Real Anomalies

Jules H. van Binsbergen* Christian C. Opp†

October 7, 2016

Preliminary, comments welcome.

Abstract

We examine the importance of asset pricing anomalies for the real economy. When firms interpret public information in the way it is reflected in market prices, informational inefficiencies manifesting in financial markets as anomalies can cause material real inefficiencies. We estimate the joint dynamic distribution of firm characteristics that have been linked to anomalies and other firm variables, such as investment, capital, and value added. Based on a model that matches these joint dynamics, we then evaluate the counterfactual dynamic distribution of these quantities in an informationally efficient economy, and find significant deviations. Our results suggest that if financial- and academic institutions helped reduce and/or eliminate such anomalies they could provide large value added to the economy. We show that informational inefficiencies are particularly destructive for high Tobin's q firms, and that the persistence and the amount of mispriced capital are major determinants of the real economic consequences.

JEL classification: D22, D24, D53, D92, E22, G2, G30

*University of Pennsylvania, The Wharton School, email: julesv@wharton.upenn.edu.
†University of Pennsylvania, The Wharton School, email: opp@wharton.upenn.edu.

1. Introduction

In the past few decades a vast literature has developed that attempts to document and explain the behavior of asset prices both in the cross section and the time series. The seminal paper on excess volatility (Shiller (1981)) has spurred a literature that attempts to explain why stock markets are so volatile and whether or not such volatility is excessive (irrational), relative to the existing models. Similarly, many different "anomalies" have been uncovered in the cross-section of asset prices, such as the value premium puzzle, the investment anomaly, the profitability anomaly, and momentum.[1] One important question that follows from these empirical findings is how harmful these patterns in expected returns would be for the real economy if they indeed reflected informational inefficiencies. In this paper we ask what the economy-wide real implications are when firms interpret public information in a way that is consistent with observed market prices, and financial markets are subject to anomalies.

In the finance literature, mispricings are typically estimated based on realizations of so-called alphas, that is, deviations of average returns from a benchmark asset-pricing model. While realizations of alpha indicate that imperfections exist, they can be poor indicators of the economic importance of anomalies for at least three reasons. First, they only represent changes of asset mispricings. Mispricing is an inherently dynamic phenomenon — as alphas are realized over time, firms are only temporarily affected by distortions. As a consequence, we need to also consider the persistence of the mispricing. For example, one may wonder whether for the aggregate economy it is worse for firms to have a very short-lived alpha of 5% versus a very persistent alpha of 1%. From an investors' point of view, the short-lived alpha may seem more interesting, but for the firms investment decisions, it seems hard to imagine that such short-lived mispricings matter. Second, as alphas are *return* measures, they do not give an accurate representation of the *value* of the mispricing. Just as the internal rate of return cannot be used to measure the value of an investment opportunity (it is the net present value that does), the alpha cannot be used to measure the economic importance of an anomaly.[2] Thirdly, and most importantly, it is not clear from studying alphas to what extent mispricings translate into real investment and value added distortions.

[1] See papers as early as Rosenberg, Reid, and Lanstein (1985) for value and Jegadeesh and Titman (1993) for momentum. For a recent overview of value and momentum in various asset classes see Asness, Moskowitz, and Pedersen (2013).

[2] See also Berk and van Binsbergen (2015) who use this same argument to show that the alpha of a mutual fund manager is a poor measure of the manager's skill.

In order to assess the real effects of documented anomalies quantitatively, we estimate the joint dynamic distribution of firm characteristics that have been linked to mispricing and other firm variables, such as investment, capital, and output. Based on a structural model that matches these joint dynamics we then evaluate the counterfactual dynamic distribution of capital, investment, and value added absent anomalies. Deviations between the counterfactual and the actual distribution allow us to assess the magnitude of real inefficiencies caused by asset-pricing anomalies. As such, while the existing literature evaluates asset pricing anomalies primarily based on the statistical significance of alphas, we aim to provide a framework to gauge economic significance, as measured by distorted investment and the present value of losses in value added.

