Beyond Q: The Marginal Value of Capital and Corporate Investment.

Xuesi Wang*

Leonidas G. Barbopoulos†

Khaladdin Rzayev‡

This version
November 9, 2025

*University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: X.Wang-344@sms.ed.ac.uk. Corresponding author.

†University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: leonidas.barbopoulos@ed.ac.uk. Tel: +44 131 650 8344.

‡University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: khaladdin.rzayev@ed.ac.uk. Koç University, Istanbul, Turkey, and Systemic Risk Centre, London School of Economics, London, UK.

Beyond Q: The Marginal Value of Capital and Corporate Investment

Abstract

Tobin's q exhibits poor predictive power for corporate investment despite being the cornerstone of neoclassical investment theory and extensively used in empirical literature. We show that q embeds information about two distinct forward-looking channels: the expected cost of capital and the expected marginal value of capital (MVK). Using a log-linear approximation of the investment Euler equation under quadratic adjustment costs, we derive a testable prediction: holding current q constant, future investment should depend *negatively* on expected MVK. We test this prediction empirically by developing a machine-learning approach that combines neural networks with automatic differentiation to estimate firm-specific, time-varying expected MVK without imposing restrictive functional form assumptions. Incorporating expected MVK increases the explanatory power of aggregate investment regressions fivefold and raises the within-firm R^2 by about 15 percentage points. Consistent with the model, expected MVK is negatively related to future investment. These results persist after controlling for cash flow and extend to intangible capital, suggesting q's empirical failure reflects omitted variable bias rather than theoretical breakdown.

Keywords: Tobin's q; Investment; Marginal value of capital; Machine learning; Cost of capital.

JEL Classification Codes: E22; G17; G31; G32.

1. Introduction

How do firms make investment decisions? According to the neoclassical investment-q theory, value-maximizing firms determine their optimal future investment based on the marginal value of capital to future operating profits, discounted by the expected cost of capital. However, these marginal valuations are not directly observable to researchers or market participants, presenting a fundamental measurement challenge for empirical work.

Hayashi (1982) shows that under certain conditions – including perfect competition, constant returns to scale, and frictionless capital markets – Tobin's q (the ratio of market value to capital stock) can serve as a proxy for the unobservable marginal value of capital. In this framework, average q alone should determine investment, with no role for other variables. Over the past four decades, q has become the dominant empirical proxy for investment opportunities in corporate finance research (Bartlett and Partnoy, 2020).

Yet despite its theoretical elegance and widespread use, q exhibits poor empirical performance. Firm-level investment regressions typically yield low R^2 values, and the estimated coefficient on q varies dramatically across samples and time periods (Erickson and Whited, 2000; Andrei et al., 2019). This persistent gap between theory and evidence raises a fundamental question: why does q fail to predict investment? The literature offers two main explanations. First, average q may be mismeasured and does not contain accurate information about the true marginal value of capital to profit (hereafter, we express it in short as MVK)(Abel and Blanchard, 1983; Erickson and Whited, 2000). Measurement error can arise from market mispricing, intangible assets, or violations of Hayashi's assumptions such as constant returns to scale. Second, investment may be driven or constrained by factors outside the q framework, such as financial frictions or cash flow constraints, which are more important than the MVK itself (Abel, 2018; Andrei et al., 2019).

We offer a different perspective. We argue that the expected MVK is indeed a fundamental determinant of future investment, but this information is not fully captured by q alone. Using a log-linear approximation to the investment Euler equation, we first show that q embeds information about two forward-looking channels: the cost of capital (the discount-rate channel) and the marginal value of capital (the profit channel) – consistent with the recursive framework in Lettau and Ludvigson (2002). Next, to derive testable predictions, we assume standard quadratic adjustment costs and that firms form rational expectations about future investment optimality. Under these conditions, we obtain a key result: holding current q constant, future investment should be positively related to the expected cost of capital but negatively related to the expected MVK.

Both the negative coefficient on expected MVK and the positive coefficient on expected cost of capital may appear counterintuitive. The key to understanding these predictions is that they are conditional on current q, arising from the forward-looking Euler equation rather than the contemporaneous investment optimality condition. Current q already embeds the market's expectations about future profitability, costs, and optimal investment. When we condition on q and examine the incremental effects of expected MVK and cost of capital, we are asking: given a fixed current valuation, what does higher expected productivity imply about the investment path that was embedded in that valuation? Mechanically, this negative association arises because when q is held constant, higher expected capital productivity implies lower expected investment was embedded in that fixed valuation. Economically, it reflects that when higher future capital productivity is already capitalized into the value of existing capital, the optimal response is to moderate rather than accelerate investment expansion. Empirical evidence supports this view through plant-level data showing lumpy investment under frictions. Cooper and Haltiwanger (2006) find that U.S. plants often stay inactive until making large adjustments, consistent with dynamic optimization that favors intermittent rather than continuous expansion – even in high-profit periods. Likewise, Caballero and Engel (1999) show that fixed costs generate threshold dynamics with prolonged inactivity during productivity shocks, supporting moderated paths when future gains are already reflected in valuations.

To empirically test this prediction, we require an accurate ex-ante measure of the expected marginal value of capital (MVK) – a quantity that is unobservable to econometricians. Existing literature offers two main approaches, each with significant limitations. The first approach specifies parametric forms for the production and adjustment cost functions, allowing MVK to be derived analytically as the partial derivative of profit with respect to capital. For example, under a Cobb-Douglas production function, MVK equals an elasticity parameter multiplied by the revenue-to-capital ratio. While this classical approach has been widely adopted for its simplicity (Merz and Yashiv, 2007; Liu et al., 2009; Belo et al., 2014), it faces specification risks and requires strong homogeneity assumptions that may not hold in practice. The second approach assumes limited time-series dynamics for MVK – either treating it as constant over time (Frank and Shen, 2016; Gennaioli et al., 2016) or modeling it as a simple AR(1) process (Abel and Blanchard, 1983; Frank and Shen, 2016). However, such parsimonious approximations rely on historical information and suppress forward-looking content, potentially leading to substantial measurement error, particularly for firms with rapidly evolving capital stocks and cash flows (Peters and Taylor, 2017; Andrei et al., 2019).

We propose a novel machine-learning-based approach to estimate firm-specific, time-varying expected MVK that addresses these limitations. Specifically, we employ a neural network (NN) model to approximate the relationship between future operating profit and all relevant state variables observable at the present time, using cross-sectional data. The nonparametric nature of the NN allows the estimation to be data-driven and free from restrictive functional-form assumptions, enabling it to capture potentially complex and nonlinear relationships between inputs and future outcomes (Gu et al., 2020; Kaniel et al., 2023). We then apply automatic differentiation (AD) to compute the partial derivatives of predicted future profit with respect to capital. AD, which is based on backpropagation and serves as the central algorithm in neural network training, provides an efficient and precise means of computing partial derivatives numerically by systematically applying the chain rule (Rumelhart et al., 1986; Baydin et al., 2018; Margossian, 2019).[1] Because each combination of state variables reflects a firm's unique operating environment, the estimated gradients provide firm-specific MVK measures. The entire training and prediction procedure is implemented in a rolling-window fashion over time, generating time-varying MVK estimates that are constructed purely ex-ante using only historical information.

Our empirical findings strongly support the theoretical predictions at both the aggregate time-

[1] Unlike symbolic differentiation, which requires explicit functional form specification, or finite-difference numerical approximations that introduce discretization error, AD combines flexibility with accuracy.To demonstrate AD's applicability for estimating economic marginal effects, we conduct simulation experiments under different production function specifications and sample sizes. The results confirm AD's accuracy in approximating the true MVK and its effectiveness across diverse scenarios.

series and firm-panel levels. At the aggregate level, we find that market-wide q alone exhibits limited explanatory power for aggregate investment, with an R^2 of only 0.056. However, including the aggregate expected MVK – measured as the median across S&P 500 constituents – increases the R^2 to 0.253, nearly five times higher. The estimated coefficient on MVK is significantly negative, consistent with our prediction. At the firm level, panel regressions similarly confirm that expected MVK has substantial incremental predictive power beyond q: including MVK increases the within-firm R^2 by approximately 15 percentage points. The coefficient on MVK is significantly negative and economically meaningful, while the cost of capital, measured using various proxies including WACC based on CAPM and Fama-French three-factor models, as well as implied cost of capital, exhibits limited incremental explanatory power.

These results are robust to alternative specifications and persist after controlling for cash flow, which has been commonly used to explain q's predictive failures (Fazzari et al., 1987; Abel, 2018). We find that MVK and cash flow contain distinct information: MVK consistently exerts a negative effect on investment even after controlling for cash flow, whereas cash flow has a positive effect, suggesting it represents an external financing channel outside the intertemporal optimization framework. Finally, we extend our theory and estimation to settings where firms jointly manage multiple forms of capital – both traditional physical and increasingly important intangible capital. Motivated by the rising prominence of intangibles over the past two decades (Peters and Taylor, 2017; Woeppel, 2022), we apply the same machine-learning strategy to estimate the marginal value of intangible capital. Our results show that the extended framework remains valid: the marginal value of intangible capital predicts investment negatively, consistent with our theory, supporting the universal applicability of our approach as firms' capital structures evolve.

Our paper makes several contributions to the corporate investment literature. First, we contribute to understanding why Tobin's q fails empirically. Prior work identifies two main explanations: q is mismeasured due to market noise, intangibles, or violations of Hayashi's assumptions such as constant returns to scale (Hall, 2001; Gomes, 2001; Cooper and Haltiwanger, 2006), or investment is driven by financial constraints outside the q framework (Fazzari et al., 1987; Abel, 2018). We offer a complementary explanation: even when properly measured, q embeds multiple forward-looking channels – the cost of capital and marginal value of capital – that must be separately identified. By decomposing q through the forward-looking Euler equation, we show that q's empirical failure reflects omitted variable bias rather than theoretical breakdown. Additionally, our theoretical framework predicts that expected MVK should negatively predict future investment when conditioning on current q, and we confirm this prediction empirically – a relationship that, while initially counterintuitive, arises because when current valuations already capitalize expected future productivity, higher expected MVK implies that lower investment was embedded in that valuation.

Second, we contribute methodologically by developing a nonparametric machine-learning approach to estimate firm-specific, time-varying expected MVK. Classical approaches either impose restrictive assumptions such as Cobb-Douglas production functions (Merz and Yashiv, 2007; Liu et al., 2009) or rely on simple time-series models that suppress forward-looking information (Abel and Blanchard, 1983; Frank and Shen, 2016). Recently, Gandhi et al. (2020) estimate production function elasticities using nonparametric approach but requires intermediate – input data not consistently available across firms and time. Our NN+AD framework provides a data-driven, nonparametric method that directly estimates expected MVK using only widely available Compustat data, without imposing functional form restrictions. The approach combines neural networks to approximate the unobservable profit function with automatic differentiation to compute partial derivatives, potentially excelling at capturing complex nonlinearities compared to traditional econometric tools (van Binsbergen et al., 2022; De Silva and Thesmar, 2024). While current AD applications in finance primarily target derivative markets – enabling precise evaluation of first-order risks and addressing complex PDEs for option pricing (Capriotti, 2010; Fries, 2019; Capriotti and Giles, 2024) – our approach extends AD's utility to estimating partial sensitivities of firm-level fundamentals in dynamic optimization problems. Although our focus is MVK estimation, the NN+AD framework has broader applicability for tackling high-dimensional, nonlinear challenges across finance and economics (Berg et al., 2022; Fuster et al., 2022).

The rest of the paper is organized as follows. Section 2 reviews the investment–q theory and develops a model linking q, MVK, and the cost of capital. Section 3 introduces the machine learning estimation of firm-year MVK. Section 4 presents data and summary statistics, and Section 5 reports empirical results. Section 6 relates the findings to financial constraints and firms' evolving capital structures. Section 7 concludes.

2. Model

This section develops our model. We first define the dynamic optimization problem of a value-maximizing firm and derive the first-order conditions linking investment, q, MVK, and the cost of capital. We then discuss the conditions required for constructing a measurable ex-ante expectation of MVK.

2.1. Firm's dynamic problem

Let Ω_t denote the information set available to the firm at time t. At the beginning of each period t, the firm chooses investment (I_{it}) to maximize the expected present value of future operating profits by solving the following problem:

$$\begin{aligned} V_{it} = \max_{\{I_{it+s},\ K_{it+s+1}\}} \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta_{it+s}\, \Pi_{it+s} \middle| \Omega_t \right], \\ \text{s.t.} \quad \Pi_{it+s} = f(K_{it+s}, N_{it+s}) - \Phi(I_{it+s}, K_{it+s}). \end{aligned} \tag{1}$$

We drop the subscript i for simplicity in the remainder of this section. β_t denotes the stochastic discount factor (SDF) from t to $t+1$. Firms use capital K_t and labor N_t to generate revenue $f(K_t, N_t)$ and incur total adjustment costs $\Phi(I_t, K_t)$, yielding operating profits Π_t. Following Abel (2018) and Gonçalves et al. (2020), we assume costless labor adjustment so that, making adjustment costs Φ depends only on investment and capital. This focuses the analysis on how the marginal value of capital governs investment decisions.

Firm capital depreciates at a constant rate δ, and the capital accumulation constraint is:[2]

$$K_{t+1} = (1-\delta)K_t + I_t. \tag{2}$$

Let Q_t be the Lagrange multiplier associated with Eq. (2). The first-order conditions (FOC) with respect to I_t and K_{t+1} are:

$$Q_t = \Phi_I(I_t, K_t), \tag{3}$$

[2]In the empirical analysis of intangible investment-(q) theory, we estimate the depreciation rate at the industry level to better capture the specific depreciation patterns of intangible capital, following Peters and Taylor (2017) and Woeppel (2022).

$$Q_t = \mathbb{E}_t\left[\beta_{t+1}\left(\Pi_K' + (1-\delta)Q_{t+1}\right) \mid \Omega_t\right], \tag{4}$$

where Φ_I is the partial derivative of Φ_t with respect to beginning-of-period investment I_t, capturing the marginal adjustment cost of installing new capital. Π_K' is the partial derivative of Π_{t+1} with respect to the end-of-period capital stock K_{t+1}, capturing the marginal contribution of installed capital to next period's operating profits.

Eq. (3) states that the firm installs capital up to the point where the marginal installation cost equals its marginal benefit. Eq. (4) is the Euler equation capturing the recursive relationship of Q_t; the shadow value of capital reflects the expected discounted payoff from additional capital and its continuation value.

Empirical tests of the linear investment–q model typically focus on Eq. (3), assume a quadratic adjustment-cost function to obtain a closed form for $\Phi_I(I_t, K_t)$, and proxy for the unobservable q. However, Eq. (3) is purely contemporaneous and does not impose a forward-looking restriction. Time-differenced regressions (e.g., I_t on Q_{t-1}) cannot fully capture the intertemporal nature of optimal investment decisions. Moreover, linear regressions based on Eq. (3) rely on strong assumptions of firm homogeneity and frictionless capital markets (Hayashi, 1982), limiting their general applicability.[3]

By contrast, Eq. (4) is derived directly from value maximization and is robust to assumptions about q, firm heterogeneity, and functional forms (Gala et al., 2020). It characterizes an explicit intertemporal link: today's Q_t embeds the optimal expectation of next-period SDF, MVK, and continuation value conditional on Ω_t. Thus, Q_t from Eq. (4) is an ex-ante measure of the rational expectation of future capital returns. Hence, we focus on extending the nonlinear forward-looking relation implied by Eq. (4), following Lettau and Ludvigson (2002) and Gala et al. (2020).

2.2. Tobin's q, MVK, and investment-capital ratio

We decompose Q_t to characterize its information content using a log-linear approximation of Eq. (4), following Lettau and Ludvigson (2002). We rewrite the SDF in cost-of-capital form, i.e., $\beta_{t+1} = 1/(1+R_{t+1})$, and express variables in logs: $r_{t+1} = \log(1+R_{t+1})$, $q_t = \log(Q_t)$, and $\pi_{t+1} = \log(\Pi_K)$.

[3] See Gomes (2001) and Cooper and Haltiwanger (2006) for further discussion.

As shown in the Online Appendix B.1, first- and second-order Taylor expansions yield the following log-linearized representation of Eq. (4):

$$q_t \approx \kappa + \tau \mathbb{E}_t[q_{t+1}] - \mathbb{E}_t[r_{t+1}] + (1-\tau)\mathbb{E}_t[\pi_{t+1}] + \xi_t, \tag{5}$$

where κ collects constants, and ξ_t captures approximation errors. The parameter $\tau \equiv \frac{(1-\delta)\exp(\bar{q})}{(1-\delta)\exp(\bar{q})+\exp(\bar{\pi})}$ is a weighting factor strictly less than one, determined by the steady-state values of q and π, and reflects their relative importance in the capital valuation process.

Ignoring approximation errors ξ_t, Eq. (5) implies that q_t embeds expectations of future capital returns through either the cost-of-capital channel or the marginal value of capital channel (or both), analogous to the dividend-price ratio identity, which decomposes its predictive content into a discount-rate news channel (future returns) and a cash-flow news channel (future dividend growth) (Campbell and Shiller, 1988; Koijen and Van Nieuwerburgh, 2011; De La O and Myers, 2021). Applying the law of iterated expectations and the no-bubble condition $\lim_{T\to\infty} \tau^T \mathbb{E}_t[q_{t+T}] = 0$, we obtain the infinite-horizon identity:

$$q_t \approx -\sum_{j=0}^{\infty} \tau^j \mathbb{E}_t[r_{t+j+1}] + (1-\tau)\sum_{j=0}^{\infty} \tau^j \mathbb{E}_t[\pi_{t+j+1}], \tag{6}$$

where constants are suppressed. Eq. (6) highlights that higher q_t reflects either lower expected discount rates (the "cost of capital channel") or higher expected MVK (the "marginal value channel"), consistent with the interpretation in Lettau and Ludvigson (2002).

Until now, the discussion has relied solely on the capital Euler equation and does not explicitly link to investment. To do so, we must simultaneously apply the FOCs for capital and investment and impose a functional form on adjustment costs. We summarize these modeling choices in Assumption 1.1.

