equation (A17) as \( W_{t+1} \) that is negative. Now substituting \( K_{t+1} = [G^{-1}(\alpha Q_t, K_t^{-1}) + (1-\delta)] K_t \) and \( I_{t+1} = I_{t+1} - (1-\delta) [G^{-1}(\alpha Q_t, K_t^{-1}) + (1-\delta)] K_t \) into equation (A17) and differentiating the equation with respect to \( Q_t \) yield:

\[
\frac{\partial^2 E_t[U_{t+1}]}{\partial Q_t^2} = -\alpha^2 K_t^{-2(1-\alpha)} K_t \frac{G''(I_t/K_t)}{[G'(I_t/K_t)]^2} E_t[W_{t+1}] + \frac{\alpha K_t^{1-\alpha} K_t}{G'(\alpha Q_t, K_t^{-1})} \frac{\partial E_t[W_{t+1}]}{\partial Q_t} \quad (A19)
\]

To show \( \frac{\partial E_t[W_{t+1}]}{\partial Q_t} \geq 0 \), it suffices to show that \( \frac{\partial E_t[W_{t+1}]}{\partial Q_t} \geq 0 \) because the first term in equation (A19) is nonnegative (Lemma 3 and Assumption 5 imply that \( W_{t+1} < 0 \) and \( G''(\cdot) \geq 0 \) because \( \Phi_{111} \geq 0 \)). But, \( \frac{\partial E_t[W_{t+1}]}{\partial Q_t} = \frac{\alpha K_t^{1-\alpha} K_t}{G'(\alpha Q_t, K_t^{-1})} E_t[\Pi_{t1}(K_{t+1}, X_{t+1}) - 3(1-\delta)^2\Phi_{12}(I_{t+1}, K_{t+1}) + (1-\delta)^3\Phi_{111}(I_{t+1}, K_{t+1})] \geq 0 \), where the inequality follows from Assumption 5 and Lemma 3.

**Proof of Proposition 8** First, when \( \Pi_t - \Phi_t \leq 0 \) or \( C_t = 0 \), the two derivatives in the proposition are exactly zero. Now consider the case when \( C_t > 0 \); so I can ignore the indicator function. Equation (28) implies that

\[
\frac{C_t}{K_t} = \frac{\Pi_t}{K_t} - \frac{G(\frac{I_t}{K_t})}{K_t^{\alpha-1}} \quad \text{or} \quad \frac{I_t}{K_t} = G^{-1}\left((\frac{\Pi_t - C_t}{K_t})/K_t^{\alpha-1}\right) \quad (A20)
\]

47
where $G^{-1}(\cdot)$ is the inverse function of $G$, and is also an increasing function because $G$ is around the neighborhood of optimal investment rate.

Now by the chain rule, $\frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (C_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{K_t^{1-\alpha}}{G'(I_t/K_t)} > 0$, where the inequality follows because Proposition 6 says that $\frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} < 0$. Next, again by the chain rule, $\frac{\partial^2 E_t[r_{t+1}]}{\partial (I_t/K_t) \partial Q_t} = -\frac{\partial E_t[r_{t+1}]}{\partial Q_t} \frac{\partial Q_t}{\partial (C_t/K_t)} + \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{\partial^2 Q_t}{\partial (C_t/K_t) \partial Q_t} = \frac{\partial^2 E_t[r_{t+1}]}{\partial (I_t/K_t) \partial Q_t} < 0$, where the inequality follows from the inequality (A18) and equation (A22). $\blacksquare$

**Proof of Proposition 9** When $O_t=0$, the two derivatives in the proposition are exactly zero. Consider the case when $O_t>0$. Now equation (29) implies that

$$O_t/K_t = G\left(\frac{I_t}{K_t}\right) K_t^{\alpha-1} - \frac{\Pi_t}{K_t} \text{ or } \frac{I_t}{K_t} = G^{-1}\left[\left(\frac{O_t}{K_t} + \frac{\Pi_t}{K_t}\right) K_t^{1-\alpha}\right] \quad (A21)$$