We find that cross-sectional anomalies can have important effects on value added and investment. Even if we assume that the aggregate firm's cash flows are correctly priced by the market (the Capital Asset Pricing Model (CAPM) alpha of the index is by definition 0), cross-sectional anomalies have important effects. Due to distortions in the cost of capital for individual firms, some firms overinvest and others underinvest, both leading to suboptimal investment decisions and value destruction. We show that in our current calibration, cross-sectional distortions can represent a significant fraction of several percentage points of value added. As a consequence, if the financial sector eliminated these mispricings it could justifiably extract large rents from this process.

Rather than taking a stance on which asset pricing model is the correct one, we aim to provide a flexible methodology that allows assessing real distortions from pricing errors conditional on a variety of standard asset pricing models. It is well known since Fama (1970) that the informational efficiency of prices per se is not testable. It must be tested jointly with some postulated asset-pricing model. As such, pricing errors are always estimated *conditional* on the pricing model that an econometrician imposes. Yet, just as the joint-hypothesis problem has not invalidated the empirical asset pricing literature, it does not invalidate estimating the real implications of potential anomalies.

Our calculations shed light on another important debate in the literature on financial intermediation. One often heard critique of active mutual funds is Sharpe's arithmetic. Sharpe divided all investors into two sets: people who hold the market portfolio, whom he called "passive" investors, and the rest, whom he called "active" investors. Because market clearing requires that the sum of active and passive investors' portfolios is the market portfolio, the sum of just active investors' portfolios must also be the market portfolio. This observation is used to imply that the abnormal return of the average active investor must be zero, what has become known as Sharpe's critique.[3] The problem with this logic is that it does not take into account what the market portfolio would have looked like under the counterfactual of no active management. Our paper suggests that absent alphas, firms' investment decision are better, thus leading to more real value creation in the economy. To the extent that active mutual funds trade on and thereby reduce alphas, this leads to a more valuable market portfolio. Sharpe's arithmetic is thus not informative regarding the question of whether or not active management adds value to the economy. Put differently, there is a free-riding problem that allows passive investors to benefit from the price corrections induced by active investors. By simply comparing the performance of active and passive investors (the financial arithmetic), these gains from altering real economic outcomes are not taken into account.

2. Related Literature

To our knowledge we are the first to quantitatively assess the real value losses associated with cross-sectional financial market anomalies (alphas). Our focus on the real effects connects our study to the literature in macroeconomics quantifying efficiency losses due to capital misallocations (Hsieh and Klenow, 2009).[4]

The fact that we study the real implications of informational inefficiencies in financial markets fundamentally differentiates our study from the large literatures in macroeconomics and corporate finance that study external financing frictions that drive wedges between internal and external funds (Whited, 1992, Kiyotaki and Moore, 1997, Gomes, Yaron, and Zhang, 2003, Hennessy and Whited, 2007). For example, these wedges take the form of leverage constraints and issuance cost that limit insiders' ability to raise funds externally, and can potentially constrain investment. Yet they may also involve information asymmetries that can lead insiders to use *private* information to raise external funds from markets at opportune times.[5]

[3] Berk and van Binsbergen (2014) provide other arguments for why Sharpe's arithmetic is flawed.

[4] See also Eisfeldt and Rampini (2006) for evidence on the amount of capital reallocation between firms and the cost of reallocation.

[5] Gilchrist, Himmelberg, and Huberman (2005) argue that dispersion in investor beliefs and short-selling constraints can lead to stock market bubbles and that firms, unlike investors, can exploit such bubbles by issuing new shares at inflated prices. This lowers the cost of capital and increases real investment. They use the variance of analysts' earnings forecasts to proxy for the dispersion of investor beliefs, and find that increases in dispersion cause increases in new equity issuance, Tobin's q, and real investment, as predicted by their model. Baker, Stein, and Wurgler (2003) test the prediction that stock prices have a stronger impact on the investment of equity-dependent firms – firms that need external equity to finance marginal investments — and find strong support for it.

In contrast, the informational inefficiencies we study are measured with respect to publicly available information. By definition insiders and outsiders have symmetric access to this type of information — some of the most prominent anomalies even rely only on salient market and accounting information, such as book and market values. As it is not obvious whether insiders or outsiders are better in processing all publicly available information relevant for prices, we do not impose that either party is better at this task. For example, it is not clear whether hedge funds or firm managers know better in what percentile of the book-to-market distribution a firm is, or what a firm's current exposure (beta) to various priced risk factors is, or what the fair risk prices in light of all public information about current macroeconomic conditions are. The friction we study is therefore not driving a wedge between internal and external funds, but instead creates a wedge between the efficient and actual use of all publicly available information, which lies at the heart of an asset pricing anomaly.