Assumption 1. *Adjustment costs are quadratic in the investment–capital ratio:* $\Phi(K_t, I_t) = (a/2)(I_t/K_t)^2 K_t + I_t$, $a > 0$. *The optimal ex-ante expectation of* q_{t+1} *equals the conditional expectation of the marginal adjustment cost:* ϕ'_I*:*

$$\mathbb{E}_t[q_{t+1}] = \mathbb{E}_t[\phi'_I] = \mathbb{E}_t[log(a\frac{I_{t+1}}{K_{t+1}} + 1)]. \tag{7}$$

This specification follows Frank and Shen (2016) and Abel and Panageas (2022), enabling an analytical and empirically testable mapping between q and the investment-capital ratio. While richer specifications exist (e.g., Hall, 2001; Barnett and Sakellaris, 1998; Merz and Yashiv, 2007), they either require additional firm-level inputs or yield more complex predictive structures – considerations beyond the scope of this paper. In the following section of this paper, we provide a more detailed discussion on the effects of the setting of adjustment cost functions.

Assumption 1 uses rational expectations rather than imposing contemporaneous optimality from Eq. (3). That is, we do not require $q_t = \Phi_I$, but instead assume that next-period optimality informs today's expectations of future investment. Similar forward-looking treatments appear in Liu et al. (2009) and Belo et al. (2013).

Substituting Eq. (7) into Eq. (5), defining $\Gamma_{t+1} = I_{t+1}/K_{t+1}$ and $\gamma_{t+1} = \log(\Gamma_{t+1})$, and applying log-linear approximation yields:

$$\mathbb{E}_t[\gamma_{t+1}] = \vartheta + \frac{1}{\nu} q_t + \frac{1}{\nu}\mathbb{E}_t[r_{t+1}] - \frac{1-\tau}{\nu}\mathbb{E}_t[\pi_{t+1}], \tag{8}$$

where ϑ contains constants, $\nu \equiv \tau \frac{a\exp(\bar{\gamma})}{a\exp(\bar{\gamma})+1}$, and the approximation error terms are suppressed. Since $\tau < 1$ and $\nu > 0$, Eq. (8) predicts that the expected investment–capital ratio increases with q_t and the expected cost of capital, and decreases with expected MVK. Like the dividend-price identity, this result applies both in the aggregate and cross-section.[4]

For a rational firm manager who knows the parameters ν and τ, next period's optimal investment–capital ratio is determined using current q_t together with the expected cost of capital and expected MVK, where ν and τ serve as weights reflecting the relative importance of each component. To put it differently, from a firm's perspective, ν and τ determine how much information current q_t conveys about future investment. From an econometric perspective, Eq. (8) provides a direct explanation for the empirical under-performance of q_t as a predictor of investment. The investment–capital ratio depends not only on q_t, but also on the expected cost of capital and the expected MVK. If either of these components contains independent forward-looking information relevant for investment, then regressions that include only q_t will omit relevant state variables. This omission

[4] See Koijen and Van Nieuwerburgh (2011) for aggregate evidence and Delao et al. (2025) for cross-sectional tests.

biases the estimated q_t coefficient and reduces the explanatory power of the regression.

The severity of this omitted-variable bias depends on the cross-sectional distribution of firms. When firms differ widely in their expected discount rates or marginal productivity, the dispersion in q_t reflects not just investment expectations, but also heterogeneous cost-of-capital and cash-flow news. In this case, the same observed q_t value could imply very different future investment outcomes across firms, making the q_t-investment linkage appear weaker, but dependent on the comovement between the observed q_t and ignored MVK.

In contrast, when q_t already fully incorporates information about future investment – because expected cost of capital and MVK are largely homogeneous or strongly correlated with q_t – including additional predictors should not materially improve predictive performance. Thus, Eq. (8) highlights when q_t should be informative about investment, and when it should not, providing a unified interpretation for both strong and weak empirical results in the literature. As we shown in the following sub-sample tests, this mechanism offers a natural explanation for the puzzling empirical finding in Andrei et al. (2019) that the predictive coefficient on q_t declines precisely among firms with greater dispersion in q_t.

Why would a higher expected marginal value of capital lead to lower expected future investment? Two perspectives can clarify this seemingly counterintuitive result. First, when investment involves convex adjustment costs, a firm expecting high future marginal productivity may already be operating near its optimal adjustment level. If the expected gains are not driven by lower investment costs, the firm can choose to restrain future investment to avoid excessive adjustment costs. Empirically, investment often declines under adjustment-cost uncertainty, even when productivity is strong (Gordon, 2004; Bloom et al., 2007). Second, this negative relation arises only after conditioning on present optimality, captured by the known and fixed value of q_t. The capital first-order condition embeds expectations of future investment into the current value of q_t. Holding q_t and financing costs constant, a higher expected MVK indicates that firms have already advanced investment to the present period in anticipation of higher future productivity. As a result, less additional investment will be required later. Consistent with this, the literature shows that firms build capital earlier so it is in place when productivity increases, leading to lower expected investment in the next period

(Beaudry and Portier, 2006).

2.3. The conditional expectation of MVK

The preceding discussion highlights that empirical implementation of Eq. (8) requires a reliable estimate of the conditional expectation of the marginal value of capital, $\mathbb{E}_t[\pi_{t+1}]$, which is not directly observable. To construct such a measure, we impose the following assumptions on the firm's operating profit:

Assumption 2. *The profit function* Π_t*:*

(i) is smooth and continuously differentiable in K_t*;*

(ii) has bounded gradients with respect to K_t*; and*

(iii) evolves as an unobservable random process governed by the state variables K_t, N_t*, and* I_t*.*

We impose no concavity restriction on Π with respect to capital; thus, the marginal value of capital may be non-diminishing or even negative. Assumption 2 places only general restrictions on regularity and boundedness and does not constrain the functional form of Π. These conditions ensure the existence of finite first-order derivatives.

We next impose a timing assumption to guarantee that the conditional expectation of forward-looking profits is measurable given information at time t. Specifically, consistent with Gala et al. (2020),[5] investment (or hiring) undertaken at the beginning of period t affects the end-of-period capital stock K_{t+1} (or labor N_{t+1}), but does not contribute to current production. By contrast, adjustment costs are incurred immediately in period t and reduce current operating profits. Under this structure, the end-of-period capital stock is deterministic at time t conditional on the information set Ω_t. We denote this as $\tilde{K}_{t+1} = (1-\delta)K_t + I_t$ and $\tilde{N}_{t+1} = (1-\delta)N_t + H_t$.

As shown in Online Appendix B.2, when $\tilde{K}_{t+1}$ is deterministic under the filtration Ω_t, the expected marginal value of capital can be approximated by the partial derivative of $\mathbb{E}_t[\Pi_{t+1} \mid \Omega_t]$

[5] This timing assumption is standard in the investment-q literature, which typically adopts a lag–lead structure between capital stocks and investment decisions.

with respect to $\tilde{K}_{t+1}$:

$$\mathbb{E}_t[\Pi'_K \mid \Omega_t] = \mathbb{E}_t\left[\frac{\partial \Pi_{t+1}}{\partial K_{t+1}}\middle|\Omega_t\right] \approx \frac{\partial\, \mathbb{E}_t[\Pi_{t+1} \mid \Omega_t]}{\partial \tilde{K}_{t+1}}. \tag{9}$$

This approximation relies on interchanging conditional expectation and differentiation, justified by Assumption 2, which requires that Π_{t+1} is sufficiently smooth and that the filtration Ω_t is well defined. The result is closely related to Fries (2019), who set similar conditions for stochastic systems. Although this approach substantially simplifies the estimation of MVK, it does not eliminate approximation errors. Such errors represent one source of the measurement term ξ_t in Eq. (5). When nontrivial, ignoring these errors in empirical implementation may lead to deviations from the theoretical predictions.

3. Machine learning estimation of expected MVK

This section introduces our approach to estimating firm-specific, time-varying expected MVK. We propose a nonparametric method combining neural networks with automatic differentiation, offering flexibility relative to conventional approaches while avoiding restrictive functional form assumptions.

3.1. Estimating MVK – conventional vs. machine learning approach

As illustrated by Eq. (9),estimating the conditional expectation of MVK is approximately equivalent to estimating the marginal effect of forward capital on the conditional expectation of future operating profit. The conventional approach assumes a specific production function with constant returns to scale, yielding an analytical expression for MVK:

$$\mathbb{E}_t[\Pi'_K \mid \Omega_t] = \mathbb{E}_t[f'_K - \Phi'_K \mid \Omega_t] = \mathbb{E}_t\left[\alpha\frac{Y_{t+1}}{K_{t+1}} + \frac{a}{2}\left(\frac{I_{t+1}}{K_{t+1}}\right)^2\middle|\Omega_t\right],$$

where the production function f is typically specified in Cobb-Douglas form, i.e., $Y_t = AK_t^{\alpha}L_t^{1-\alpha}$. Under this specification, f'_K equals the sales-to-capital ratio scaled by the elasticity parameter α, while Φ'_K is derived from the quadratic adjustment cost function in Assumption 1.

While constant returns to scale makes MVK theoretically measurable, accurate implementation is challenging in practice. It requires reliable proxies for the expected sales-to-capital and squared investment-to-capital ratios, as well as ex-ante estimates of the unknown parameters α and

a. Moreover, the empirical validity of the Cobb-Douglas specification is questionable. Its restrictive assumptions – including constant returns to scale and homogeneous production technology across firms – make it more suitable as a simplifying device for analytical tractability than as an accurate empirical measure (Biddle, 2012; Gopinath et al., 2017).

Alternatively, MVK can be approximated using the profit-capital ratio $\frac{\mathbb{E}_t[\Pi_{t+1}]}{K_{t+1}}$, circumventing restrictive functional form assumptions. This approach requires only a proxy for $\mathbb{E}_t[\Pi_{t+1}]$, typically obtained from analyst earnings forecasts (Cummins et al., 2006; Gennaioli et al., 2016) or by assuming profits follow an AR(1) process (Frank and Shen, 2016). However, this simplification raises two concerns. First, firms with negative earnings forecasts are commonly excluded to ensure positive MVK, yet persistently negative expected earnings are frequent across industries and do not necessarily imply negative marginal value of capital. Second, analyst forecasts suffer from well-documented biases, including excessive optimism, conservatism, and herding behavior, that cast doubt on their reliability as proxies for rational expectations (Hou et al., 2012; Kothari et al., 2016; De Silva and Thesmar, 2024).

To address these concerns, we propose a novel estimation strategy using neural networks combined with automatic differentiation (AD). This approach simultaneously approximates the unobserved profit function and computes its partial derivative with respect to capital, yielding firm-specific estimates of MVK. Specifically, we train a neural network to model the relationship between future operating profits and current state variables. During training, AD computes numerical partial derivatives with respect to capital at each firm-year observation, generating time-varying, firm-specific MVK estimates in a data-driven manner without imposing functional form restrictions.

To implement this estimation, we must specify which state variables determine future operating profit. Forward profit Π_{t+1} depends on: (i) productive inputs K_{t+1} and N_{t+1}, which generate revenue; and (ii) adjustment costs, which depend on K_{t+1} and I_{t+1}. While forward capital and labor are deterministic at time t through their accumulation processes, future investment I_{t+1} is not directly observable. However, we can capture the information content of I_{t+1} using present state variables. Specifically, from Eq. (8), optimal investment depends on $\{Q_t, \mathbb{E}_t[MVK], \mathbb{E}_t[r_{t+1}]\}$, and Q_t itself can be expressed as a function of $\{K_t, I_t\}$ through the investment FOC (Eq. (3)). Therefore, the information relevant for I_{t+1} is fully captured by the observable state vector $\{K_{t+1}, \mathbb{E}_t[r_{t+1}], K_t, I_t\}$. . This allows us to approximate the forward profit function using only current observables: $\Pi_{t+1} = f\big(\tilde{K}_{t+1}, \tilde{N}_{t+1}, \tilde{r}_{t+1}, K_t, I_t\big)$.

This approximation strategy is consistent with Gala et al. (2020) and Gala et al. (2022), who project investment onto present state variables using polynomial approximations. We extend this intuition to recover the information content necessary for estimating unobservable future investment, rather than estimating structural investment equations explicitly. Empirically, we train a neural network on the following projection using historical cross-sectional data:

$$\Pi_{t-s+1} = \mathcal{F}_\Theta\big(\tilde{K}_{t-s+1},\, \tilde{N}_{t-s+1},\, \tilde{r}_{t+s+1},\, I_{t-s},\, K_{t-s}\big),$$

where s denotes the length of the training window. The network parameters Θ are estimated by minimizing a loss function ℓ – the mean absolute error between fitted values $\mathcal{F}_\Theta$ and actual profits $\Pi_{i,t-s+1}$:

$$\Theta^* = \arg\min_\Theta \sum_n \ell\big(\Pi_{i,t-s+1},\, \mathcal{F}_\Theta\big).$$

Once the neural network $\mathcal{F}_{\Theta^*}$ is fitted, its functional form and parameters are fixed. AD, applied during training, preserves the learned relationship between inputs and outputs, enabling direct extraction of marginal effects. Given new state variables observed at time t, we apply AD to the fitted network to obtain an ex-ante estimate of expected MVK:

$$\mathbb{E}_t[\widehat{\Pi}'_K] = \frac{\partial\, \mathcal{F}_{\Theta^*}\big(\tilde{K}_{t+1},\, \tilde{N}_{t+1},\, \tilde{r}_{t+1},\, I_t,\, K_t\big)}{\partial \tilde{K}_{t+1}}.$$

Ultimately, expected MVK is estimated ex-ante using nonparametric methods – neural network approximation of the profit function combined with AD (NN+AD). The estimated MVK quantifies how incremental changes in capital stock affect expected profits, conditional on other state variables, based on the technological relationships learned from historical data.

We implement this estimation empirically following a three-step procedure: (i) Initial calibration (using data from 1970 to 1974): Select optimal hyperparameters by training on 1970-1972 data and testing on 1974 out-of-sample, i.e., the optimal hyperparameters will enable the model obtained from the 1970-1972 data to minimize forecast error of Π on the test period 1974. The year between the end of train and test period (here 1973) is skipped to ensure that at the test time, the one-year-ahead realized profit using in the training periods are observable, and to avoid look-ahead bias.[6] (ii) Rolling estimation: For each year t from 1975 onwards: with the obtained optimal hyperparameters, train neural network $\mathcal{F}_{\Theta^*}$ on past 3 years of cross-sectional data (years $t-4$ to $t-2$, year $t-1$ skipped).[7] For each firm i at time t, compute firm-specific MVK, $\mathbb{E}_t[\widehat{\Pi}'_K] = \partial \mathcal{F}_{\Theta^*}/\partial \tilde{K}_{t+1}$, using automatic differentiation. (iii) Repeat: Move forward one year and repeat step (ii) until the last year 2023. As a result, this procedure generates a panel of firm-year MVK estimates from 1975 to 2023, constructed purely ex-ante using only historical information available at time t.

To illustrate how AD works, Table 1 shows an example of using a profit function with Cobb-Douglas production and quadratic adjustment costs. A neural network is first trained to approximate the relationship between profits and state variables (K, N, I). The trained network evaluates the function at given inputs $(v_{-1} = K, v_0 = N, v_{-2} = I)$, sequentially computing intermediate values (v_1, v_2, v_3) in the forward pass. Reverse-mode AD then propagates gradients backward through these nodes, efficiently computing input-level gradients $(v'-1, v'0, v'-2, v'-3)$ (reverse trace). The AD-estimated gradients $(\partial_K \mathcal{F}_\Theta, \partial_N \mathcal{F}_\Theta, \partial_I \mathcal{F}_\Theta)$ equals 0.13, 1.49, and -0.60, respectively, closely match the analytical derivatives $(\partial_K F, \partial_N F, \partial_I F) = (0.11, 1.52, 0.50)$, confirming AD's accuracy.

Overall, our estimation leverages AD to compute exact derivatives through the neural network's computational graph, avoiding the approximation errors inherent in finite-difference methods, such as round-off error from finite precision arithmetic (Rumelhart et al., 1986; LeCun et al., 2015). The flexibility of neural networks accommodates highly nonparametric functional forms, allowing the profit-capital relationship to be learned directly from data without imposing parametric restrictions. Consequently, MVK estimates are firm-specific and time-varying, capturing heterogeneous marginal valuations. To our knowledge, this is the first application of AD for estimating firm-level marginal values in a fully nonparametric corporate finance setting.

More broadly, our MVK estimation approach extends the growing application of machine learning

[6]Following Gu et al. (2020) and Chen and McCoy (2024), we focus on key hyperparameters known to significantly affect neural network performance. Details of the hyperparameter tuning procedure are provided in the Online Appendix C.2.

[7]We use three years of training length to ensure sufficient data in the model estimation. We find that longer training lengths, like five years, do not significantly affect the estimation of MVK.

in empirical finance, where flexible nonparametric methods often outperform traditional parametric approaches in high-dimensional, nonlinear settings (Gu et al., 2020; van Binsbergen et al., 2022; De Silva and Thesmar, 2024). Neural networks are particularly well-suited for this task: they can approximate complex, nonlinear relationships without assuming specific functional forms (Hornik et al., 1989), and through backpropagation and gradient descent, they iteratively learn patterns in the data that are difficult to capture with linear models or simple polynomials.

3.2. Simulation test

We validate the accuracy of our gradient estimation approach using Monte Carlo simulations. The procedure consists of three steps. First, we specify several data-generating processes (DGPs) for the profit function Π and derive the analytical partial derivatives with respect to K, which serve as the true MVK. Second, we construct a neural network to approximate the functional relationship between Π and the state variables using simulated data. To ensure the approach is applicable to conditional forecasting, we estimate MVK on out-of-sample data; the model is trained on a subset of the data and then used to estimate MVK on the remaining observations. Finally, we compute estimated MVK using both AD and the conventional finite-difference method, and compare their accuracy by measuring deviations from the true MVK. For the finite-difference benchmark, we adopt the central difference (CD) approximation:

$$\frac{\partial h(x_0)}{\partial x_0} \approx \frac{h(x_0 + n) - h(x_0 - n)}{2n} + O(n^p),$$

where $n > 0$ is a small step size and $O(n^p)$ denotes the truncation error from neglecting higher-order terms in the Taylor expansion.

We use three classical production function specifications to generate output: linear, Cobb-Douglas, and constant elasticity of substitution (CES). These functional forms have been widely tested in prior studies and allow analytical solutions for derivatives (Shapiro, 1986; Gandhi et al., 2020). Parameter values are calibrated based on existing empirical estimates. In all cases, the adjustment cost function Φ follows the standard quadratic form specified in Assumption 1.