Now by the chain rule and Proposition 6, $\frac{\partial E_t[r_{t+1}]}{\partial (O_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} = \frac{\partial E_t[r_{t+1}]}{\partial (I_t/K_t)} \frac{K_t^{1-\alpha}}{G'(I_t/K_t)} < 0$. And again by the chain rule, $\frac{\partial^2 E_t[r_{t+1}]}{\partial (O_t/K_t) \partial P_t} = -\frac{\partial E_t[r_{t+1}]}{\partial Q_t} \frac{\partial Q_t}{\partial (O_t/K_t)} - \frac{\partial^2 E_t[r_{t+1}]}{\partial (O_t/K_t) \partial Q_t}$. To show the left-hand-side is negative, it suffices to show that $\frac{\partial^2 E_t[r_{t+1}]}{\partial (O_t/K_t) \partial P_t} < 0$ because $\frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t \partial P_t} > 0$ from equation (A24), $\frac{\partial Q_t}{\partial (O_t/K_t)} = \frac{\partial (I_t/K_t) \partial P_t}{\partial (O_t/K_t)} > 0$, and $\frac{\partial^2 Q_t}{\partial (O_t/K_t) \partial P_t} < 0$. But $\frac{\partial^2 Q_t}{\partial (O_t/K_t) \partial P_t} = \frac{\partial (1/K_t+1)}{\partial (O_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} < 0$, where the inequality follows from equation (A23) and $\frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} > 0$. $\blacksquare$

**Proof of Proposition 10** Equation (9) and Proposition 1 imply that $Q_t = G'(I_t/K_t) K_t^{\alpha-1/\alpha}$, which in turn implies that

$$\frac{\partial Q_t}{\partial (I_t/K_t)} = \frac{1}{\alpha} G''\left(\frac{I_t}{K_t}\right) K_t^{\alpha-1} > 0 \quad (A22)$$

$$\frac{\partial^2 Q_t}{\partial (I_t/K_t) \partial P_t} = -\frac{\partial (1/K_t+1)}{\partial (I_t/K_t)} \frac{\partial (I_t/K_t)}{\partial (O_t/K_t)} = -\frac{K_t}{K_t} < 0 \quad (A23)$$

Now, $\frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t \partial P_t} = -\frac{\partial E_t[r_{t+1}]}{\partial Q_t} \frac{\partial Q_t}{\partial (I_t/K_t)} - \frac{\partial^2 E_t[r_{t+1}]}{\partial (I_t/K_t) \partial Q_t}$. From $\frac{\partial E_t[r_{t+1}]}{\partial Q_t} < 0$ and equations (A22) and (A23), to show the second inequality in the proposition, it suffices to show $\frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t \partial P_t} > 0$. But,

$$\frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t \partial P_t} = \frac{\partial (\frac{\partial E_t[r_{t+1}]}{\partial Q_t} \frac{\partial Q_t}{\partial P_t})}{\partial Q_t} = \frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t^2} \frac{1}{K_t} + \frac{\partial E_t[r_{t+1}]}{\partial Q_t} \frac{\partial^2 Q_t}{\partial Q_t \partial P_t} = \frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t^2} \frac{1}{K_t} > 0$$

(A24)

where the inequality follows from equation (A18).
Moreover, \( \frac{\partial^2 E_t[r_{t+1}]}{\partial Q_t^2} \left( \frac{\partial Q_t}{\partial (I_t/K_t)} \right) \) because \( \frac{\partial^2 Q_t}{\partial (I_t/K_t)} \left( \frac{\partial Q_t}{\partial (I_t/K_t)} \right) < 0 \), where the equality follows.

B A Brief Review of the Anomalies Literature

I briefly review the list of anomalies that motivates my theoretical work in this paper. This list covers a broad scope of the literature on empirical asset pricing, empirical corporate finance, and capital markets research in accounting. See Fama (1998) and Schwert (2003) for much more complete surveys.