In contrast, Warusawitharana and Whited (2016) evaluate the shareholder value implications of an informational wedge between insiders and outsiders, where managers are better informed and therefore perceive different valuations. Further, whereas we aim to provide a methodology to estimate the real effects of anomalies measured with respect to standard asset pricing models, Warusawitharana and Whited (2016) consider misvaluation shocks affecting a representative firm that are measured based on hedonic regressions. Our paper provides a methodological contribution by presenting a simple continuous time framework that features many of the important characteristics that have been posed in the literature, yet still maintains a simple tractable setting that allows us to evaluate the stationary cross-sectional distribution of the quantities of interest in closed-form.

Rather than processing all public information in the same way as market participants, managers might simply use market prices as a signal that aggregates relevant information. The notion that managers rely on market prices when making investment decisions is central to the literature on the real feedback effects of financial markets. The influence of prices on the allocative efficiency of resources goes at least as far back as Hayek (1945). Further, the allocational role of prices in secondary financial markets and their influence on real investment has been studied theoretically in papers such as Leland (1992), Dow and Gorton (1997), Subrahmanyam and Titman (2001), and Goldstein, Ozdenoren, and Yuan (2013).

On the empirical side, several papers also find evidence consistent with firm behavior that responds to information encoded in financial market prices. Barro (1990) shows that changes in stock prices have substantial explanatory power for U.S. investment, especially for long-term samples, and even in the presence of cash flow variables. The specification he employs outperforms standard Tobin's q regressions. Chen, Goldstein, and Jiang (2007) show that two measures of the amount of new information in stock prices — price nonsynchronicity and probability of informed trading — have a strong positive effect on the sensitivity of corporate investment to stock prices. They argue that firm managers learn from the information about fundamentals encoded in stock prices and incorporate this information in corporate investment decisions. Polk and Sapienza (2009) use discretionary accruals as a proxy for mispricing and find a positive relation between abnormal investment and discretionary accruals. Edmans, Goldstein, and Jiang (2012) identify a strong effect of market prices on takeover activity, and conclude that that financial markets have real effects by affecting managers' behavior.

Hou, Xue, and Zhang (2015, 2016) show that sorting firms on investment yields large spreads in alphas, no matter if one imposes the CAPM, the Fama-French 3 factor model, the Carhart model, or the Pastor-Stambaugh model. These results suggest that firm investment indeed responds to the abnormal components of discount rates encoded in market prices. Bai, Philippon, and Savov (2016) find evidence that market-based information production has increased since 1960 — prices have become a stronger predictor of investment and investment a stronger predictor of cash flows. Price informativeness has increased at longer horizons, in particular among firms with greater institutional ownership and share turnover, firms with options trading, and growth firms.

David, Hopenhayn, and Venkateswaran (2016) estimate that a firm's inability to perfectly predict its own productivity when making ex ante investment decisions causes sizable output losses. The authors also estimate that firm's learn more about their firm-specific productivity from private signals than from market prices. Conceptually we are interested in a different question — the starting point for our analysis is the large literature documenting asset pricing anomalies, which are estimated conditional on an asset pricing model such as the CAPM. The counterfactual we evaluate is not a world where firms can perfectly predict their idiosyncratic productivity but one where prices are informationally efficient with respect to public information. For our question, we can be agnostic if managers learn from market prices, or if they simply interpret public information in the same way as market participants, and thus assign the same distorted valuations as the market. The fact that studies like Morck, Shleifer, and Vishny (1990) find weak *incremental* explanatory power of stock prices for investment spending is therefore not per se inconsistent with our analysis. What does matter is that managers choose investment policies that maximize the observed market prices of their firms, such that policies — just like market prices — might be informationally inefficient. It is important to note that in our dynamic model with lumpy asset growth and adjustment cost, this firm behavior is fully consistent with a highly noisy investment-q relation and wide dispersion in measures of marginal revenue products.

3. Reduced-Form Estimates and Their Shortcomings

In this section we present reduced-form measures of mispricing. We start by computing the alphas of the decile-sorted portfolios, the Markov transition matrices of the decile portfolios, as well as the dollar-values represented by these portfolios.