Table **2** summarizes the Monte Carlo results. Panel A displays the production function specifications and parameter values. Panels B and C report summary statistics for actual and estimated MVK (from CD and AD), along with two accuracy metrics: mean absolute error (MAE) and mean absolute percentage error (MAPE). Overall, when the model is trained with sufficient data (using 10000 data points), AD-estimated MVK closely matches the true distribution across all production functions, with similar means, standard deviations, and medians. By contrast, CD estimates deviate substantially from true values.

Regarding accuracy, AD consistently outperforms CD across all specifications. With 2,500 simulated observations, AD achieves MAE values of 0.037, 0.029, and 0.022 and MAPE values of 2.3%, 4.5%, and 3.5% for the linear, Cobb-Douglas, and CES specifications, respectively. In contrast, CD produces MAE more than ten times higher and MAPE more than twenty times higher. When sample size increases to 10,000 observations, AD accuracy improves further (MAE from 0.012 to 0.016, MAPE from 1.3% to 3%), while CD errors remain largely unchanged. These results demonstrate that with sufficient training data, AD delivers stable, accurate estimates regardless of functional form complexity. Moreover, out-of-sample MVK estimates maintain high accuracy and in some cases even outperform in-sample estimates, confirming the approach's reliability for ex-ante prediction.

Overall, these simulations validate the NN+AD approach: AD-based gradient estimation is robust, accurate across diverse functional forms, and performs well out-of-sample. In contrast, conventional finite-difference methods exhibit persistent large errors unsuitable for reliable MVK estimation.

4. Data and summary statistics

We obtain firm-year accounting data from Compustat to generate the annually estimation of MVK. We use operating income before depreciation (Compustat item *oibdp*) to donate as the operating profit Π. This is because *oibdp* is one of the commonest measurement of firm's profitability or cash flow among empirical corporate finance studies (Frank and Shen, 2016; Gala et al., 2022). Additionally, data for *oibdp* is relatively readily available and is usually non-missing from Compustat.

Next, we construct all the necessary state variables $\tilde{K}_{t+1}, \tilde{N}_{t+1}, \tilde{r}_{t+1}, K_t, I_t$ used to fit the profit function by the following process. The one-year-ahead planned physical capital value is calculated by capital accumulation process: $\tilde{K}_{t+1} = (1-\delta)K_t + I_t$, with K_t measured by the book value of gross property, plant, and equipment (item *ppegt*); I_t is capital expenditure (item *capx*), and δ taken fixed value of 0.1 consistent with Cummins et al. (2006) and Hall (2001). To calculate the forward labor force, we follow a similar accumulation process based on the present data: $\tilde{N}_{t+1} = (1-\delta^N)N_t + H_t$, with N_t is the number of employees (item *emp*). δ^N is the quit rate set to be 0.22 to match the result from the Job Openings and Labor Turnover Survey program at the Bureau of Labor Statistics, as suggested by Belo et al. (2023). As the firm-level labor hiring is unavailable from Compustat, we first calculate a last period's net hiring by reverting the labor accumulation process, i.e., taking the difference between this year's number of employees and lagged number of employees times one minus quit rate. Next, the present hiring is assumed to be a rolling average hiring numbers computed by the past 10-year-period's (contain at least one-year non-missing values) hiring amount. This method is consistent with Belo et al. (2023), ensuring our measure of labor force totally based on *emp* without introducing extra variables.

We then fit the neural network model to non-parametrically approximate the relationship between future profit and these present measurable state variables, and obtain the AD estimation of $\tilde{K}_{t+1}$'s gradients as MVK, using steps described in Section 3. Our sample include all available Compustat firms excluding financial (SIC codes 6000–6799), utilities (SIC codes 4900–4999), and public administration and non-classified firms (SIC codes 9000–9999). Besides, we require the operating profit and all the four state variables are non-missing. Eventually, we obtain 210,000 firm-year conditional expectation of MVK, spans from 1975 to 2023.

Lastly, we estimate three cost of capital measures and use their present value as proxy of $\mathbb{E}_t[r_{it+1}]$. By doing this, we are assuming the expected cost of capital follows an AR (1) process, as assumed by Abel and Blanchard (1983) and Frank and Shen (2016). Specifically, we estimate two measures of cost of equity (r^e_{it}) based on expected returns estimated using CAPM and Fama-French three-factor model, respectively. Next, we follow Gonçalves et al. (2020) to estimate pre-tax cost of debt (r^d_{it} as the ratio of total interest and total debt. This approach helps to generate a broader sample of valid cost of debt and is less biased by estimation error, compared to the prior studies which relying on firm-level credit scoring data.[8] With r^e_{it} and r^d_{it}, WACC is computed as:

$$r_{it} = (1 - \frac{D}{V}) \times r^e_{it} + \frac{D}{V}(1 - Lev) \times r^d_{it},$$

where $\frac{D}{V}$ is firm's market leverage, measured by book value of debt divided by market value. Lev is firm's average tax rate (item *txt* / *pi*). Using cost of equity estimated by CAPM and Fama-French three-factor model, we obtain two WACC, donated as r^{CAPM}_{it} and $r^{\text{FF-3}}_{it}$, respectively.

One potential issue for the classical WACC measures is the estimation of cost of equity relies substantially on historical data, which may not accurately capture the forward-looking information content embedded in cost of capitals (Pástor et al., 2008; Lee et al., 2020). To address this concern, we also estimate the implied cost of capital (ICC) as one alternative measure, based on analyst forecast of earnings data from I/B/E/S, using a finite Gordon dividend discount model. The detailed estimation of these three tested cost of capital measures are shown in Online Appendix D.

In the time-series test at aggregate market level, the aggregate market's investment-capital ratio and Tobin's q using data from Federal Reserve Economic Data (FRED), following the method by Andrei et al. (2019).[9] Next, we compute an aggregate expected MVK by the median of S&P 500 constitutes, where we assume the S&P 500 universe as a proxy for the aggregate market. In each year, the aggregate MVK is formed using on average around 350 firms from the index constitute. As for the aggregate market's cost of capital, we employ the following four measures: (i) $\mathbb{E}_t[r^{LS}_{t+1}]$, the economists' expected forward stock market rate of return, calculated as the ratio of median forecast of S&P 500 index to the last year's realized value, where the economists' forecast data is obtained from Livingston Survey by Philadelphia Fed; (ii) r^{CAPM}_t, the aggregate median of S&P 500 constitutes' WACC with cost of equity estimated using CAPM; (iii) $r^{\text{FF-3}}_t$, the aggregate median of S&P 500 constitutes' WACC with cost of equity estimated using Fama-French three-factor model; and (iv) r^{ICC}_t the implied cost of capital estimated from the dividend discount model.

[8]As a common approach, firms with no credit ratings data in Compustat are usually imputed by using the coefficients estimated from the ordered probit regression of observed credit rating and several explanatory variables, such as Liu et al. (2009). However, this imputation process is likely to introduce more measurement errors (Gonçalves et al., 2020).

[9]Andrei et al. (2019) originally conduct quarterly variables using data from Fed's Flow of Funds and the Bureau of Economic Analysis (BEA). However, due to several modifications afterwards, some data are no longer available. We then use the updated code shared online by William Mann to replicate the same computations using data from FRED. The detailed definition of variables can be found in our Online Appendix A.

Other variables used in the frim-level test are defined as the following. The Tobin's average q is firm's market value scaled by physical capital stock, where the firm's market value is market value of outstanding equity (item *prcc_f* times *csho*), plus the book value of debt (item *dlcc* + *dlt*, set to be zero if missing), minus the current asset value (item *act*). Investment-capital ratio is defined as the investment expenditure (item *capx*) scaled by the physical capital stock (item *ppegt*).

Table **3** presents the summary statistics for our estimated MVK and all other relevant variables used in investment predictive regressions, both at aggregate time-series and firm-level data. The investment-capital ratio demonstrate stable values at both aggregate and firm-level, with 0.009 and 0.119 standard deviation, respectively. In comparison, both macro- and firm-level Tobin's q exhibit much greater variation, in consistent with the common empirical findings (e.g., Bakke and Whited (2010); Liao et al. (2020)). Our machine learning estimation of the conditional expected MVK also show a more balanced distribution at both aggregate and firm-level, with mean value of 0.295 and 0.217, respectively. Over the whole firm-level sample, the majority estimation MVK us lower than 0.4, showing that the partial benefits from capital stock is expected to generate lower-than-unity future profit. This result is unsurprising, as the magnitude of capital stock is usually three times larger than operating profit. As for the cost of capital measures, their distribution is analogous to the existing empirical estimations (e.g., Li et al. (2013); Frank and Shen (2016)).

Figure 1 displays the industry distribution over our sample and their median estimated MVK. Sectors such as manufacturing and durables exhibit markedly higher marginal value of capital, consistent with their intensive use of tangible assets that can be rapidly scaled and converted into operating profits with relatively low adjustment costs. In contrast, HighTech and Health show much lower MVK, reflecting their heavy reliance on intangible capital such as R&D, software, and organizational know-how (see (McGrattan and Prescott, 2010)), as well as faster depreciation and longer implementation lags that dampen the measured return on physical investment. This cross-industry pattern accords with prior evidence on heterogeneous investment frictions and non-convex adjustment costs ((Asker et al., 2015)) and thus provides an external validity check: the estimated MVK captures economically meaningful differences across sectors rather than pure noise, lending additional support to the reliability of our MVK estimates.

5. Empirical test of investment predictive regression

In this section, we empirically test the predicted links among investment, q, the conditional expectation of MVK, and the cost of capital, as outlined in Section 2. We first present time-series evidence from investment–predictive regressions using aggregate market and macroeconomic data. Next, we report panel regression results at the firm–year level constructed by Compustat data. Finally, we conduct sub-sample tests to examine the magnitude of omitted-variable bias when q is used as the sole regressor while MVK is ignored.

5.1. Aggregate expected MVK and macro-economic

Before turning to firm-level analysis, it is useful to examine the aggregate co-movement of investment, Tobin's q, and the marginal value of capital, which reflects the intertemporal trade-offs underlying capital formation and, ultimately, output and employment (Bloom, 2009). Time-series tests of Eq. (8) provide a structural decomposition of aggregate q into "cost-of-capital news" and "marginal-value news," clarifying how shifts in financial conditions and uncertainty propagate to real activity.

Figure 2 plots and compares the one-period-ahead aggregate investment–capital ratio, current q, and expected MVK. Consistent with the quarterly patterns documented by Andrei et al. (2019), we find that annual q and investment have exhibited an increasingly positive correlation since the early 2000s. By contrast, our aggregate MVK shows a consistently negative relationship with investment, in line with our theoretical prediction. Moreover, MVK typically rises at the onset of recessions while investment declines. This pattern is intuitive: during recessions, firms often cut investment to reduce exposure to uncertainty and preserve liquidity, which causes the capital stock to shrink through normal depreciation. As a result, the potential value of an additional unit of capital becomes high, increasing its expected marginal value. Conversely, during expansions, when firms have already accumulated substantial capital, the marginal value of further increasing capital is likely to be low.

We then estimate the following time-series regression of the aggregate investment–capital ratio on q, the cost of capital $\mathbb{E}_t[r_{t+1}]$, and expected MVK $\mathbb{E}_t[\pi_{t+1}]$:

$$\gamma_{t+1} = b_0 + b_1 q_t + b_2 \mathbb{E}_t[r_{t+1}] + b_3 \mathbb{E}_t[\pi_{t+1}] + \epsilon_t,$$

where all variables are in logarithms.

Table 4 reports the time-series regression results. When macro-q is used as the sole regressor, its explanatory power for investment is limited: the R^2 is only 0.056, and the estimated coefficient is marginally significant at the 10% level. Once the marginal value of capital (MVK) is added, the R^2 rises to 0.253—almost five times the explanatory power of q alone. The estimated MVK coefficient is negative and highly significant at the 1% level, indicating that, holding q constant, a one-percent increase in the marginal value of capital reduces the investment–capital ratio by about 0.07% (in log–log terms). Moreover, the predictive role of MVK remains robust when including various measures of the cost of capital: adding r slightly reduces the magnitude of the MVK coefficient but leaves its statistical significance unchanged. In addition, incorporating $r^{\text{FF-3}}$ or r^{ICC} substantially improves the R^2, whereas r^{LS} or r^{CAPM} provide only limited incremental explanatory power. The sign of r measured using accounting-based data (i.e., the two WACC and ICC measures) shows positive predictive power for investment, consistent with our theoretical prediction; whereas r derived from economists' expectation data yields a negative sign, contrary to the prediction. This result highlights the differing information content of alternative r measures at the aggregate level.

Overall, the time-series evidence supports our main aggregate-level predictions: (i) univariate q fails to capture the variation in investment; (ii) expected MVK contains significant incremental information and negatively predicts investment beyond q; and (iii) the cost of capital positively predicts investment, although the magnitude of its effect depends on how it is measured.

5.2. Test of investment–predictive regression: firm-level data

We next examine the relationship between investment, Tobin's q, MVK, and the cost of capital at the firm–year level. To construct the sample, we follow standard approach and exclude firms with non-positive total assets, non-positive sales, or *ppegt* below \$5 million, as in Peters and Taylor (2017). Observations with negative investment–capital ratios, MVK, or Tobin's q are also excluded to ensure valid logarithmic transformations. The final sample contains 110,977 firm–year observations, covering 12,471 unique firms from 1975 to 2023. All regression variables are winsorized at the 1% level.

Specifically, we estimate the following panel regression using both the ordinary least squares (OLS) estimator and the fourth-order cumulant estimator:

$$\gamma_{it+1} = b_0 + b_q q_{it} + b_1 \mathbb{E}_t[r_{it+1}] + b_2 \mathbb{E}_t[\pi_{it+1}] + FE + \epsilon_{it},$$

where γ_{it+1}, q_{it}, and $\mathbb{E}_t[\pi_{it+1}]$ are measured using either total capital or physical capital, and the cost of capital $\mathbb{E}_t[r_{it+1}]$ is taken from one of three alternative measures, i.e., $\mathbb{E}_t[r_{it+1}] \in \{r_{it}^{\text{CAPM}}, r_{it}^{\text{FF-3}}, r_{it}^{\text{ICC}}\}$. The fixed effects term FE includes both firm and time fixed effects. For estimation using the cumulant estimator, all variables are demeaned by firm and by year to account for these fixed effects prior to estimation.

Table 5 reports the panel regression results. Overall, our predictions are confirmed in the firm-level data. As shown in Panel A, MVK exhibits a significantly negative partial effect on future investment when holding q constant: a one-percent increase in MVK reduces the investment–capital ratio by about 0.172%. Moreover, including MVK consistently increases the model's explanatory power, as reflected by a 15% rise in R^2, from 0.25 to 0.283. Although the positive effect of the cost of capital (when measured by r^{CAPM} or $r^{\text{FF-3}}$) aligns with our theory, adding the cost of capital provides little incremental explanatory power for investment. Furthermore, the inclusion of the cost of capital does not materially affect the estimated coefficients of either q or MVK. When using ICC as the proxy for the cost of capital, however, the coefficient becomes negative—opposite in sign to the coefficients of r^{CAPM} and $r^{\text{FF-3}}$. This difference is likely driven by the distinct estimation approaches of these measures as well as the different information they embed.

Table 5 Panel B reports the same specification estimated with the error-correction method, replacing the OLS estimator with the higher-order cumulant estimator. The signs of all estimated coefficients remain consistent with those in Panel A; however, the magnitudes of the q and cost-of-capital coefficients change substantially, indicating significant bias when OLS is used. This finding is consistent with the evidence in Erickson and Whited (2012). In comparison, the estimated coefficient for MVK slightly reduce to -0.207, showing its relative robustness regarding to the use of estimators. Moreover, the coefficient of determination (ρ^2)—equivalent to the within-firm R^2 in panel regressions—increases by 26% after including MVK. It is also noteworthy that when q is the sole independent variable, the Hansen–J test yields a p-value of 0.02, rejecting the null hypothesis of valid over-identifying restrictions at the 5% significance level. This suggests potential model misspecification or omitted variables, implying that additional relevant state variables may be required. By contrast, when MVK is added as an additional regressor, the p-value rises to 0.768, failing to reject the null and suggesting that the specification becomes consistent with the over-identifying restrictions, thereby alleviating concerns about omitted-variable bias.

The panel regression results further confirm our predictions at the firm level. While the expected marginal value of capital provides substantial incremental information about future investment, the cost of capital—representing the discount-rate channel—does not have a significant effect. This diminished role of the cost side in investment decisions aligns with the recent survey evidence of Sharpe and Suarez (2021), who document that firms' investment plans respond only moderately to changes in interest rates, while CFOs place greater emphasis on the stability of cash flows and profitability. Our empirical findings therefore provide additional evidence consistent with these observations.

5.3. Measuring the omitted-variable bias: subgroup test

Our previous analyses highlight the importance of including MVK as a separate predictor in the investment–q regression, while showing only a negligible effect from adding the cost of capital. This is evident from both the biased estimation of the q coefficient and the changes in R^2. Moreover, we find that including MVK not only improves the overall predictability of investment but also alters the sensitivity of investment to q: the estimated q coefficient declines once MVK is added. This outcome is not surprising—if MVK is an important state variable, omitting it will induce omitted-variable bias. In this subsection, we investigate how firm-level heterogeneity in MVK or q amplifies or mitigates the degree of bias caused by excluding MVK.

Our subgroup test is motivated by the findings of Andrei et al. (2019), who show that the stronger investment–q relationship observed in recent decades can be attributed to the increasing within-firm volatility of q. They find that investment–q regressions for firms with higher q variance tend to exhibit higher R^2 but, puzzlingly, lower estimated q coefficients. While they attribute the rising explanatory power to firms' growing innovation activity and improved knowledge about future profitability—factors that increase q's volatility—the declining estimated slopes, i.e., the weakening effect of q on investment as q's volatility rises, remain unexplained. In this subsection, we offer an explanation for this mixed performance of q across volatility groups, emphasizing the omitted-variable bias that arises when MVK is excluded from the univariate investment–q regression.