The Investment Anomaly Firms that disinvest earn higher future stock returns on average (e.g., Miles and Rosenfeld (1983), Cusatis, Miles, and Woolridge (1993)). Firms with high investment-to-asset earn lower returns on average than firms with low investment-to-asset (e.g., Richardson and Sloan (2003), Titman, Wei, and Xie (2004), Lyandres, Sun, and Zhang (2005), Xing (2005), and Anderson and Garcia-Feijóo (2006)). Titman et al. also show that the investment-return relation is stronger in firms with higher operating income-to-asset ratios. Cusatis et al. attribute their evidence to market underreaction. Richardson and Sloan and Titman et al. interpret their evidence as investors underreacting to overinvestment behavior of empire-building managers. Anderson and Garcia-Feijóo interpret their evidence as consistent with the real options models of Berk, Green, and Naik (1999) and Gomes, Kogan, and Zhang (2003). Xing and Lyandres et al. interpret their evidence as consistent with optimal investment as in Zhang (2005).

The Value Anomaly Value stocks earn higher average returns than growth stocks (e.g., Graham and Dodd (1934), Rosenberg, Reid, and Lanstein (1985), Fama and French (1992, 1993), and Lakonishok, Shleifer, and Vishny (1994)). Fama and French (1993) also document that the value anomaly is stronger in small firms. A closely related anomaly is the long-term reversal anomaly documented by De Bondt and Thaler (1985), i.e., losers over the past five years earn higher returns on average in the subsequent three to five years than winners over the past five years. Fama and French (1996) show that the value anomaly subsumes the reversal anomaly.

Fama and French (1993, 1996) argue that the value strategy captures common variations in stock returns, a source of undiversifiable risk in the ICAPM- or APT-framework. De Bondt and Thaler (1985) and Lakonishok, Shleifer, and Vishny (1994) instead argue that investors tend to overreact temporarily to a long string of past earnings news in the same direction, and that this mispricing tends to revert in the long run.

The Payout Anomaly Anomalous long-term positive abnormal returns apply to firms distributing capital back to shareholders. Lakonishok and Vermaelen (1990) find positive long-term abnormal returns when firms tender for their stocks. Ikenberry, Lakonishok, and
Vermaelen (1995) find that the average abnormal four-year return after the announcements of open market share repurchases is significantly positive. And the average abnormal return is much higher for value firms, but is negative although insignificant for growth firms. Finally, Michaely, Thaler, and Womack (1995) find that stock prices underreact to the negative information in dividend omissions and the positive information in initiations.

The Seasoned-Equity-Offering Anomaly Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995) document that firms conducting seasoned equity offerings earn much lower returns on average over the next three to five years than nonissuing firms with similar characteristics. Brav, Geczy, and Gompers (2000) and Eckbo, Masulis, and Norli (2000) find that this underperformance is more pronounced for small firms. A frequent conclusion in this literature is that firms time their equity-issuance decisions to exploit the systematic mispricing of their securities in capital markets (e.g., Ritter (2003)).

The seasoned-equity-offering anomaly belongs to a class of anomalies related to external financing including long-term stock underperformance following initial public offerings (e.g., Ritter (1991)), straight and convertible debt offerings (e.g., Spiess and Affleck-Graves (1999)), private placements of equity (e.g., Hertzel, Lemmon, Linck, and Rees (2002)), and bank loan announcements (e.g., Billett, Flannery, and Garfinkel (2005)).

The Expected-Profitability Anomaly Investors seem to underreact to new information on future cash flows. Shocks to expected cash flows are positively correlated with shocks to expected returns (e.g., Frankel and Lee (1998), Dechow, Hutton, and Sloan (1999), Cohen, Gompers, and Vuolteenaho (2002), Vuolteenaho (2002), and Fama and French (2006)). Cohen et al. and Vuolteenaho also find that the magnitude of this correlation is higher in small firms.

The Profitability Anomaly Controlling for market price relative to cash flows or book equity, more profitable firms earn higher average returns (e.g., Haugen and Baker (1996) and Piotroski (2000)). Piotroski also shows that this relation is stronger in small firms.

The Post-Earnings-Announcement Drift Ball and Brown (1968) and Bernard and Thomas (1989, 1990) document that stock price drifts in the direction of earnings surprise, defined as the scaled change in earnings. Bernard (1993) shows that the magnitude of the drift is inversely related to the market value. Chan, Jegadeesh, and Lakonishok (1996) confirm this evidence using time series and cross-sectional regressions. This anomaly is often interpreted as underreaction to earnings news. Because this anomaly is exceptionally robust, Fama (1998) calls it the “granddaddy” of anomalies.