3.1. Alphas

First we replicate CAPM alphas on four well-known anomalies for the sample period 1975-2014. We sort firms into decile portfolios based on their (1) lagged book-to-market ratio, (2) investment as measured by asset growth, (3) gross profitability, as well as (4) their past annual return (momentum) and form 10 value-weighted portfolios each month. We then regress the portfolio excess returns (returns on decile i denoted by R_{it} minus the risk free rate R_{ft}) on the excess return of the market $R_{mt} - R_{ft}$:

$$R_{i,t+1} - R_{ft} = \alpha_{i,btm} + \beta_{i,btm}(R_{m,t+1} - R_{ft}) + \varepsilon_{i,t+1} \tag{1}$$

The results are summarized in Panels A and B of Table 1. The panels confirm the findings in the literature that there are return spreads that are not explained by the CAPM. Firms with high book-to-market ratios, low investment, high gross profitability and high past returns earn high average returns over this sample period with annual return spreads ranging from 3% (profitability) to 12% (momentum). To get a first sense of the aggregate mispricing (MP) we compute the average absolute value of the alpha. That is, we define the mispricing measure MP for anomaly j as:

$$MP_j = \frac{\sum_{i=1}^{10} \mid \alpha_{i,j} \mid}{10}. \tag{2}$$

The results are summarized in the second column of Table 2. The numbers range from 1.1% for profitability to 3.6% for momentum. One potential downside of this measure of aggregate mispricing, however, is that it does not properly account for size differences

Decile	1	2	3	4	5	6	7	8	9	10
					Panel A: Raw Returns					
BtM	0.0100	0.0092	0.0104	0.0104	0.0112	0.0113	0.0124	0.0139	0.0138	0.0167
Invest	0.0133	0.0120	0.0123	0.0127	0.0117	0.0111	0.0106	0.0104	0.0113	0.0089
Profitability	0.0099	0.0109	0.0112	0.0101	0.0109	0.0113	0.0115	0.0106	0.0115	0.0125
Momentum	0.0051	0.0085	0.0083	0.0108	0.0094	0.0100	0.0118	0.0126	0.0133	0.0155
					Panel B: CAPM Alphas					
BtM	-0.0014	-0.0014	0.0001	0.0000	0.0011	0.0010	0.0020	0.0034	0.0030	0.0059
Invest	0.0019	0.0013	0.0021	0.0030	0.0018	0.0013	0.0002	-0.0004	0.0000	-0.0033
Profitability	-0.0020	0.0004	0.0014	-0.0002	0.0006	0.0009	0.0008	0.0000	0.0013	0.0015
Momentum	-0.0101	-0.0041	-0.0032	0.0003	-0.0008	0.0001	0.0019	0.0025	0.0027	0.0039
					Panel C: Time Series Average of Decile's Equity Value as Fraction of Total					
BtM	0.1420	0.1445	0.1325	0.1208	0.1152	0.0993	0.0931	0.0776	0.0545	0.0204
Invest	0.0219	0.0505	0.0889	0.1204	0.1312	0.1434	0.1491	0.1223	0.1037	0.0687
Profitability	0.0529	0.0839	0.1016	0.1112	0.1424	0.1109	0.0976	0.1052	0.1129	0.0814
Momentum	0.0173	0.0520	0.0856	0.1107	0.1269	0.1396	0.1438	0.1378	0.1197	0.0665
					Panel D: Time Series Average of Decile's Aggregate Firm Value (Equity plus Debt) as Fraction of Total					
BtM	0.0556	0.0625	0.0687	0.0804	0.1038	0.1230	0.1508	0.1962	0.1299	0.0290
Invest	0.0252	0.0566	0.0962	0.1217	0.1305	0.1490	0.1511	0.1216	0.0914	0.0568
Profitability	0.2436	0.2142	0.1188	0.0834	0.0896	0.0643	0.0511	0.0497	0.0503	0.0350
Momentum	0.0255	0.0608	0.0923	0.1179	0.1332	0.1445	0.1434	0.1308	0.1047	0.0468
					Panel E: Persistence as Measured by Diagonal Element of Decile in Markov Matrix					
BtM	0.5771	0.3524	0.2890	0.2583	0.2461	0.2444	0.2634	0.3026	0.3379	0.5274
Invest	0.2750	0.1789	0.1628	0.1513	0.1531	0.1496	0.1496	0.1604	0.1854	0.2274
Profitability	0.6569	0.6135	0.5384	0.4556	0.3997	0.3762	0.3747	0.4054	0.4700	0.6639
Momentum	0.1758	0.1150	0.1033	0.1055	0.1079	0.1218	0.1131	0.1080	0.0973	0.1158