To start, consider the classical univariate investment–q regression that omits the expected marginal value of capital, $\mathbb{E}_t[\pi_{it+1}]$:

$$\gamma_{it+1} = \beta_0 + \beta_1 q_{it} + \epsilon_{it+1}.$$

Suppose the true data-generating process for investment follows

$$\mathbb{E}_t[\gamma_{t+1}] = \vartheta + \frac{1}{\nu} q_t - \frac{1-\tau}{\nu}\, \mathbb{E}_t[\pi_{t+1}],$$

which implies that the true coefficient on q is $\frac{1}{\nu}$. As shown in the Online Appendix B.3, estimating an univariate investment-q regression without including the expected MVK yields a biased slope coefficient:

$$\hat{\beta}_1 = \frac{1}{\nu} - \frac{1-\tau}{\nu} \frac{\mathrm{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\mathrm{Var}(q_t)}. \tag{10}$$

Eq. (10) shows that the bias in the estimated q coefficient of the univariate investment-q regression depends on the sensitivity of q_t to the change of expected MVK, i.e., $\beta_{q,\pi} \equiv \frac{\mathrm{Cov}(q_t, \mathbb{E}_t[\pi_{t+1}])}{\mathrm{Var}(q_t)}$. If the MPK-$q$ sensitivity is positive (negative), i.e., q_t positively (negatively) covary with expected MVK, the univariate regression will generate a downward (upward) biased estimation of q_t coefficient.

To further examine how each component of the sensitivity term affects the investment–q regression, Eq. (10) can be rewritten as

$$\tilde{\beta}_1 = \frac{1}{\nu} - \frac{1-\tau}{\nu}\, \rho_{q,\pi}\, \frac{\sigma_\pi}{\sigma_q}, \tag{11}$$

where $\rho_{q,\pi}$ denotes the within-firm correlation between q_t and expected MVK, σ_q is the within-firm standard deviation of q_t, and σ_π is the within-firm standard deviation of $\mathbb{E}_t[\pi_{t+1}]$. For instance, if we isolate the numerator $\rho_{q,\pi}\sigma_\pi$, the variation of biased coefficient is supposed to be attribute to the change from σ_q.

Hereafter, we aim to test the patterns of the estimated coefficient from the investment–q regression by sub-sample analysis. First, we compute each firm's within-firm standard deviation of q_t and $\mathbb{E}_t[\pi_{t+1}]$, denoted σ_q and σ_π, respectively, as well as their within-firm correlation $\rho_{q,\pi}$. Next, we form four groups ranked by the MVK–q sensitivity, defined as $\frac{\mathrm{Cov}\left(q_t, \mathbb{E}_t[\pi_{t+1}]\right)}{\sigma_q^2}$. According to Eq. (10), the biased coefficient of q, $\hat{\beta}_1$, should be lower for the group with higher MVK–q sensitivity.

For the second test of heterogeneity in the estimated coefficients, we form sixteen groups by first ranking firms into four groups based on $\rho_{q,\pi}\sigma_\pi$ and then interacting these rankings with four groups based on σ_q, separately for the cases where $\rho_{q,\pi} > 0$ and $\rho_{q,\pi} < 0$. This double-sorting approach helps us isolate the effects of each component of the sensitivity term. For the sub-samples with $\rho_{q,\pi} < 0$, holding $\rho_{q,\pi}\sigma_\pi$ constant, $\tilde{\beta}_1$ is expected to decrease as σ_q increases, which is consistent with the findings of Andrei et al. (2019). Conversely, for the sub-samples with $\rho_{q,\pi} > 0$, $\tilde{\beta}_1$ is predicted to increase as σ_q rises, a pattern opposite to the results in Andrei et al. (2019) and not readily explained by firms' learning or innovation effects.

Lastly, we test the heterogeneity of the omitted-variable bias with respect to the regression R^2. The biased R^2 for the univariate regression model can be expressed as

$$R^2 = \frac{\left[1 - (1-\tau)\,\dfrac{\mathrm{Cov}\left(q_t, \mathbb{E}_t[\pi_{t+1}]\right)}{\mathrm{Var}(q_t)}\right]^2}{1 - 2(1-\tau)\,\dfrac{\mathrm{Cov}\left(q_t, \mathbb{E}_t[\pi_{t+1}]\right)}{\mathrm{Var}(q_t)} + (1-\tau)^2\,\dfrac{\mathrm{Var}\left(\mathbb{E}_t[\pi_{t+1}]\right)}{\mathrm{Var}(q_t)}}.$$

It is straightforward to see that $R^2 < 1$ whenever q_t and $\mathbb{E}_t[\pi_{t+1}]$ are not perfectly correlated (i.e., $\rho_{q,\pi} \neq 1$). On the other hand, when q is almost insensitive to changes in MVK (i.e., $|\beta_{q,\pi}| \approx 0$), R^2 reduces to a simple function of the variance ratio σ_π^2/σ_q^2:

$$R^2 \approx \frac{1}{1 + (1-\tau)^2\dfrac{\sigma_\pi^2}{\sigma_q^2}} \qquad \text{when } \rho_{q,\pi} \approx 0. \tag{12}$$

Eq. (12) implies that the regression R^2 increases monotonically with the variance ratio σ_π^2/σ_q^2 when the sensitivity of MVK to q is negligible.[10] To empirically test this prediction, we select the sub-sample of firms with absolute q–MVK sensitivity below 0.1 and then form four groups ranked by

[10] In contrast, when the correlation magnitude $|\rho_{q,\pi}|$ is non-negligible, the relationship between R^2 and σ_π^2/σ_q^2 becomes nonlinear, and no clear monotonic pattern can be expected.

the variance ratio σ_π^2/σ_q^2. We estimate the univariate investment–q regression within each group and compare their resulting R^2.

Table **6** reports the sub-sample test results, comparing the patterns of the estimated coefficients and R^2 across different groups. As shown in Panel A, for groups with higher within-firm MVK–q sensitivity, the investment–q relationship becomes weaker — reflected by a decline in the estimated q coefficient from 0.456 (group 1, with the lowest $\beta_{q,\pi}$) to 0.168 (group 4, with the highest $\beta_{q,\pi}$). These results confirm Eq. (10) and suggest that the omitted-variable bias becomes stronger when MVK and q are more strongly co-varying.

Panels B and C present the double-sorted sub-sample tests. The sixteen groups are formed by first splitting firms into four groups based on q volatility (σ_q) and intersecting them with four groups based on the product of MVK volatility and MVK–q correlation ($\rho_{q,\pi}\sigma_\pi$), separately for firms with positive correlation ($\rho_{q,\pi} > 0$, Panel B) and negative correlation ($\rho_{q,\pi} < 0$, Panel C). Importantly, when MVK and q are positively correlated and $\rho_{q,\pi}\sigma_\pi$ is high — for example, in the row corresponding to the highest $\rho_{q,\pi}\sigma_\pi$ group — the estimated q coefficient increases with greater q volatility, rising from 0.065 to 0.144. In contrast, when MVK and q are negatively correlated, the slope decreases more clearly and monotonically as q volatility increases, holding $\rho_{q,\pi}\sigma_\pi$ constant. Taken together, these findings show that the declining slope with increasing q volatility echoes the results of Andrei et al. (2019), while the opposite increasing trend in the positively correlated subsample reveals a mechanism not explained by the firms' learning or innovation activities proposed in their study.

Panel D conducts the final sub-sample test, where firms are grouped into four bins ranked by the relative importance of MVK and q volatility, measured by σ_π^2/σ_q^2, using only firms with very low MVK–q sensitivity ($|\beta_{\pi,q}| < 0.1$). Across these four bins, when the MVK–q sensitivity is negligible, the investment–q regression R^2 decreases as the relative volatility σ_π^2/σ_q^2 rises, falling from 0.276 to 0.215. This result is consistent with the prediction from Eq. (12).

Overall, our sub-group analysis provides an explanation for the observed dynamics in the investment–q relationship from the perspective of omitted-variable bias. We show that the strength of the investment–q link — reflected in both the slope and the fitted R^2 — depends not only on the volatility of q and

MVK but also on their correlation. These findings help the literature better understand the variation in the investment–q relationship within the standard intertemporal optimization framework, without introducing additional state variables or alternative mechanisms.

6. The role of cash flows and intangible capitals

In this section, we discuss two additional channels affecting firms' investment decisions that the neoclassical investment–q theory fails to fully explain. First, a large empirical literature finds that cash flow often contains substantial incremental information for investment beyond q, challenging the theoretical prediction that q alone should embed all relevant information about investment opportunities (Erickson and Whited, 2000; Gilchrist and Himmelberg, 1995). Here, we are primarily interested in examining whether cash flow can either serve as a valid proxy for the marginal value of capital (MVK) or potentially crowd out the information content provided by our machine-learning-based measure of MVK.

Second, we analyze the impact of changes in firms' capital structure on our theoretical link between investment, q, MVK, and the cost of capital. This motivation stems from empirical evidence documenting the rising importance of intangible capital in firms' balance sheets and the associated increase in R&D spending (Peters and Taylor, 2017). Our main prediction above focuses on the role of the marginal value of physical capital in driving investment in physical assets. In this section, we extend the framework to incorporate intangible capital into the firm's optimization problem and, by analogy, construct a machine-learning-based estimate of the marginal value of intangible capital to re-examine our main predictions.

6.1. Can cash flow explain the information content of MVK?

We measure cash flow as the ratio of operating profit (Compustat item *oibdp*) to capital stock (item *ppegt*), following Frank and Shen (2016), and include it as an additional explanatory variable in the investment prediction model, alongside q, $\mathbb{E}_t[\pi_{t+1}]$, and r^{COC}:

$$\gamma_{it+1} = b_0 + b_q \times q_{it} + b_1 \times CF_{it} + b_2 \times \mathbb{E}_t[\pi_{it+1}] + b_3 \times \mathbb{E}_t[r_{it+1}] + FE + \epsilon_{it}.$$

This specification allows us to test two hypotheses. First, if the cash flow measure—i.e., scaled operating profit—is a sufficient proxy for the marginal value of capital, its coefficient is expected to be negative, and our machine-learning-based MVK measure should add no incremental explanatory power. In this case, constructing a separate machine learning measure of MVK would be unnecessary. Second, if there exist other factors affecting firms' value-maximization problems that are not captured by Eq. (1)—such as external financing constraints (Fazzari et al., 1987; Chen and Chen, 2012)—then including cash flow should provide additional information about investment beyond the channels represented by MVK and q.

Table **7** presents the panel regression results. Consistent with existing findings, cash flow exhibits a positive coefficient, and its inclusion significantly enhances the model's explanatory power beyond what q alone captures. However, adding cash flow does not diminish the contribution of MVK: the coefficient of MVK remains negative and magnitude changes slightly compared to the results without cash flow (from -0.17 to -0.15), and the R^2 increases by an additional 15% when MVK is included alongside q and cash flow. Lastly, although the positive sign and statistical significance of the cost of capital remain unchanged, its incremental explanatory power for investment variation is still limited.

The results confirm that the common cash flow measure — i.e., the ratio of operating profit to capital stock — does have a positive effect on investment, but through a channel unrelated to the existing capital value–cost optimization framework. The persistent negative effect of the marginal value of capital provides additional information about investment that is not captured by cash flow. Moreover, this contrast in performance indicates that the current profit-to-capital ratio is not an appropriate proxy for the expected marginal value of capital, underscoring the importance of constructing an ex-ante measure of these unobservable marginal values.

6.2. Shifting capital structure: tangible vs. intangible capital

The role of intangible capital in the neoclassical investment–q theory has attracted increasing academic attention, especially given the rising importance of intangible assets for firms' profitability over the past two decades (Corrado and Hulten, 2010; Peters and Taylor, 2017). If the share of intangible capital is non-negligible, estimating its marginal value to profits becomes equally important for evaluating the validity of the neoclassical investment–q framework. Several key questions naturally arise: Does the marginal value of intangible capital carry the same information content as that of physical capital? Are firms' real investment decisions in intangible assets truly influenced by these marginal values? If the theoretical framework developed for physical capital remains valid, can it be sensibly extended to other forms of capital as well? We aim to address these questions in this subsection.

We re-define the firm's operating profit by allowing physical and intangible capital to enter separately as distinct state variables:[11]

$$\Pi_t = f\big(K_t^{\text{PHY}}, K_t^{\text{INT}}, N_t\big) - \Phi_t^{\text{PHY}} - \Phi_t^{\text{INT}},$$

where the two adjustment cost functions are defined as

$$\Phi_t^{\text{PHY}} = \frac{a}{2}\frac{(I_t^{\text{PHY}})^2}{K_t^{\text{PHY}} + K_t^{\text{INT}}} + I_t^{\text{PHY}}, \qquad \Phi_t^{\text{INT}} = \frac{a}{2}\frac{(I_t^{\text{INT}})^2}{K_t^{\text{PHY}} + K_t^{\text{INT}}} + I_t^{\text{INT}},$$

which is the same with Peters and Taylor (2017): when firms employ both capital types to generate profits, the total installation cost depends on the combined capital stock, $K_t^{\text{PHY}} + K_t^{\text{INT}}$.

Using the same log-linear approximation as above (see Online Appendix B.4), future total investment can be expressed as a function of the observable proxies of average q for physical and intangible capital, their expected marginal values, and the overall cost of capital. To test this prediction, we estimate the following predictive regression model:

$$\iota_{it+1} = b_0 + b_1 \times q_{it}^{\text{PHY}} + b_2 \times q_{it}^{\text{INT}} + b_3 \times \mathbb{E}_t[\pi_{it+1}^{\text{PHY}}] + b_4 \times \mathbb{E}_t[\pi_{it+1}^{\text{INT}}] + b_5 \times \mathbb{E}_t[r_{it+1}] + FE + \epsilon_{it}, \quad (13)$$

where ι_{it+1} is the sum of investments in intangible and physical capital, scaled by the total installed capital; q_{it}^{INT} is the average q for intangible capital, calculated as the ratio of firm value to the stock of intangible capital; $\mathbb{E}_t[\pi_{it+1}^{\text{PHY}}]$ and $\mathbb{E}_t[\pi_{it+1}^{\text{INT}}]$ denote the conditional expectations of the marginal values of physical and intangible capital, respectively. Under the log-linear approximation, the coefficients on both q terms are expected to be positive, while those on both MVK terms are expected to be negative; that is, $\hat{b}_1, \hat{b}_2 > 0$ and $\hat{b}_3, \hat{b}_4 < 0$.

[11] Our specification of the revenue function $f(\cdot)$ differs from that in Peters and Taylor (2017), who treat the sum of physical and intangible capital as a single input to production. In contrast, we aim to uncover the distinct information content of each capital type, which is more realistic when firms choose their investments in physical and intangible assets separately.

To estimate the MVK of intangible capital, we require data on the predetermined value of intangible capital, which we assume evolves in the same manner as physical capital, consistent with Peters and Taylor (2017):

$$\tilde{K}_{t+1}^{\text{INT}} = (1 - \delta^{\text{INT}})K_t^{\text{INT}} + I_t^{\text{INT}}.$$

The firm-year intangible capital stock K_t^{INT} is estimated following Peters and Taylor (2017) and obtained from WRDS. Investment in intangible capital I_t^{INT} is measured as R&D expenditure (Compustat item *xrd*) plus 0.3 times selling, general, and administrative expense (item *xsga*). The depreciation rate δ^{INT} is taken from the BEA's industry-specific R&D depreciation rates, and missing values are fixed at 0.15.

Analogous to the derivation of the present-state information set for physical capital, future operating profits can be approximated using the state variables $\{\tilde{K}_{t+1}^{\text{PHY}}, \tilde{K}_{t+1}^{\text{INT}}, \tilde{N}_{t+1}, \tilde{r}_{t+1}, \tilde{K}_t^{\text{PHY}}, \tilde{K}_t^{\text{INT}}, I_t^{\text{PHY}}, I_t^{\text{INT}}\}$. Where $\{\tilde{K}_{t+1}^{\text{PHY}}, \tilde{K}_{t+1}^{\text{INT}}, \tilde{N}_{t+1}\}$ governed the future production revenue generation process; while the rest of present state variables are used to approximate future installment cost process for investment on physical and intangible capital. AD is then applied to the trained neural network to compute the gradients with respect to $\tilde{K}_{t+1}^{\text{PHY}}$ and $\tilde{K}_{t+1}^{\text{INT}}$, which provide the conditional expected MVK for physical and intangible capital, respectively.

Table 8 presents the panel regression results. When using OLS estimator (Panel A). In line with our prediction, the marginal value of intangible capital is estimated to have a statistically significant negative effect (-0.215) on total investment, stronger than the negative effect (-0.084) found for the marginal value of physical capital. Also, the sign and statistical significance of the estimated coefficients for the marginal value of physical capital and the cost of capital remain unchanged.

Interestingly, the sign of the marginal value of intangible capital turns positive when we switch to the high-order cumulant estimator, in contrast to our theoretical prediction. We attribute this discrepancy relative to the OLS result to potential measurement error in the marginal value of intangible capital. As highlighted by Gonçalves et al. (2020), both intangible capital and its investment flows are often constructed under ad-hoc assumptions — such as fixed depreciation rates and fixed proportions of selling, general, and administrative expenses used to approximate intangible investment — which may lead to systematic mismeasurement. As a result, when key state variables are mis-specified, the estimated marginal value will also be biased. This hypothesis is further supported by the predicted negative effect from intangible q obtained with the high-order cumulant estimator. In comparison, the consistently estimated sign for the marginal value of physical capital across different estimators indicates that its measurement is relatively robust.

Nevertheless, the incremental information content provided by the two MVK measures beyond their corresponding q — reflected in the substantial increase of ρ^2 from 0.081 to 0.194 — remains non-negligible. As intangible capital becomes increasingly important in shaping firms' capital structures, our test suggests the potential broad applicability of the proposed framework to various forms of capital. At the same time, the mixed results highlight the growing need for more reliable data and more robust theoretical modeling of the mechanisms through which intangible capital affects firms' value-maximization processes.

7. Conclusion

In this paper, we augment the neo-classical investment–q theory by applying a log-linear approximation to the first-order conditions for firms' optimal investment decisions. This yields a new link between expected investment, Tobin's q, the expected marginal value of capital to operating profit, and the expected cost of capital. Unlike the existing corporate investment literature, which typically assumes present-period optimality and directly relates investment solely to q, our theoretical framework relies only on rational conditional expectations and preserves the forward-looking nature of the information embedded in q.