Table 1

Anomalies: The table reports several characteristics of often-studied anomalies. We sort stocks into portfolios based on (1) their lagged book-to-market ratio (value-growth), (2) their investment (percentage change in total assets), (3) their operating profitability and (4) their past 12-month return (momentum). Panel A reports average monthly returns for each decile portfolio. Panel B reports monthly CAPM alphas. Panel C reports the deciles average weight in terms of equity outstanding. That is, for each month we compute the amount of equity outstanding in the decile and divide this by the total amount of equity across all deciles. We then take a time series average of these weights. Panel D reports the same quantities as Panel C but using total firm value (debt plus equity). Panel E reports the diagonal element of the decile in the annual Markov transition matrix.

across the deciles. That is, if most (least) capital is concentrated in the deciles with the least mispricing, the measure in (2) overstates (understates) the amount of mispricing. To address this issue, we recompute the measures above using the aggregate market value of equity (E) of the decile as weights in the computation. Define the weight of decile i for anomaly j as:

$$we_{i,j} = \frac{1}{T} \sum_{t=1}^{T} \frac{E_{i,j}}{\sum_{i=1}^{10} E_{i,j}}, \tag{3}$$

then the equity-weighted mispricing measure is given by:

$$EMP_j = \sum_{i=1}^{10} we_{i,j} \mid \alpha_{i,j} \mid . \tag{4}$$

The weights $we_{i,j}$ are summarized in Panel C of Table 1. The results are summarized in the third column of Table 2. Interestingly, for all anomalies this weighted average (EMP) is lower than the simple average (MP) and particularly so for momentum and book-to-market, suggesting that more mispricing occurs in deciles with lower market capitalizations. Another important question that naturally arises is whether the mis-

	MP	EMP	VMP
BtM	0.0231	0.0165	0.0232
Invest	0.0185	0.0166	0.0166
Profitability	0.0109	0.0100	0.0124
Momentum	0.0355	0.0241	0.0243

Table 2
An overview of aggregate mispricing measures for different anomalies. The measure MP is simply the average absolute value of the CAPM alpha cross the deciles. The measure EMP computes an average absolute alpha as well but on a weighted basis. The weights of each decile are determined by the amount of equity capital in that decile. The measure VMP is the same as the measure EMP but uses total firm value to compute the weights.

pricing only applies to the equity portion of the balance sheet or to the debt portion as well. Put differently, what if the debt fraction of the firm is similarly mispriced as the equity portion? To assess the importance of mispricing in that case, we recompute the value-weighted measure using total firm value in the weights. That is, we compute the weights:

$$wm_{i,j} = \frac{1}{T} \sum_{t=1}^{T} \frac{E_{i,j} + L_{i,j}}{\sum_{i=1}^{10} E_{i,j} + L_{i,j}}, \tag{5}$$

where $L_{i,j}$ is the book value of the liabilities of each firm, and define the firm-value-weighted mispricing measure as:

$$VMP_j = \sum_{i=1}^{10} wm_{i,j} \mid \alpha_{i,j} \mid . \tag{6}$$

The results are summarized in the last column of Table 1. Interestingly, we find that by weighting by total firm size, the weighted mispricing of book-to-market sorted portfolios is about the same size as the unweighted one. The mispricing measure of momentum remains significantly lower on a weighted basis, regardless of the weighting scheme.

Even though these measures could be useful as first estimates of the importance of mispricing, they miss one important feature which is the persistence of the mispricing. Momentum is a short-lived phenomenon, whereas value is a long-lived one. One way to assess the persistence is to compute Markov transition matrices that summarize how firms migrate across deciles. We compute for each anomaly an annual Markov transition matrix. The diagonal elements of these matrices are summarized in Panel E of Table 1.[6] As expected, the table shows that momentum is a much shorter-lived anomaly than the value anomaly. The average diagonal element for value is 0.34, with values above 0.5 for the two extreme portfolios. For momentum the average diagonal element is 0.12 and the first diagonal element (the (1,1) element) equals 0.17. Given that momentum is so much less persistent than value we would expect this lack of persistence to substantially lower the influence of momentum on firms' decisions. The most persistent anomaly is profitability with an average diagonal Markov element of 0.50 across the deciles.

There also seems to be a negative relationship between the mispricing measures and their persistence. The three panels in Figure I plot the mispricing measured against the persistence of the anomalies as measures by the average diagonal elements of the 1st and 10th deciles. For all mispricing measures the relationship is negative: the short-lived anomalies have high mispricing measures but low persistence.

[6]The full Markov matrices are listed in the Appendix.

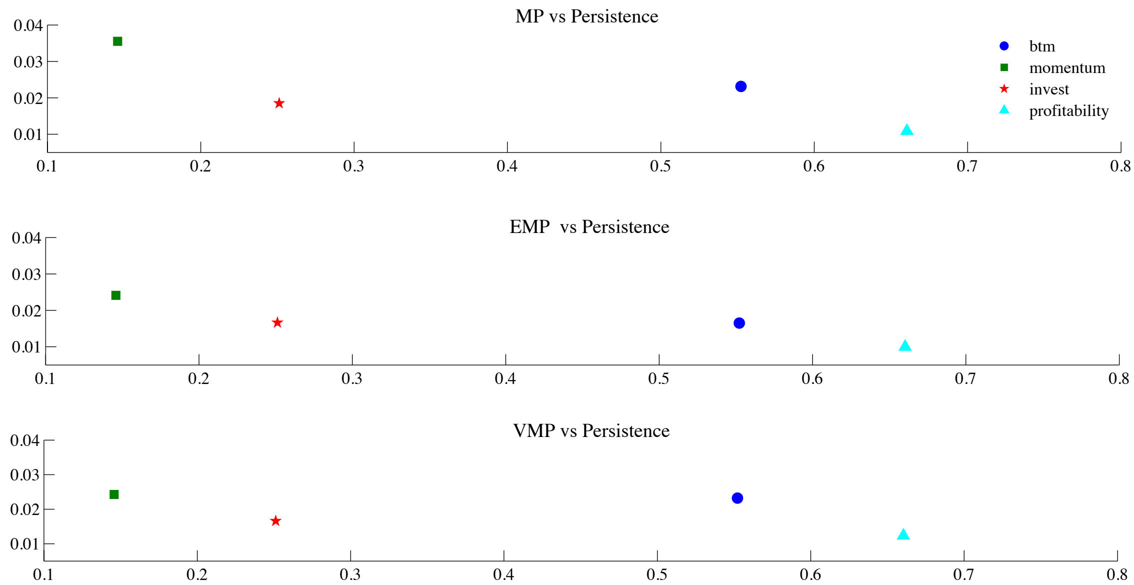

FIGURE I

Mispricing Measures vs Persistence. The graphs plots for each anomaly the mispricing measure against the persistence of the anomaly. The persistence is measures as the average of the (1,1) and the (10,10) element of the annual Markov transition matrix of firms across the deciles. Panel A uses MP as the mispricing measure. Panel B and C use EMP and VMP respectively.

We have now presented a range of reduced-form measures of mispricing. Even though all these measures give an impression of how important cross-sectional mispricing can be, none of them address arguably the most important question. What is the influence of these anomalies on real economic quantities?

3.2. A Different Counterfactual

To better understand why alpha measures by themselves (even the weighted ones) are not informative regarding economic losses, consider the following one-period example. Consider a firm i that generates cash flows at time 1, denoted by CF_{i1}. The value at time 0 that the market places on these cash flows is a function of α_{i0} and given by:

$$V_{i0}(\alpha_{i0}) = \frac{E_0\left[CF_{i1}\right]}{1 + r_f + \beta_i \psi_0 + \alpha_{i0}}. \tag{7}$$

where r_f is the risk free rate, ψ_0 is a measure of the risk price and β_i measures the usual scaled covariation with the stochastic discount factor. Interpreting α_{i0} as a distortion in the discount rate is isomorphic to a distortion in the beliefs (probabilities) regarding the cash flows. Suppose we observe $\alpha_{i0} = \phi$ in the data. It is clear that the value of the firm V_{i0} is affected by this distortion at time 0. However, as long as the actual cash flows of the firm are not affected by α_{i0}, and thus real economic quantities are unaffected, the misvaluation will resolve itself at time 1 through the higher return, and no further losses to the economy occur. Computing the ratio of $V_{i0}(\alpha_{i0} = 0)$ to $V_{i0}(\alpha_{i0} = \phi)$ in this one-period example is equally informative as the value of ϕ itself.