Our contribution to the practitioners lies in proposing a novel nonparametric estimation of the marginal value of capital using a machine-learning approach — specifically, neural network modeling combined with automatic differentiation to compute gradients. Without relying on ad-hoc functional assumptions (e.g., constant returns to scale) or restrictive distributional forms, we provide a flexible technique to generate ex-ante expected marginal values of capital, which can be readily applied in practice as long as the relevant state variables are observable.

Empirically, using both aggregate time-series data and firm-level panel data, we confirm the main theoretical predictions: forward investment is positively related to current q and the cost of capital but negatively related to the expected marginal value of capital. While the incremental information content from the marginal value of capital beyond q is consistently strong, we find that the role of the cost of capital in driving firms' investment decisions is generally weak. As our paper focuses primarily on the marginal value of capital channel, the diminishing link between investment and the cost channel remains an open question for future research.

References

Abel, A. B., 2018. The effects of q and cash flow on investment in the presence of measurement error. Journal of Financial Economics 128, 363–377.

Abel, A. B., Blanchard, O. J., 1983. The present value of profits and cyclical movements in investment. Tech. rep., National Bureau of Economic Research.

Abel, A. B., Panageas, S., 2022. An analytic framework for interpreting investment regressions in the presence of financial constraints. The Review of Financial Studies 35, 4055–4104.

Andrei, D., Mann, W., Moyen, N., 2019. Why did the q theory of investment start working? Journal of Financial Economics 133, 251–272.

Asker, J., Farre-Mensa, J., Ljungqvist, A., 2015. Corporate investment and stock market listing: A puzzle? The Review of Financial Studies 28, 342–390.

Bakke, T.-E., Whited, T. M., 2010. Which firms follow the market? an analysis of corporate investment decisions. The Review of Financial Studies 23, 1941–1980.

Barnett, S. A., Sakellaris, P., 1998. Nonlinear response of firm investment to q:: Testing a model of convex and non-convex adjustment costs. Journal of Monetary Economics 42, 261–288.

Bartlett, R., Partnoy, F., 2020. The misuse of tobin's q. Vand. L. Rev. 73, 353.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., Siskind, J. M., 2018. Automatic differentiation in machine learning: a survey. Journal of Marchine Learning Research 18, 1–43.

Beaudry, P., Portier, F., 2006. Stock prices, news, and economic fluctuations. American economic review 96, 1293–1307.

Belo, F., Donangelo, A., Lin, X., Luo, D., 2023. What drives firms' hiring decisions? an asset pricing perspective. The Review of Financial Studies 36, 3825–3860.

Belo, F., Lin, X., Bazdresch, S., 2014. Labor hiring, investment, and stock return predictability in the cross section. Journal of Political Economy 122, 129–177.

Belo, F., Xue, C., Zhang, L., 2013. A supply approach to valuation. The Review of Financial Studies 26, 3029–3067.

Berg, T., Fuster, A., Puri, M., 2022. Fintech lending. Annual Review of Financial Economics 14, 187–207.

Biddle, J., 2012. Retrospectives: The introduction of the cobb–douglas regression. Journal of Economic Perspectives 26, 223–236.

Bloom, N., 2009. The impact of uncertainty shocks. econometrica 77, 623–685.

Bloom, N., Bond, S., Van Reenen, J., 2007. Uncertainty and investment dynamics. The review of economic studies 74, 391–415.

Caballero, R. J., Engel, E. M., 1999. Explaining investment dynamics in us manufacturing: a generalized (s, s) approach. Econometrica 67, 783–826.

Campbell, J. Y., Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1, 195–228.

Capriotti, L., 2010. Fast greeks by algorithmic differentiation. Available at SSRN 1619626 .

Capriotti, L., Giles, M., 2024. 15 years of adjoint algorithmic differentiation (aad) in finance. Quantitative Finance pp. 1–27.

Chen, A. Y., McCoy, J., 2024. Missing values handling for machine learning portfolios. Journal of Financial Economics 155, 103815.

Chen, H. J., Chen, S. J., 2012. Investment-cash flow sensitivity cannot be a good measure of financial constraints: Evidence from the time series. Journal of Financial Economics 103, 393–410.

Cooper, R. W., Haltiwanger, J. C., 2006. On the nature of capital adjustment costs. The Review of Economic Studies 73, 611–633.

Corrado, C. A., Hulten, C. R., 2010. How do you measure a "technological revolution"? American Economic Review 100, 99–104.

Cummins, J. G., Hassett, K. A., Oliner, S. D., 2006. Investment behavior, observable expectations, and internal funds. American Economic Review 96, 796–810.

De La O, R., Myers, S., 2021. Subjective Cash Flow and Discount Rate Expectations. The Journal of Finance 76, 1339–1387.

De Silva, T., Thesmar, D., 2024. Noise in expectations: Evidence from analyst forecasts. The Review of Financial Studies 37, 1494–1537.

Delao, R., Han, X., Myers, S., 2025. The return of return dominance: Decomposing the cross-section of prices. Journal of Financial Economics 169, 104059.

Erickson, T., Whited, T. M., 2000. Measurement error and the relationship between investment and q. Journal of political economy 108, 1027–1057.

Erickson, T., Whited, T. M., 2012. Treating measurement error in tobin's q. The Review of Financial Studies 25, 1286–1329.

Fazzari, S., Hubbard, R. G., Petersen, B. C., 1987. Financing constraints and corporate investment.

Frank, M. Z., Shen, T., 2016. Investment and the weighted average cost of capital. Journal of Financial Economics 119, 300–315.

Fries, C. P., 2019. Stochastic automatic differentiation: automatic differentiation for monte-carlo simulations. Quantitative Finance 19, 1043–1059.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. Predictably unequal? the effects of machine learning on credit markets. The Journal of Finance 77, 5–47.

Gala, V. D., Gomes, J. F., Liu, T., 2020. Investment without q. Journal of Monetary Economics 116, 266–282.

Gala, V. D., Gomes, J. F., Liu, T., 2022. Marginal q. Jacobs Levy Equity Management Center for Quantitative Financial Research Paper .

Gandhi, A., Navarro, S., Rivers, D. A., 2020. On the identification of gross output production functions. Journal of Political Economy 128, 2973–3016.

Gennaioli, N., Ma, Y., Shleifer, A., 2016. Expectations and investment. NBER Macroeconomics Annual 30, 379–431.

Gilchrist, S., Himmelberg, C. P., 1995. Evidence on the role of cash flow for investment. Journal of monetary Economics 36, 541–572.

Gomes, J. F., 2001. Financing investment. American Economic Review 91, 1263–1285.

Gonçalves, A. S., Xue, C., Zhang, L., 2020. Aggregation, capital heterogeneity, and the investment capm. The Review of Financial Studies 33, 2728–2771.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., Villegas-Sanchez, C., 2017. Capital allocation and productivity in south europe. The Quarterly Journal of Economics 132, 1915–1967.

Gordon, R. J., 2004. Five puzzles in the behavior of productivity, investment, and innovation.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Hall, R. E., 2001. The stock market and capital accumulation. American Economic Review 91, 1185–1202.

Hayashi, F., 1982. Tobin's marginal q and average q: A neoclassical interpretation. Econometrica: Journal of the Econometric Society pp. 213–224.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural networks 2, 359–366.

Hou, K., van Dijk, M. A., Zhang, Y., 2012. The implied cost of capital: A new approach. Journal of Accounting and Economics 53, 504–526.

Kaniel, R., Lin, Z., Pelger, M., Van Nieuwerburgh, S., 2023. Machine-learning the skill of mutual fund managers. Journal of Financial Economics 150, 94–138.

Koijen, R. S., Van Nieuwerburgh, S., 2011. Predictability of Returns and Cash Flows. Annual Review of Financial Economics 3, 467–491.

Kothari, S. P., So, E., Verdi, R., 2016. Analysts' forecasts and asset pricing: A survey. Annual Review of Financial Economics 8, 197–219.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436–444.

Lee, C., Ng, D., Swaminathan, B., 2009. Testing international asset pricing models using implied costs of capital. The Journal of Financial and Quantitative Analysis 44, 307–335.

Lee, C. M. C., So, E. C., Wang, C. C. Y., 2020. Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects. The Review of Financial Studies 34, 1907–1951.

Lettau, M., Ludvigson, S., 2002. Time-varying risk premia and the cost of capital: An alternative implication of the q theory of investment. Journal of Monetary Economics 49, 31–66.

Li, Y., Ng, D. T., Swaminathan, B., 2013. Predicting market returns using aggregate implied cost of capital. Journal of Financial Economics 110, 419–436.

Liao, S., Nolte, I., Pawlina, G., 2020. Can capital adjustment costs explain the decline in investment–cash flow sensitivity? Journal of Financial and Quantitative Analysis pp. 1–26.

Liu, L. X., Whited, T. M., Zhang, L., 2009. Investment-based expected stock returns. Journal of Political Economy 117, 1105–1139.

Margossian, C. C., 2019. A review of automatic differentiation and its efficient implementation. Wiley interdisciplinary reviews: data mining and knowledge discovery 9, e1305.

McGrattan, E. R., Prescott, E. C., 2010. Technology capital and the us current account. American Economic Review 100, 1493–1522.

Merz, M., Yashiv, E., 2007. Labor and the market value of the firm. American Economic Review 97, 1419–1431.

Pástor, L., Sinha, M., Swaminathan, B., 2008. Estimating the intertemporal risk–return tradeoff using the implied cost of capital. The Journal of Finance 63, 2859–2897.

Peters, R. H., Taylor, L. A., 2017. Intangible capital and the investment-q relation. Journal of Financial Economics 123, 251–272.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back-propagating errors. nature 323, 533–536.

Shapiro, M. D., 1986. The dynamic demand for capital and labor. The Quarterly Journal of Economics 101, 513–542.

Sharpe, S. A., Suarez, G. A., 2021. Why isn't business investment more sensitive to interest rates? evidence from surveys. Management Science 67, 720–741.

van Binsbergen, J. H., Han, X., Lopez-Lira, A., 2022. Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases. The Review of Financial Studies 36, 2361–2396.

Woeppel, M., 2022. Using patent capital to estimate tobin's q. Journal of Financial and Quantitative Analysis 57, 2929–2967.

Figure 1: Industry distribution and MPK

The figure plots the industry distribution (the bars) and their median conditional expected marginal value of capital (the value points). Industry is classified by Fama-French ten-industry definition and exclude “Other” and “Utility”.

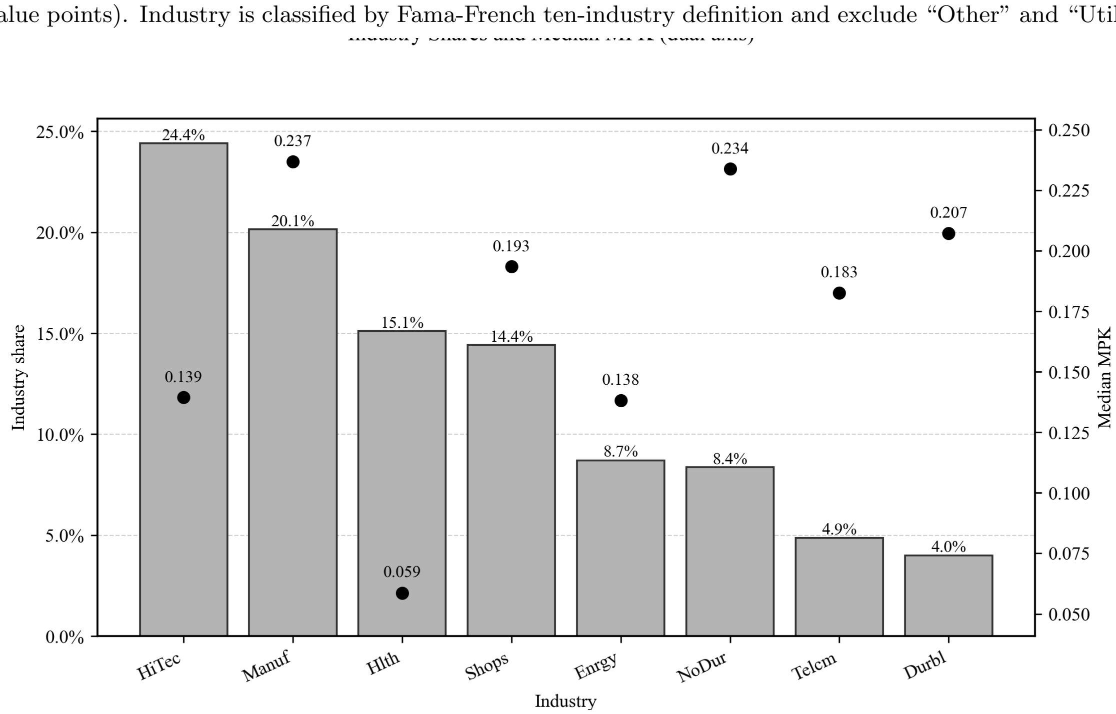


Figure 2: Time-series aggreagte investment, Tobin's Q and MVK

The figure plots the aggregate annually time series data. The black bold line plots the investment-capital ratio and gray dash line plots the macro-Q, both of them are calculated using annually data from FRED. The red dash line plots the aggregate market's conditional expected marginal value of capital, calculated by the median of S&P 500 constitutes' value covered by our machine learning estimated sample. The shaded regions on the graph correspond to NBER recession periods.

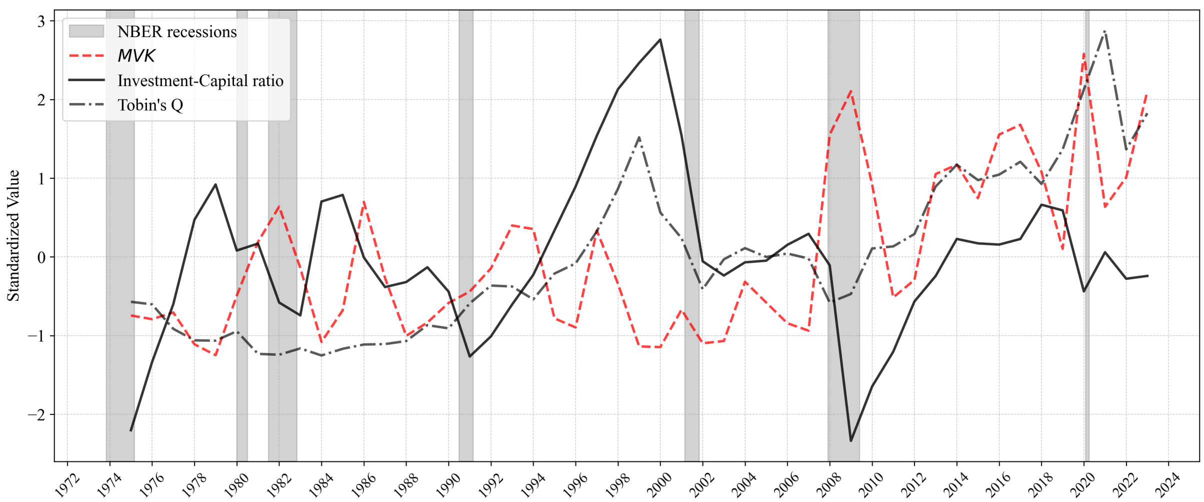


Table 1: Example of AD estimation of MVK

The table aligns forward- and reverse-mode traces that compute input-level gradients via reverse-mode automatic differentiation (AD) in a small feedforward ReLU network. The simple neural network model, consisting of two hidden layers with ReLU activations and a single linear output node, is trained using the Adam optimizer with 1000 epochs on a 10000 simulated observations from the data-generating process $F(K, N, I) = K^{0.3}N^{0.7} - I^2/K$. The left panel (**Value**) reports the numerical results of the forward pass: each row shows the value of a variable when the selected evaluation point $(K, N, I) = (400, 30, 100)$ is fed into the network. v denotes inputs, z denotes pre-activations (linear combinations of the previous layer), and a denotes post-activation values. The right panel (**Eval. point**) records the corresponding reverse-mode quantities: λ represent adjoint sensitivities that propagate gradient information backward through the network, indicating how a marginal change in each input or intermediate variable would affect the output at the evaluation point. The last rows compare the AD-based input gradients obtained from the neural network with the exact analytic partial derivatives of the underlying data-generating process.

Forward (primal) trace		**Reverse (adjoint) trace**		
Definition	**Value**	**Derivative**	**Reverse formula**	**Eval. point**
$v_{-2} = K$	400	v'_{-2}	$v'_{-2} = W_{1,K,\cdot}\, \lambda_{z1}$	**0.13**
$v_{-1} = N$	30	v'_{-1}	$v'_{-1} = W_{1,N,\cdot}\, \lambda_{z1}$	**1.49**
$v_0 = I$	100	v'_0	$v'_0 = W_{1,I,\cdot}\, \lambda_{z1}$	**-0.60**
Hidden layer 1 (linear → ReLU)				
$z_1 = W_1^\top [v_{-2}, v_{-1}, v_0] + b_1$	[-239.10, 10.30, -281.25, -63.03]	λ_{z1}	$\lambda_{z1} = \lambda_{a1} \odot \text{ReLU}'(z_1)$	[0.00, -7.97, 0.00, 0.00]
$a_1 = \text{ReLU}(z_1)$	[0.00, 10.30, 0.00, 0.00]	λ_{a1}		[-2.39, -7.97, -4.58, -1.02]
Hidden layer 2 (linear → ReLU)				
$z_2 = W_2^\top a_1 + b_2$	[-4.42, 19.20, 6.17]	λ_{z2}	$\lambda_{z2} = \lambda_{a2} \odot \text{ReLU}'(z_2)$	[0.00, -1.31, 9.75]
$a_2 = \text{ReLU}(z_2)$	[0.00, 19.20, 6.17]	λ_{a2}	$\lambda_{a1} = W_2\, \lambda_{z2}$	[0.22, -1.31, 9.75]
Output layer (linear)				
$z_{\text{out}} = W_3^\top a_2 + b_3$	39.38	$\lambda_{z,\text{out}}$	$\lambda_{z,\text{out}} = \lambda_{a,\text{out}} \odot 1$	[1]
$a_{\text{out}} = \text{linear}(z_{\text{out}})$	39.38	$\lambda_{a,\text{out}}$	$y' = 1$	[1]
$y = \mathcal{F}_\theta = a_{\text{out}}$	39.38			
AD estimation (final)		$(\partial_K \mathcal{F}_\theta,\ \partial_N \mathcal{F}_\theta,\ \partial_I \mathcal{F}_\theta) = (0.13,\ 1.49,\ -0.60)$		
Analytical solution		$(\partial_K F,\ \partial_N F,\ \partial_I F) = (0.11,\ 1.52,\ -0.50)$		

Table 2: Simulation Results

This table presents the simulation test results for MVK estimated using neural network plus auto-differentiation approach. Panel A summarizes the form of data generation processes. The profit function Π is generated by three different forms of production function: linear, Cobb-Douglas, and constant elasticity of substitution (CES), minus a quadratic adjustment cost function $(\alpha/2 \cdot (I^2/K))$. Π is generated with setting parameter values and Capital (K), investment (I), and labor (L) are drawn from log-normal distributions: $\log K \sim \mathcal{N}(\log(1000),\ 0.5^2)$, $\log I \sim \mathcal{N}(\log(200),\ 0.1^2)$, and $\log L \sim \mathcal{N}(\log(500),\ 0.25^2)$. Panel B, C reports the summary statistics of actual and two estimated MVK (by central differentiation (CD) and auto-differentiation (AD), as well as the evaluation of accuracy by comparing the estimated value with actual value solved analytically from the data-generating process, using 2,500 and 10,000 simulated data points, respectively. MAE (MAPE) denotes the mean absolute (percentage) error between estimated and true MVK. MAE^{OOS} (MAPE^{OOS}) evaluate the accuracy using an out-of-sample approach, which using 80% data to fit the model and the rest of 20% used to estimation and evaluation.

Panel A: Production Function Specifications

Function	**Specification**	**Parameters**
Linear	$\Pi = a_1 K + b_1 L - \alpha/2 \cdot (I^2/K)$	$\alpha = 10$; $a_1 = 0.4$; $b_1 = 1.2$
Cobb–Douglas	$\Pi = \lambda_1 L^{\lambda_2} K^{1-\lambda_2} - \alpha/2 \cdot (I^2/K)$	$\alpha = 10$; $\lambda_1 = 1.5$; $\lambda_2 = 0.28$
CES	$\Pi = c_1 \left[d_1 K^{1-\frac{1}{\nu}} + (1-d_1) L^{1-\frac{1}{\nu}} \right]^{\frac{\theta}{1-\frac{1}{\nu}}} - \alpha/2 \cdot (I^2/K)$	$\alpha = 10$; $c_1 = 3$; $d_1 = 0.36$; $\nu = 0.5$; $\theta = 0.85$

Panel B: 2500 Data Points

	Linear			Cobb-Douglas			CES		
	Actual	AD	CD	Actual	AD	CD	Actual	AD	CD
Mean	0.676	0.650	0.690	0.361	0.345	0.362	0.292	0.277	0.285
Std	0.360	0.262	0.432	0.387	0.314	0.466	0.376	0.305	0.455
Median	0.561	0.561	0.610	0.240	0.233	0.343	0.172	0.174	0.248
MAE		0.037	0.299		0.029	0.286		0.022	0.267
MAPE		2.3%	51.3%		4.5%	176.1%		3.5%	328.2%
MAE^{OOS}		0.053	0.319		0.040	0.282		0.035	0.307
MAPE^{OOS}		2.8%	52.7%		4.7%	162.4%		4.2%	341.7%

Panel C: 10000 Data Points

	Linear			Cobb-Douglas			CES		
	Actual	AD	CD	Actual	AD	CD	Actual	AD	CD
Mean	0.681	0.676	0.713	0.367	0.365	0.367	0.297	0.291	0.298
Std	0.394	0.353	0.506	0.420	0.406	0.552	0.409	0.362	0.431
Median	0.559	0.561	0.610	0.240	0.239	0.305	0.171	0.171	0.229
MAE		0.016	0.290		0.012	0.277		0.012	0.175
MAPE		1.3%	51.5%		3.0%	181.0%		2.5%	212.8%
MAE^{OOS}		0.011	0.289		0.011	0.282		0.010	0.287
MAPE^{OOS}		1.1%	52.3%		2.9%	185.6%		4.5%	390.9%

Table 3: Summary Statistics for Aggregate Market and Firm-Level Panel

This table presents summary statistics for the variables used in investment predictive regression, for time-series aggregate data (Panel A) and for firm-year panel data (Panel B). In Panel A, the aggregate market's I/K and Macro q are constructed following Andrei et al. (2019), using annual data from FRED, and the detailed definitions are shown in Online Appendix Table A1. $\mathbb{E}_t[\Pi'_K]$ donates the aggregate market's conditional expected marginal value of capital, calculated by the median of S&P 500 constitutes' value covered by our machine learning estimated sample. r^{LS} represents the economists' expectation of one-year-ahead S&P 500 return, calculated as the ratio of one-year-ahead median expected value for S&P 500 index to the present year's realized index value, using data from Livingston Survey by Federal Reserve Bank of Philadelphia. r^{CAPM} and $r^{\text{FF-3}}$ are two measures of WACC calculated using cost of equity estimated from CAPM and Fama-French-three-factor model, respectively. r^{ICC} is implied cost of capital estimated from finite Gordon dividend discount model, using analyst forecast of earnings per share data. For firm-level variables, I/K is defined as one-year-ahead value of the investment expenditure (Compustat item *capx*) scaled by the physical capital stock (item *ppegt*). Tobin's q is defined as firm's market value (item *prcc_f* times *csho*), plus the book value of debt (item *dlcc* + *dlt*, set to be zero if missing), minus the current asset value (item *act*), scaled by capital stock (item *ppegt*). The sample spans from 1975 to 2023, covers 49 years, with exception for r^{ICC}, spans from 1976 to 2023, covers 48 years.

Panel A. Aggregate market

	Mean	**Std.**	**25th**	**50th**	**75th**
I/K	0.112	0.009	0.108	0.112	0.115
q^{macro}	1.369	0.513	0.908	1.327	1.809
$\mathbb{E}_t[\Pi'_K]$	0.295	0.166	0.164	0.243	0.410
r^{LS}	0.090	0.063	0.051	0.079	0.104
r^{CAPM}	0.110	0.027	0.088	0.104	0.131
$r^{\text{FF-3}}$	0.116	0.028	0.091	0.120	0.135
r^{ICC}	0.082	0.010	0.076	0.079	0.087

Panel B. Firm-year panel data

	Mean	**Std.**	**25th**	**50th**	**75th**
I/K	0.127	0.119	0.059	0.098	0.160
Tobin's q	5.160	23.470	0.409	1.134	3.514
$\mathbb{E}_t[\Pi'_K]$	0.217	0.276	0.080	0.199	0.369
r^{CAPM}	0.113	0.037	0.088	0.110	0.134
$r^{\text{FF-3}}$	0.124	0.049	0.092	0.118	0.150
r^{ICC}	0.122	0.056	0.080	0.107	0.145

Table 4: Time-series regression results

This table presents the time-series regression results using the aggregate market variables, with the following model:

$$\gamma_{t+1} = b_0 + b_1 \times X_t + \epsilon_t,$$

where γ_{t+1} is the log forward investment-capital ratio. The independent variable vector X_t includes univariate log q^{macro} for column (1), log q^{macro} and log $\mathbb{E}_t[\pi_{t+1}]$ for column (2), and adding one of the four aggregate cost of capital measure: r^{LS}, r^{CAPM}, $r^{\text{FF}-3}$, r^{ICC}, for column (3), (4), (5), (6), respectively. All cost of capital measures are added one and taken logarithms before the estimation. The forward investment-capital ratio and q^{macro} are constructed following Andrei et al. (2019), using annually data from FRED, and the detailed definitions are shown in Online Appendix Table A1. $\mathbb{E}_t[\Pi'_K]$ donates the aggregate market's conditional expected marginal value of capital, calculated by the median of S&P 500 constitutes' value covered by our machine learning estimated sample. r^{LS} represents the economists' expectation of one-year-ahead S&P 500 return, calculated as the ratio of one-year-ahead median expected value for S&P 500 index to the present year's realized index value, using data from Livingston Survey by Federal Reserve Bank of Philadelphia. r^{CAPM} and $r^{\text{FF-3}}$ are two measures of WACC calculated using cost of equity estimated from CAPM and Fama-French-three-factor model, respectively. r^{ICC} is implied cost of capital estimated from finite Gordon dividend discount model, using analyst forecast of earnings per share data. Standard errors are reported in in parentheses. The test (1) - (5) spans from 1975 to 2023, covers 49 years; the test for (6), spans from 1976 to 2023, covers 48 years. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
q^{macro}	0.053*	0.099***	0.087***	0.213***	0.182***	0.119***
	(0.032)	(0.032)	(0.031)	(0.060)	(0.041)	(0.033)
$\mathbb{E}_t[\pi_{t+1}]$		-0.071***	-0.068***	-0.071***	-0.057***	-0.076***
		(0.020)	(0.020)	(0.019)	(0.019)	(0.018)
r^{LS}			-0.343*			
			(0.184)			
r^{CAPM}				1.996**		
				(0.905)		
$r^{\text{FF}-3}$					1.802***	
					(0.620)	
r^{ICC}						1.857
						(1.265)
Obs.	49	49	49	49	49	48
R^2	0.056	0.253	0.306	0.326	0.371	0.336

Table 5: Panel regression results

This table reports panel regression estimates of the investmemt predictive regression, using the following model:

$$\gamma_{it+1} = b_0 + b_1 \times X_{it} + FE + \epsilon_{it},$$

where γ_{it+1} is firm's one-year-ahead log investment-capital ratio, calculated as investment expenditure (Compustat item *capx*) scaled by the physical capital stock (item *ppegt*). The independent variable vector X_t includes univariate log present Tobin's q (firm's market value (item *prcc_f* times *csho*), plus the book value of debt (item *dlcc* + *dlt*, set to be zero if missing), minus the current asset value (item *act*), scaled by capital stock (item *ppegt*)), log present Tobin's q plus conditional expected marginal value of capital, and plus one of the three measures of cost of capital: r^{CAPM}, $r^{\text{FF}-3}$, r^{ICC}. r^{CAPM} and $r^{\text{FF-3}}$ are two measures of WACC calculated using cost of equity estimated from CAPM and Fama-French-three-factor model, respectively. r^{ICC} is implied cost of capital estimated from finite Gordon dividend discount model, using analyst forecast of earnings per share data. The detailed calculation of cost of capital measures is shown in Online Appendix. FE donates both firm- and year-level fixed effects. Panel A shows OLS fixed-effect estimates including estimated coefficients and with-in-firm R-square. Panel B reports measurement error-correction model using high-order cumulant GMM estimates with associated ρ^2, which represents the explanatory power from the overall independent variables, Sargan-Hansen J-statistic and p-values testing over-identifying restrictions. Standard errors (in parentheses) are clustered at both firm and year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. OLS estimator

	q only	with MVK	with COC		
			CAPM	FF-3	ICC
q	0.262***	0.239***	0.235***	0.238***	0.239***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
MVK		-0.172***	-0.173***	-0.173***	-0.172***
		(0.004)	(0.004)	(0.004)	(0.004)
COC			0.962***	0.420***	-0.225***
			(0.106)	(0.062)	(0.073)
Fixed effects	YES	YES	YES	YES	YES
N	110977	110977	110977	110977	110977
R-sq	0.250	0.283	0.284	0.283	0.283

Panel B. High-order cumulant estimator

	q only	with MVK	with COC		
			CAPM	FF-3	ICC
q	0.659***	0.651***	0.645***	0.642***	0.626***
	(0.028)	(0.028)	(0.029)	(0.028)	(0.027)
MVK		-0.207***	-0.182***	-0.191***	-0.191***
		(0.005)	(0.005)	(0.005)	(0.005)
COC			5.298***	3.748***	3.478***
			(0.147)	(0.116)	(0.168)
Fixed effects	YES	YES	YES	YES	YES
N	110977	110977	110977	110977	110977
ρ^2	0.267	0.336	0.366	0.356	0.331
J-stat	7.491	0.768	7.836	2.661	0.170
p-val	0.024	0.681	0.020	0.264	0.918

Table 6: Univariate and Double-Sorted Portfolio Analysis

This table reports panel regression estimates of the investment predictive regression using only q as independent variable, using samples constructed by different within-firm attributes:

$$\gamma_{it+1} = b_0 + b_1 \times q_{it} + FE + \epsilon_{it},$$

where γ_{it+1} is firm's one-year-ahead log investment-capital ratio, calculated as investment expenditure (Compustat item *capx*) scaled by the physical capital stock (item *ppegt*). q donate as the log present Tobin's q (firm's market value (item *prcc_f* times *csho*), plus the book value of debt (item *dlcc* + *dlt*, set to be zero if missing), minus the current asset value (item *act*), scaled by capital stock (item *ppegt*)). FE donates both firm- and year-level fixed effects. Panel A shows the regression results for four groups of firms sorted by the within-firms' MVK-q sensitivity, $\beta_{\pi,q}$, defined as $\frac{\mathrm{Cov}\left(q_t, \mathbb{E}_t[\pi_{t+1}]\right)}{\mathrm{Var}(q_t)}$. Panel B (C) presents the estimated coefficients of q for 16 groups intersected by four groups sorted by $\rho_{q,\pi}\sigma_\pi$ and four groups sorted by σ_q, for the firms with $\rho_{q,\pi} > 0$ ($\rho_{q,\pi} < 0$). Standard errors (in parentheses) are clustered at both firm and year level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Four Groups Univariate Sorted by $\beta_{\pi,q}$

Group by $\beta_{\pi,q}$	1	2	3	4
$\hat{b_1}$	0.456***	0.290***	0.208***	0.168***
	(0.005)	(0.004)	(0.003)	(0.005)

Panel B: 4×4 Groups Double Sorted by σ_q and σ_π: $\rho_{q,\pi} > 0$

$\hat{b_1}$	Group by σ_q			
Group by σ_π	1	2	3	4
1	0.289***	0.238***	0.194***	0.202***
	(0.026)	(0.018)	(0.013)	(0.009)
2	0.249***	0.252***	0.197***	0.167***
	(0.027)	(0.018)	(0.014)	(0.009)
3	0.182***	0.223***	0.183***	0.155***
	(0.036)	(0.017)	(0.013)	(0.009)
4	0.065**	0.122***	0.135***	0.144***
	(0.031)	(0.017)	(0.015)	(0.010)

Panel C: 4×4 Groups Double Sorted by σ_q and σ_π: $\rho_{q,\pi} < 0$

$\hat{b_1}$	Group by σ_q			
Group by σ_π	1	2	3	4
1	0.706***	0.509***	0.436***	0.338***
	(0.027)	(0.015)	(0.011)	(0.008)
2	0.539***	0.429***	0.336***	0.263***
	(0.020)	(0.015)	(0.010)	(0.008)
3	0.438***	0.352***	0.286***	0.221***
	(0.020)	(0.013)	(0.011)	(0.007)
4	0.419***	0.300***	0.267***	0.209***
	(0.021)	(0.014)	(0.011)	(0.007)

Panel D: Four Groups Univariate Sorted by σ_π^2/σ_q^2 when $\beta_{\pi,q} = 0.1$

Group by σ_π^2/σ_q^2	1	2	3	4
R^2	0.276	0.237	0.208	0.215

Table 7: Panel regression test with cash flows

This table reports panel regressions of the one-year-ahead investment–capital ratio

$$\gamma_{it+1} = b_0 + b_1 X_{it} + FE + \epsilon_{it},$$

where γ_{it+1} is the log of total investment divided by capital stock. Total investment equals R&D expenditure (*xrd*) plus 0.3× selling, general, and administrative expenses (*xsga*) plus capital expenditures (*capx*), scaled by physical capital stock (*ppegt*). The first specification includes only log Tobin's q. The second adds firm cash flow, defined as log operating income before depreciation (*oibdp*) scaled by capital stock. The third augments the model with the machine-learning estimate of the marginal value of capital (MVK). The fourth further adds one of three cost-of-capital measures: r^{CAPM}, $r^{\text{FF-3}}$, or r^{ICC}. Panel A reports fixed-effects OLS estimates, including coefficients, two-way clustered standard errors (by firm and year), and within-firm R^2. Panel B reports measurement-error-corrected estimates from the high-order cumulant GMM estimator, presenting ρ^2 (variance explained by regressors), the Sargan–Hansen J-statistic, and p-values for over-identifying restrictions. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. OLS estimator

	q only	q + CF	with MVK	with COC		
				CAPM	FF-3	ICC
q	0.228***	0.167***	0.149***	0.143***	0.148***	0.149***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
CF		0.185***	0.172***	0.171***	0.171***	0.172***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
MVK			-0.156***	-0.157***	-0.157***	-0.157***
			(0.003)	(0.003)	(0.003)	(0.003)
COC				1.245***	0.376***	-0.329***
				(0.103)	(0.060)	(0.070)
Fixed effects	YES	YES	YES	YES	YES	YES
N	92706	92706	92706	92706	92706	92706
R-sq	0.255	0.286	0.319	0.321	0.319	0.319

Panel B. High-order cumulant estimator

	q only	q + CF	with MVK	with COC		
				CAPM	FF-3	ICC
q	0.566***	0.579***	0.522***	0.499***	0.513***	0.504***
	(0.030)	(0.033)	(0.032)	(0.032)	(0.031)	(0.031)
CF		-0.088***	-0.083***	-0.074***	-0.077***	-0.060***
		(0.022)	(0.021)	(0.021)	(0.020)	(0.020)
MVK			-0.227***	-0.198***	-0.209***	-0.210***
			(0.004)	(0.004)	(0.004)	(0.004)
COC				5.338***	3.542***	2.856***
				(0.154)	(0.126)	(0.171)
Fixed effects	YES	YES	YES	YES	YES	YES
N	92706	92706	92706	92706	92706	92706
ρ^2	0.208	0.183	0.258	0.296	0.285	0.262
J-stat	6.495	48.438	38.918	25.378	34.792	38.538
p-val	0.039	0.000	0.000	0.000	0.000	0.000

Table 8: Panel regression test including intangible capital

This table reports panel regressions of the one-year-ahead total investment–capital ratio:

$$\gamma_{it+1} = b_0 + b_1 X_{it} + FE + \epsilon_{it},$$

where total investment equals R&D expenditure (*xrd*) plus 0.3× SG&A (*xsga*) plus capital expenditures (*capx*), scaled by the sum of physical capital stock (*ppegt*) and estimated intangible stock. The first specification uses two Tobin's q measures: q^{PHY} (firm value over physical capital) and q^{INT} (firm value over intangible capital). The second adds the machine-learning estimates of the marginal value of physical and intangible capital (MVK^{PHY}, MVK^{INT}). The third further includes one of three cost-of-capital measures: r^{CAPM}, $r^{\text{FF}-3}$, or r^{ICC}. Panel A reports fixed-effects OLS estimates with within-firm R^2. Panel B reports measurement-error-corrected estimates using the high-order cumulant GMM estimator, including ρ^2 (variance explained by regressors), the Sargan–Hansen J-statistic, and p-values for over-identifying restrictions. Standard errors (in parentheses) are clustered by firm and year. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. OLS estimator

	q only	with MVK	with COC		
			CAPM	FF-3	ICC
q^{PHY}	0.082***	0.092***	0.091***	0.092***	0.092***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
q^{INT}	0.052***	0.037***	0.036***	0.037***	0.036***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
MVK^{PHY}		-0.084***	-0.084***	-0.084***	-0.084***
		(0.004)	(0.004)	(0.004)	(0.004)
MVK^{INT}		-0.215***	-0.213***	-0.214***	-0.215***
		(0.009)	(0.009)	(0.009)	(0.009)
COC			0.729***	0.269***	-0.206***
			(0.080)	(0.043)	(0.056)
Fixed effects	YES	YES	YES	YES	YES
R-sq	0.243	0.257	0.258	0.258	0.257
N	126053	126053	126053	126053	126053

Panel B. High-order cumulant estimator

	q only	with MVK	with COC		
			CAPM	FF-3	ICC
q^{PHY}	0.563***	0.559***	0.561***	0.558***	0.562***
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
q^{INT}	-0.332***	-0.299***	-0.305***	-0.300***	-0.300***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
MVK^{PHY}		-0.126***	-0.124***	-0.125***	-0.126***
		(0.005)	(0.005)	(0.005)	(0.005)
MVK^{INT}		0.422***	0.382***	0.392***	0.408***
		(0.008)	(0.008)	(0.008)	(0.008)
COC			1.845***	1.092***	0.753***
			(0.096)	(0.053)	(0.070)
Fixed effects	YES	YES	YES	YES	YES
ρ^2	0.081	0.194	0.205	0.201	0.196
J-stat	47.356	71.049	70.355	70.426	70.212
p-val	0.000	0.000	0.000	0.000	0.000
N	126053	126053	126053	126053	126053

Online Appendix for

"Predicting Investment with a Machine Learning Estimate of Expected Marginal Value of Capital"

A. Data for aggregate market test

We define the relevant aggregate investment and Tobin's q using in our time-series test following Andrei et al. (2019). Since their variables are constructed at quarterly frequency, we find all the corresponding variables at annually frequency from FRED to construct annual measures. We modify the codes from William Mann's website to generate the measures.