In this paper, we are interested in the value distortions that happen when the cash flows of the firm *are* affected by α_{i0}. That is, we model the real investment decisions of the firm as a function of α_{i0}, i.e. $CF_{i1}(\alpha_{i0})$.[7] We then compare the valuation under the actual firm policies:

$$V_{i0}^{act} = \frac{E_0\left[CF_{i1}(\alpha_{i0} = \phi)\right]}{1 + r_f + \beta_i \psi_0} \tag{8}$$

to the valuation under the optimal firm policies:

$$V_{i0}^{opt} = \frac{E_0\left[CF_{i1}(\alpha_{i0} = 0)\right]}{1 + r_f + \beta_i \psi_0}. \tag{9}$$

Note that in both cases we discount by the "true" discount rate ($\alpha = 0$), not the discount rate that is distorted ($\alpha = \phi$). Finally, in this one-period example there was no dynamic resolution of alpha over time, but in reality different types of mispricing resolve over different time horizons. We thus need a model that allows for such dynamic resolution.

In summary, the model that we need has three requirements. First, it is has to be dynamic. Second, it should be easy-to-solve, in the sense that given a policy we would like to obtain closed-form solutions of the stationary cross-sectional distribution of the variables of interest. Third, it should be easy to estimate in the data. We describe such a model in the next section.

4. The Model

The economy we study is in continuous time. A cross-section of firms operate technologies with decreasing returns to scale and capital adjustment costs. The structural parameters of the model are governed by a set of state variables, which are described in detail in Section 4.2 below. For notational convenience, we will omit parameters'

[7] Again, interpreting α_{i0} as a distortion in the firm's discount rate is isomorphic to interpreting it as a distortion in the firm's beliefs.

functional dependence on these states elsewhere in the model description.

4.1. Firm Technology

The firm generates an output flow rate AK^{η}, where K denotes the firm's capital stock, and incurs a proportional cost of production at rate $c_f K$. The capital stock is affected by firm investment I_+, disinvestment I_-, and depreciation shocks. Characterizing cross-sectional firm dynamics along dimensions such as investment and valuation ratios is essential for assessing the influence of cross-sectional mispricings on real economic quantities. We propose a novel specification of the investment technology that yields closed-form solutions for conditional and stationary distributions of all quantities of interest, allowing us to side-step simulations when estimating the model. Specifically, capital K takes values in a discrete set indexed by $\kappa \in \Omega_\kappa = \{1, 2, ..., N_\kappa\}$, where K is given by:

$$K(\kappa) = K_l e^{(\kappa-1)\cdot\Delta}. \tag{10}$$

By choosing Δ small enough the model can approximate a model with a continuous support for K arbitrarily well. The discrete state space structure, however, increases the tractability of the model and allows obtaining exact solutions.[8]

Firms can search for opportunities to upgrade their capital stock. Each firm chooses its expected investment rate $i_+ \equiv \frac{E[I_+]}{K}$ and stochastically succeeds in upgrading its capital K to the next-higher level, that is, by an amount:

$$I_+ = Ke^{\Delta} - K, \tag{11}$$

with a Poisson arrival rate $\frac{i_+}{e^{\Delta}-1}$.[9] When choosing $i_+ \geq 0$ a firm incurs search and investment-related cost that are quadratic in this expected investment rate: $(c_1 i_+ + c_2 i_+^2)K$. When a firm reaches the upper bound of the capital stock support, $K(N_\kappa)$, investment is assumed to be ineffective in generating further increases in capital. By choosing N_κ high enough, this boundary will have no effects on the results, as optimal investment will be zero above some endogenous threshold for capital.

[8]Models with a continuous support are in any case approximated by a discrete state space model when solved numerically.

[9]Thus, the expected rate at which capital grows due to investment is given by $(e^{\Delta} - 1) \cdot \frac{i_+}{e^{\Delta}-1} = i_+$.