Investment is defined as the private nonresidential fixed investment (FRRD item *PNFI*), divided by the net capital stock (*K1NTOTL1ES000*). Both of them are adjusted by the implicit price deflator (*A008RD3A086NBEA*).

The numerator of Tobin's q is the aggregate market value of enterprise: equity (*NCBEILA027N*) plus liabilities (*BOGZ1FL104190005A*), minus financial assets (*NCBTFTA027N*) and inventories (*A015RC1A027NBEA*), add the market value of outstanding bonds, and minus the book value of outstanding bonds. The book value of outstanding bonds is the sum of the outstanding amounts of taxable corporate bonds (*NCBCBIA027N*) and tax-exempt corporate bonds (*MSLBSNNCB*). The market value of bonds is computed following the algorithm in Hall (2001). Each year, all newly issued taxable and tax–exempt bonds are assumed to be ten-year, annual-coupon, bullet securities. The coupon rate is set equal to the yield of that year — the Moody's BAA yield (*BAA*) for taxable bonds, and the long-term municipal bond yield (*WSLB20*) for tax-exempt bonds. After 2017, the municipal yield is proxied by 0.7×BAA. For each subsequent year, the remaining coupon and principal payments of all outstanding vintages are discounted at the current yield to obtain their market value, ensuring that market and book values diverge after issuance. The book value of bonds equals the sum of taxable (*NCBCBIA027N*) and tax-exempt (*MSLBSNNCB*) outstanding amounts.

The denominator of Tobin's q is the replacement cost of corporate capital. Following Hall (2001), this series is initialized in 1954 using the real net stock of private nonresidential fixed assets (`K1NTOTL1ES000`), deflated by the same investment deflator as applied to the numerator. Each subsequent year, the real replacement cost is updated by adding real gross corporate fixed investment (`BOGZ1FA105013005A`) and depreciating the existing stock at an annual rate of 10%. Specifically, the replacement cost in year t equals the undepreciated portion of last year's capital plus that year's real investment. Finally, the resulting real replacement cost is re-inflated by the NIPA deflator to ensure consistency with the nominal valuation of the numerator in Tobin's q.

B. Complementary proof

B.1. Log-linear approximation of the capital valuation condition

To analyze the informational content of the shadow value of capital Q_t, we consider the first-order condition with respect to K_{t+1}, which equates the marginal cost and the expected marginal benefit of installed capital. Specifically, from the firm's dynamic optimization problem, we obtain:

$$Q_t = \mathbb{E}_t \left[\beta_{t+1} \left(\Pi'_{K,t+1} + (1-\delta) Q_{t+1} \right) \middle| \Omega_t \right], \tag{A1}$$

where $\Pi'_{K,t+1} \equiv \partial \Pi_{t+1} / \partial K_{t+1}$ denotes the marginal operating profit with respect to capital, and δ is the depreciation rate.

We follow standard convention and express all variables in natural logarithmic form. Let

$$q_t \equiv \log Q_t, \qquad r_{t+1} \equiv \log(1 + R_{t+1}), \qquad \pi_{t+1} \equiv \log \Pi'_{K,t+1},$$

where R_{t+1} denotes the gross return on capital such that the stochastic discount factor (SDF) is $\beta_{t+1} = 1/(1+R_{t+1}) = \exp(-r_{t+1})$. The logarithmic steady-state values are denoted $\bar{q} = \log \bar{Q}$, $\bar{\pi} = \log \bar{\Pi}'_K$, and $\bar{r} = \log(1+\bar{R})$, with corresponding steady-state identity:

$$\bar{Q} = \bar{\beta} \left(\bar{\Pi}'_K + (1-\delta)\bar{Q} \right), \qquad \text{with} \quad \bar{\beta} = \exp(-\bar{r}).$$

Dividing both sides of (A1) by the steady-state value $\bar{Q} = \bar{\beta}(\bar{\Pi}'_K + (1-\delta)\bar{Q})$ and applying the definitions above yields:

$$\frac{Q_t}{\bar{Q}} = \mathbb{E}_t \left[\exp(-r_{t+1}) \left(\frac{\Pi'_{K,t+1}}{\bar{\Pi}'_K} \cdot \frac{\bar{\Pi}'_K}{\bar{Q}} + \frac{(1-\delta)Q_{t+1}}{\bar{Q}} \right) \right].$$

To simplify notation, define the relative weights (based on steady-state values):

$$\tau \equiv \frac{(1-\delta)\bar{Q}}{\bar{\Pi}'_K + (1-\delta)\bar{Q}} \in (0,1), \qquad 1 - \tau = \frac{\bar{\Pi}'_K}{\bar{\Pi}'_K + (1-\delta)\bar{Q}}.$$

Using these weights, and expressing each term in log-deviation form around the steady state—namely, $\hat{q}_t \equiv q_t - \bar{q}$, $\hat{\pi}_{t+1} \equiv \pi_{t+1} - \bar{\pi}$, and $\hat{r}_{t+1} \equiv r_{t+1} - \bar{r}$—we obtain:

$$\hat{q}_t = \log \mathbb{E}_t \left[\exp\left(-\hat{r}_{t+1}\right) \left((1-\tau) \exp(\hat{\pi}_{t+1}) + \tau \exp(\hat{q}_{t+1}) \right) \right]. \tag{A2}$$

To log-linearize (A2), we proceed in two steps. First, apply a second-order log-normal approximation to the outer expectation using the identity $\log \mathbb{E}[e^x] \approx \mathbb{E}[x] + \frac{1}{2}\text{Var}(x)$ (neglecting higher-order cumulants). Second, apply a first-order Taylor expansion to the inner convex combination, which gives:

$$\log\left((1-\tau)\exp(\hat{\pi}_{t+1}) + \tau\exp(\hat{q}_{t+1})\right) \approx (1-\tau)\hat{\pi}_{t+1} + \tau\hat{q}_{t+1}.$$

Combining these approximations, we obtain:

$$\hat{q}_t \approx -\mathbb{E}_t[\hat{r}_{t+1}] + (1-\tau)\mathbb{E}_t[\hat{\pi}_{t+1}] + \tau\mathbb{E}_t[\hat{q}_{t+1}] + \xi_t, \tag{A3}$$

where ξ_t collects second-order approximation errors involving conditional variances and covariances, specifically:

$$\begin{aligned}\xi_t = \; & -\tfrac{1}{2}\text{Var}_t(\hat{r}_{t+1}) + \tfrac{1}{2}(1-\tau)^2\,\text{Var}_t(\hat{\pi}_{t+1}) \\ & + \tfrac{1}{2}\tau^2\,\text{Var}_t(\hat{q}_{t+1}) - \tau(1-\tau)\,\text{Cov}_t(\hat{q}_{t+1}, \hat{\pi}_{t+1}) \\ & - \text{Cov}_t\left(\hat{r}_{t+1},\, \tau\hat{q}_{t+1} + (1-\tau)\hat{\pi}_{t+1}\right).\end{aligned}$$

Finally, reverting from log-deviations to level variables, we arrive at the log-linearized pricing equation:

$$q_t \approx \kappa + \tau\mathbb{E}_t[q_{t+1}] - \mathbb{E}_t[r_{t+1}] + (1-\tau)\mathbb{E}_t[\pi_{t+1}] + \xi_t, \tag{A4}$$

where the constant term is given by:

$$\kappa \equiv (1-\tau)\bar{q} + \bar{r} - (1-\tau)\bar{\pi}.$$

Equation (A4) highlights the decomposition of the current shadow value of capital q_t into its forward-looking components. The term $\mathbb{E}_t[q_{t+1}]$ captures the information content of expected forward q_{t+1}; $\mathbb{E}_t[r_{t+1}]$ reflects the intertemporal cost of capital; and $\mathbb{E}_t[\pi_{t+1}]$ represents the expected marginal value of capital. The weights τ and $1-\tau$ reflect their respective importance in determining the present valuation of q at the margin, while ξ_t summarizes second-order uncertainty effects. This decomposition parallels the logic in Lettau and Ludvigson (2002) and subsequent work on the predictive structure of value ratios in dynamic investment models.

B.2. Proof of Eq (9)

To establish the equivalence between estimating the expected marginal product of capital (MVK) and estimating the partial derivative of the conditional expectation of future profits with respect to capital, we begin by defining the MVK. For firm i at time $t+1$, we let the marginal product of capital be denoted as

$$MVK_{t+1} = \frac{\partial \Pi_{it+1}}{\partial K_{it+1}},$$

where Π_{it+1} represents profits, and K_{it+1} denotes capital, which is deterministic under the information set Ω_t. Given this, we aim to show that the conditional expectation of MVK, $E_t[MVK_{t+1}|\Omega_t]$, is equivalent to the partial derivative of the conditional expectation of profits with respect to capital, that is:

$$E_t[MVK_{t+1}|\Omega_t] = E_t\left[\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}\middle|\Omega_t\right] = \frac{\partial E_t[\Pi_{it+1}|\Omega_t]}{\partial K_{it+1}}.$$

To justify that we can interchange differentiation and expectation in this context, we leverage the Leibniz integral rule for differentiation under the expectation. The general principle is that if we have a function $g(x, \eta)$ that is differentiable with respect to η and satisfies certain regularity conditions, then:

$$\frac{d}{d\eta}\mathbb{E}[g(X,\eta)] = \mathbb{E}\left[\frac{\partial g(X,\eta)}{\partial \eta}\right].$$

In our case, we consider the conditional expectation $E_t[\Pi_{it+1}|\Omega_t]$ of the profit function Π_{it+1}, which is differentiable with respect to K_{it+1}. To apply the Leibniz rule, we must ensure that the partial derivative $\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}$ exists and satisfies specific regularity conditions to allow the differentiation to be brought inside the expectation.

We make the following key assumptions to ensure this operation is valid. First, we assume that Π_{it+1} is continuously differentiable with respect to K_{it+1}, ensuring the existence and continuity of $\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}$. Second, we assume that $\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}$ is dominated by an integrable function. Specifically, there exists an integrable function g such that $\left|\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}\right| \leq g$ and $E_t[g|\Omega_t] < \infty$. This condition enables us to apply the Dominated Convergence Theorem (DCT), which ensures that we can move the differentiation operator inside the expectation safely.

Since K_{it+1} is deterministic under Ω_t, it behaves as a constant within the conditional expectation, further validating our application of the Leibniz rule. Applying the DCT and the Leibniz rule, we obtain:

$$\frac{\partial}{\partial K_{it+1}} E_t[\Pi_{it+1}|\Omega_t] = E_t\left[\frac{\partial \Pi_{it+1}}{\partial K_{it+1}}\middle|\Omega_t\right].$$

Thus, we have shown that the expected marginal product of capital, conditional on Ω_t, is equivalent to the partial derivative of the conditional expected profits with respect to capital:

$$E_t[MVK_{t+1}|\Omega_t] = \frac{\partial E_t[\Pi_{it+1}|\Omega_t]}{\partial K_{it+1}}.$$

This completes the proof, rigorously supporting the interchange of differentiation and expectation through the application of the Dominated Convergence Theorem.

B.3. Proof for omitted variable bias

We show that omitting the conditional expected marginal value of capital (MVK) from the investment regression leads to a biased slope coefficient on Tobin's q. Consider the following population model, consistent with the theoretical investment Euler equation:

$$\gamma_{t+1} = \vartheta + \frac{1}{\nu} q_t - \frac{1-\tau}{\nu} \mathbb{E}_t[\pi_{t+1}] + u_{t+1}, \tag{A5}$$

where γ_{t+1} denotes the log investment–capital ratio, q_t is (log) Tobin's q, and $\mathbb{E}_t[\pi_{t+1}]$ is the conditional expectation of the marginal value of capital (MVK). We assume the usual orthogonality condition $\mathbb{E}[u_{t+1} \mid q_t, \mathbb{E}_t[\pi_{t+1}]] = 0$.

Researchers often estimate the following misspecified univariate regression:

$$\gamma_{t+1} = \beta_0 + \beta_1 q_t + \varepsilon_{t+1}. \tag{A6}$$

Let $z_t := \mathbb{E}_t[\pi_{t+1}]$. Rewrite (A5) as

$$\gamma_{t+1} = \vartheta + b q_t + c z_t + u_{t+1}, \quad \text{where} \quad b = \tfrac{1}{\nu}, \quad c = -\tfrac{1-\tau}{\nu}.$$

The population OLS estimator in (A6) is

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(q_t, \gamma_{t+1})}{\mathrm{Var}(q_t)}. \tag{A7}$$

Substituting the true model (A5) into (A7) gives

$$\hat{\beta}_1 = b + c\frac{\operatorname{Cov}(q_t, z_t)}{\operatorname{Var}(q_t)} + \frac{\operatorname{Cov}(q_t, u_{t+1})}{\operatorname{Var}(q_t)}.$$

By the exogeneity condition $\mathbb{E}[u_{t+1} \mid q_t, z_t] = 0$, the last term vanishes. Thus,

$$\hat{\beta}_1 = b + c\,\frac{\operatorname{Cov}(q_t, z_t)}{\operatorname{Var}(q_t)} = \frac{1}{\nu} - \frac{1-\tau}{\nu}\,\frac{\operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)}, \tag{A8}$$

which is exactly Equation (10) in the main text.

The second term in (A8) is the classical omitted variable bias: it equals the true coefficient on the excluded regressor, $c = -\frac{1-\tau}{\nu}$, multiplied by the slope from regressing the omitted variable z_t on q_t, namely $\operatorname{Cov}(q_t, z_t)/\operatorname{Var}(q_t)$. If q_t and $\mathbb{E}_t[\pi_{t+1}]$ are positively correlated, the bias term is negative because $c < 0$, and the univariate estimate $\hat{\beta}_1$ will understate the true structural coefficient $1/\nu$. Unbiasedness requires either zero correlation between q_t and the expected MVK or a degenerate case with $1 - \tau = 0$.

Using the correlation coefficient $\rho_{q,\pi}$ and the standard deviations σ_q and σ_π, the covariance–variance ratio can be expressed as

$$\frac{\operatorname{Cov}(q_t, \mathbb{E}_t[\pi_{t+1}])}{\operatorname{Var}(q_t)} = \rho_{q,\pi}\frac{\sigma_\pi}{\sigma_q},$$

so the bias term can be written equivalently as

$$\hat{\beta}_1 = \frac{1}{\nu} - \frac{1-\tau}{\nu}\,\rho_{q,\pi}\,\frac{\sigma_\pi}{\sigma_q},$$

which facilitates empirical assessment by estimating the correlation $\rho_{q,\pi}$ and the standard deviation ratio from the data.

Building on the same setting, the population coefficient of determination for the univariate regression (A6) can be derived directly from the definition

$$R^2 = \frac{\operatorname{Cov}(q_t, \gamma_{t+1})^2}{\operatorname{Var}(q_t)\ \operatorname{Var}(\gamma_{t+1})}. \tag{A9}$$

Using the true model (A5), the covariance between q_t and γ_{t+1} is

$$\operatorname{Cov}(q_t, \gamma_{t+1}) = b\ \operatorname{Var}(q_t) + c\ \operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big), \tag{A10}$$

and the variance of γ_{t+1} is

$$\operatorname{Var}(\gamma_{t+1}) = b^2 \operatorname{Var}(q_t) + c^2 \operatorname{Var}\big(\mathbb{E}_t[\pi_{t+1}]\big) + 2bc \operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big) + \operatorname{Var}(u_{t+1}). \tag{A11}$$

Substituting (A10)–(A11) into (A9) and dividing numerator and denominator by $b^2 \operatorname{Var}(q_t)$, with $c/b = -(1-\tau)$, yields the general expression

$$R^2 = \frac{\left[1-(1-\tau)\dfrac{\operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)}\right]^2}{1-2(1-\tau)\dfrac{\operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)} + (1-\tau)^2\dfrac{\operatorname{Var}\big(\mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)} + \dfrac{\operatorname{Var}(u_{t+1})}{b^2 \operatorname{Var}(q_t)}}. \tag{A12}$$

When the disturbance variance is negligible ($\operatorname{Var}(u_{t+1}) = 0$), (A12) reduces to the compact closed-form expression reported in the main text:

$$R^2 = \frac{\left[1-(1-\tau)\dfrac{\operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)}\right]^2}{1-2(1-\tau)\dfrac{\operatorname{Cov}\big(q_t, \mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)} + (1-\tau)^2\dfrac{\operatorname{Var}\big(\mathbb{E}_t[\pi_{t+1}]\big)}{\operatorname{Var}(q_t)}}.$$

B.4. Proof of Eq. (13)

Allowing firm use intangible capital to generate profit, the profit function is written as

$$\Pi_t = f\big(K_t^{\text{PHY}}, K_t^{\text{INT}}, N_t\big) - \Phi_t^{\text{PHY}} - \Phi_t^{\text{INT}},$$

where the two adjustment cost functions are defined as

$$\Phi_t^{\text{PHY}} = \frac{a}{2}\frac{(I_t^{\text{PHY}})^2}{K_t^{\text{PHY}} + K_t^{\text{INT}}} + I_t^{\text{PHY}}, \qquad \Phi_t^{\text{INT}} = \frac{a}{2}\frac{(I_t^{\text{INT}})^2}{K_t^{\text{PHY}} + K_t^{\text{INT}}} + I_t^{\text{INT}}.$$

Assuming the homogeneous evolution process to bind the physical and intangible capital, the intangible Q is analogously defined as the Lagrangian multiplier for firm's intertemporal optimization problem, and obtain the first-order conditions with respect to I_t^{INT} and K_{t+1}^{INT} as shown in Eq. (3) and Eq. (4).

With the same log-linearization and assumptions regarding to the conditional expectation on the next period's Q, as shown above, we obtain the same Euler equation with respect to both the physical and intangible capital investment, respectively:

$$\mathbb{E}_t[\iota_{t+1}^{\text{PHY}}] = \vartheta_1 + \frac{1}{\nu_1} q_t^{\text{PHY}} + \frac{1}{\nu_1}\mathbb{E}_t[r_{t+1}] - \frac{1-\tau_1}{\nu_1}\mathbb{E}_t[\pi_{t+1}^{\text{PHY}}], \tag{A13}$$

$$\mathbb{E}_t[\iota_{t+1}^{\mathrm{INT}}] = \vartheta_2 + \frac{1}{\nu_2} q_t^{\mathrm{INT}} + \frac{1}{\nu_2}\mathbb{E}_t[r_{t+1}] - \frac{1-\tau_2}{\nu_2}\mathbb{E}_t[\pi_{t+1}^{\mathrm{INT}}]. \tag{A14}$$

where investment-capital ratio for physical and intangible capital is donated as $\iota_{t+1}^{\mathrm{PHY}} = \frac{I_{t+1}^{\mathrm{INT}}}{K_{t+1}^{\mathrm{PHY}}+K_{t+1}^{\mathrm{INT}}}$ and $\iota_{t+1}^{\mathrm{INT}} = \frac{I_{t+1}^{\mathrm{INT}}}{K_{t+1}^{\mathrm{PHY}}+K_{t+1}^{\mathrm{INT}}}$, respectively.

Add the left-hand-side and right-hand-side for Eq. (A13) and Eq. (A14), we obtain the optimal conditional expectation for the one-period-ahead total investment-capital ratio:

$$\mathbb{E}_t[\iota_{t+1}] = \tilde{\vartheta} + \frac{1}{\nu_1} q_t^{\mathrm{PHY}} + \frac{1}{\nu_2} q_t^{\mathrm{INT}} + (\frac{1}{\nu_1} + \frac{1}{\nu_2})\mathbb{E}_t[r_{t+1}] - \frac{1-\tau_1}{\nu_1}\mathbb{E}_t[\pi_{t+1}^{\mathrm{PHY}}] - \frac{1-\tau_2}{\nu_2}\mathbb{E}_t[\pi_{t+1}^{\mathrm{INT}}],$$

where ι_{t+1} is the total investment ratio defined by $\iota_{t+1}^{\mathrm{PHY}} + \iota_{t+1}^{\mathrm{INT}}$. Hence, the equation yields a testable panel predictive regression form as shown by Eq. (13).

C. Technique note for MVK estimation

C.1. AD estimation process

The Table **1** in our main text has displayed a brief application mechanism of automatic differentiation (AD). In this sub-section, we provide additional explanations of how AD operates and why it is effective for our neural network–based estimation of the marginal value of capital (MVK). Automatic differentiation, also called algorithmic differentiation, is a computational technique that evaluates exact derivatives of a numerical function by systematically applying the chain rule to every elementary operation executed in the program. Unlike numerical differentiation, which uses finite differences and suffers from truncation and round-off error, or symbolic differentiation, which manipulates algebraic expressions and often leads to expression swell, AD operates directly on the computational graph of the implemented algorithm and returns derivatives accurate to machine precision with only modest computational overhead.

To formalize the idea, suppose a scalar-valued function $y = f(x) : \mathbb{R}^n \to \mathbb{R}$ is computed by a sequence of intermediate variables

$$v_0 = x_0, \quad \ldots, \quad v_{n-1} = x_{n-1}, \qquad v_k = \phi_k\big(v_{i_1(k)}, \ldots, v_{i_{m(k)}}\big), \quad k = n, \ldots, L,$$

where each ϕ_k is an elementary differentiable operation and the final output is $y = v_L$. AD traverses this computational graph either forward or backward to propagate derivative information. In forward mode, for a chosen input direction e_j, each variable v_k is augmented with its tangent $\dot{v}_k = \partial v_k / \partial x_j$, updated recursively

$$\dot{v}_k = \sum_{\ell \in \text{parents}(k)} \frac{\partial \phi_k}{\partial v_\ell} \dot{v}_\ell, \quad \text{starting with } \dot{v}_i = \delta_{ij}.$$

This computes one column of the Jacobian ∇f alongside the primal evaluation. In contrast, reverse mode—which is particularly efficient when y is scalar and x is high-dimensional—first runs a forward pass to store every v_k, then propagates adjoints

$$\lambda_k = \frac{\partial y}{\partial v_k}$$

backward by the chain rule

$$\lambda_\ell = \sum_{k:\, \ell \in \text{parents}(k)} \lambda_k \frac{\partial \phi_k}{\partial v_\ell}, \qquad \text{initialized with } \lambda_L = 1.$$

When the computational graph corresponds to a feedforward neural network with layers

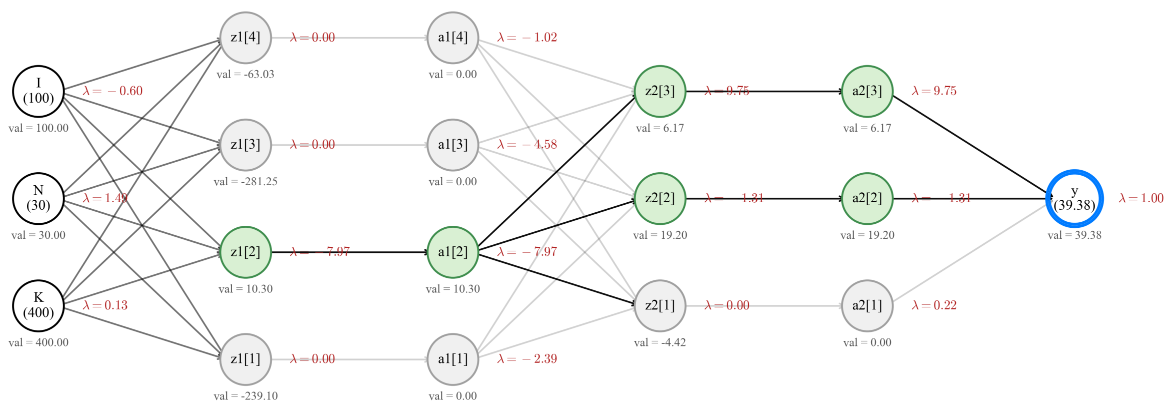
$$z_\ell = W_\ell^\top a_{\ell-1} + b_\ell, \qquad a_\ell = \sigma(z_\ell), \quad \sigma(u) = \max\{0, u\},$$

the forward pass records inputs v_{-2}, v_{-1}, v_0 (capital K, labor N, and investment I), the pre-activations z_ℓ (linear combinations of previous activations), and the post-activations a_ℓ (outputs after the ReLU nonlinearity). The backward pass initializes $\lambda_{\text{out}} = 1$ at the scalar output and propagates gradients layer by layer, using the derivative of the ReLU, $\sigma'(u) = \mathbf{1}_{\{u>0\}}$, to block gradient flow through inactive units.

Figure A1 visualizes this process for the two-layer ReLU network used in our analysis, evaluated at $(K, N, I) = (400, 30, 100)$. Each node shows its forward value (val) and its backward adjoint sensitivity (λ), while edges indicate how signals flow between layers. Green nodes represent active ReLU units (transmitting gradients), while gray nodes are inactive ($a = 0$) and contribute no gradient because $\sigma'(u) = 0$. This diagram complements the tabular presentation in Table 1 by illustrating how the forward pass builds the computational graph and how reverse-mode AD walks it backward to recover the derivatives of the network output with respect to each input.

Figure A1: Computation graph for AD

This figure illustrates the forward pass and adjoint (reverse-mode) computation in a two-layer ReLU multilayer perceptron (MLP) evaluated at input values $K = 400$ (capital), $N = 30$ (labor), and $I = 100$ (investment), yielding output $y = 39.38$. Here, $z_1[i]$ ($i = 1, \ldots, 4$) denote pre-activation values in the first hidden layer (linear combinations of inputs); $a_1[i]$ represent post-activation values via the ReLU function, $\max(0, z_1[i])$, which introduces non-linearity by zeroing negative inputs; $z_2[j]$ ($j = 1, \ldots, 3$) are pre-activation values in the second hidden layer (linear combinations of a_1); $a_2[j]$ are corresponding ReLU activations; *val* indicates forward-pass computed values at each node; λ signifies adjoint sensitivities (partial derivatives of the loss with respect to each node, propagated backward via the chain rule for gradient computation); active ReLU nodes (green, where $a > 0$) transmit gradients fully, while inactive ones (gray, where $a = 0$) prune paths due to ReLU's zero derivative for negative inputs, as detailed in the accompanying automatic differentiation example.



C.2. Hyper-parameters tuning for neural network model

The estimation of partial derivatives via automatic differentiation is quite dependent on the shape of the fitted neural network model, which is built on a complex structure of hyperparameters. As our estimation is ex ante and formed based on the information available at time t, the tuning of hyperparameters should target the model's out-of-sample performance. This design makes the model more suitable for conditional-expectation formation and alleviates concerns about overfitting that arise when optimizing only in-sample fit.

To implement this, we use the initial three-year period (1970–1972) to train the model and optimize performance on the 1974 data. Model performance is evaluated by the absolute difference between the fitted one-year-ahead operating profit and the realized Π_{it+1}. In sum, the statistically optimal combination of hyperparameters is chosen by minimizing the average absolute difference between the fitted and realized one-year-ahead profits for 1974.

Specifically, we grid-search all possible combination of the bellowing range of hyper-parameters:

```
param_grid = {
    'num_layers': [1, 2, 3, 4, 5],
    'num_units': [32, 64, 96, 128, 160],
    'activation': ['relu', 'sigmoid', 'elu'],
    'l1_reg': [1e-5],
    'use_ln': [False],
    'opt': ['adam','sgd'],
    'lr': [1e-5],
    'epochs': [100, 200, 300],
    'batch_size': [256, 512],
    'patience': [10, 20],
}
```

`num_layers` controls the depth of the neural network (1–5 fully connected layers), while `num_units` sets the width of each layer (32–160 neurons). `activation` selects the nonlinear transformation (`relu`, `sigmoid`, `elu`) that maps inputs to hidden representations. `l1_reg` adds L1 regularization to reduce overfitting by encouraging sparsity of weights. `use_ln` indicates whether to apply layer normalization, improving training stability. `opt` specifies the optimization algorithm (Adam or SGD), and `lr` sets the learning rate for parameter updates. `epochs` defines the maximum training iterations (100–300), while `batch_size` determines how many samples are used per gradient update (256 or 512). Finally, `patience` is the early-stopping patience (10 or 20 epochs) to prevent unnecessary training once the validation loss stops improving.

Note that the grid-search process is a time-consuming process, although more complex parameter ranges are available, we find they will not significantly affect the outcomes for the estimation of gradients, but just add more computation time.

D. Estimation of cost of capital

D.1. WACC

Our first two measures of cost of capital are WACC with the cost of equity component estimated using CAPM and Fama-French three-factor model, respectively. For each calendar year t, we estimate the following time-series regression for each individual stock i using all available daily excess returns:

$$r_{i,d} - r_{f,d} = \alpha_{i,t} + \beta_{i,t}\left(r_{m,d} - r_{f,d}\right) + \varepsilon_{i,d}$$

where $r_{i,d}$ is the daily stock return obtained from CRSP, $r_{f,d}$ is the daily risk-free rate proxied by the ten-year Treasury yield from the FRED, and $r_{m,d}$ is the daily value-weighted market return obtained from Ken French's website. Next, with the estimated coefficient $\hat{\beta}_{i,t}$, the annually cost of equity is then computed in each year t by:

$$r_e^{\text{CAPM}} = r_{f,t} + \hat{\beta}_{i,t} E_t\left(r_{m,t} - r_{f,t}\right),$$

where $r_{f,t}$ is annually compounded risk-free rate, $E_t\left(r_{m,t} - r_{f,t}\right)$ is the expected market premium in year t, calculated by the history average of each year's daily compound market premium.

Similarly, the cost of equity based on the Fama-French three-factor model is also estimated by this two-stage process. First, the annually factor loadings are estimated by regressing excess return on market premium, SMB, and HML:

$$r_{i,d} - r_{f,d} = \alpha_{i,t} + \beta_{i,t}^{\text{MKT}}\left(r_{m,d} - r_{f,d}\right) + \beta_{i,t}^{\text{SMB}} SMB_d + \beta_{i,t}^{\text{HML}} HML_d + \varepsilon_{i,d},$$

where the daily data for SMB and HML are downloaded from Ken French's website. Second, the annually cost of equity is then computed in each year t using the three obtained coefficient $\hat{\beta}_{i,t}^{\text{MKT}}$, $\hat{\beta}_{i,t}^{\text{SMB}}$, and $\hat{\beta}_{i,t}^{\text{HML}}$:

$$r_e^{\text{FF-3}} = r_{f,t} + \hat{\beta}_{i,t}^{\text{MKT}} E_t\left(r_{m,t} - r_{f,t}\right) + \hat{\beta}_{i,t}^{\text{SMB}} E_t[SMB_t] + \hat{\beta}_{i,t}^{\text{HML}} E_t[HML_t],$$

where $E_t[SMB_t]$ and $E_t[HML_t]$ is history average of each year's daily compound factor value of SMB and HML.

Finally, with measure of cost of equity, the two measures of WACC are formally defined by:

$$r_{it} = (1 - \frac{D}{V}) \times r_{it}^{e} + \frac{D}{V}(1 - Lev) \times r_{it}^{d},$$

where $\frac{D}{V}$ is firm's market leverage, measured by book value of debt divided by market value. Lev is firm's average tax rate (Compustat item txt / pi).

D.2. ICC

Our third measure of cost of capital is ICC estimated from Gordon's finite dividend discount model, following the approach used by Pástor et al. (2008), Lee et al. (2009), and Li et al. (2013). Specifically, ICC is estimated by solving the following model:

$$P_t = \sum_{k=1}^{T=15} \frac{\mathbb{E}_t[EPS_{t+k}] \times (1 - b_{t+k})}{(1 + ICC_t)^k} + \frac{\mathbb{E}_t[EPS_{t+k}]}{ICC_t(1 + ICC_t)^T}, \tag{A15}$$

where $\mathbb{E}_t[EPS_{t+k}]$ is the consensus median analyst forecast of earnings per share for k-years-ahead; b_{t+k} is the estimated plowback rate; P_t is stock price. The terminal value term is calculated using the no-growth perpetuity theory for the long run. The analyst forecasts data are obtained from I/B/E/S, stock price data is from CRSP.

Following Pástor et al. (2008), we obtain analyst forecast of one-year-ahead EPS ($\mathbb{E}_t[EPS_{t+1}]$) and two-year-ahead EPS ($\mathbb{E}_t[EPS_{t+2}]$) and calculate three-years-ahead EPS forecast by $\mathbb{E}_t[EPS_{t+2}] \times (1 + LTG)$, where LTG is analyst forecast of long-term EPS growth. To generate EPS forecast for $t+4$ to $t+T+1$, we assume that the earnings growth rate for year $t+3$, g_{t+3} is measured by LTG. Next, g_{t+3} will exponentially reverts to a long-run steady value represented by the long-run nominal gross domestic product (GDP) growth rate. In cases where the growth rate g_3 is negative, we replace it with the year $t+2$ earnings growth rate, denoted as g_2. Consequently, for $k = 4, ..., T+1$, we calculate the earnings growth and earnings expectations as follows:

$$\begin{aligned} g_{e,t+k} &= g_{e,t+k-1} \times exp[log(g_e/g_{e,3})/T], \\ \mathbb{E}_t[EPS_{t+k}] &= \mathbb{E}_t[EPS_{t+k-1}] \times (1 + g_{e,t+k}). \end{aligned}$$

The initial one-year ahead plowback rate, denoted as b_1, for year $t+1$ is estimated as one minus the most recent year's total payout ratio, which is available from the Financial Ratios Suite by WRDS. We assume that the plowback rates revert linearly to a steady-state plowback rate, denoted as b, following the formula: $b_{t+k} = b_{t+k-1} - (b_1 - b)/T$. According to the sustainable growth rate formula, in the steady state, the earnings growth rate equals the return on new investment (ROI) multiplied by the plowback rate. We further assume that ROI is the same as the expected return for new investments. Therefore, the plowback rates for years $t+3$ to $t+T$ can be computed as $b_{t+k} = b_{t+k-1} - (b_1 - b)/T$ for $k = 3, ..., T$.