Firms can also search for opportunities to sell their capital in order to disinvest. Choosing an expected disinvestment rate $i_- \equiv \frac{E[I_-]}{K} \geq 0$ leads to disinvestment by an amount $I_- = K - Ke^{-\Delta}$ with a Poisson arrival rate $\frac{i_-}{1-e^{-\Delta}}$. Search is costly, leading to a quadratic search cost of $c_{2-} i_-^2 K$. The expected revenue flow rate from capital sales is $c_{1-} i_- K$. We assume that disinvestment is impossible when capital reaches the lower bound K_l. Again, by choosing K_l low enough we can ensure that the firm would never optimally attempt to disinvest at the lower bound in any case, so that this restriction is also non-binding.

Capital also depreciates stochastically to the next-lower level, that is, from K to $Ke^{-\Delta}$, with a Poisson intensity $\frac{\delta}{1-e^{-\Delta}}$, except at the lower boundary K_l. Thus, the expected depreciation rate is δ, except in the lowest capital state, where it is zero.

Let N^+ and N^- denote counting processes that keep track of successful capital acquisitions and sales, and let N^δ denote a counting process that keeps track of depreciation shocks. Capital evolves according to a jump process:[10]

$$d\log(K_t) = \Delta(dN_t^+ - dN_t^- - dN_t^\delta), \tag{12}$$

where expected changes in capital are given by $\mathbb{E}[dK_t] = (i_+ - i_- - \delta)K_t dt$.

4.2. Exogenous State Variables

There are three types of stochastic processes that govern the structural parameters of the economy: a firm-specific state z, a mean-reverting aggregate state Z, and an aggregate trend factor Y.

Firm-specific states (z). The state z governs cross-sectional properties of structural parameters, such as mispricing α, depreciation δ, and total factor productivity A. The dynamics of z thus affect key endogenous objects, such as the firm-size distribution, idiosyncratic risk, exposures to aggregate risk, and growth. We assume that z follows a continuous time Markov chain that takes values in the discrete set Ω_z. Let $\Lambda_z(Z)$ denote the generator matrix that collects transition rates between firm-states z conditional on the aggregate state Z. Dependence on the macro state Z allows capturing dependencies between cross-sectional dynamics and macro-economic conditions.

[10] In the following, all processes will be right continuous with left limits. Given a process y_t, the notation y_{t-} will denote $\lim_{s\uparrow t} y_s$, whereas y_t denotes $\lim_{s\downarrow t} y_s$.

Macroeconomic state (Z). The state Z captures the mean-reverting component of the macro economic environment (e.g., booms vs. recessions). We assume that Z follows a continuous time Markov chain that takes values in the discrete set Ω_Z. Let Λ_Z denote the generator matrix that collects transition rates $\lambda(Z, Z')$, and let $\Lambda_Z(Z)$ denote the Z-th row of this generator matrix.

Aggregate trend (Y). The state Y captures an aggregate trend that follows a geometric Brownian motion:

$$\frac{dY_t}{Y_t} = \mu(Z_t)dt + \sigma(Z_t)dB_t. \tag{13}$$

The variable Y can capture a macro trend growth. Y is assumed to enter both a firm's cost function and its output linearly, that is, the cost function parameters and the TFP variable A all scale linearly with Y. The trend reflects a gradual increase in output and the price of capital.

4.3. Market Valuations

Let s denote the vector of state variables. The market values a stochastic stream of future after-tax net-payouts of a firm $\{\pi(s_\tau)\}_t^\infty$ as follows:

$$\mathbb{E}_t \int_t^\infty \frac{m(s_\tau)}{m(s_t)} e^{-\int_t^\tau \alpha(s_k)dk} \pi(s_\tau)d\tau, \tag{14}$$

where m represents the *undistorted* stochastic discount factor (SDF) that corresponds to the relevant marginal utility process of a representative household, $\mathbb{E}$ represents an unbiased rational Bayesian expectation that incorporates all public information, and α captures price distortions further discussed below. The pricing equation (14) is flexible enough to capture multiple forces that could lead to informationally inefficient market prices. For the purpose of the exposition, we will refer to deviations from perfect rational expectations that incorporate all public information, which may for example arise if agents incur cost when processing public information, or are subject to rational inattention (Sims, 2003).

Under this narrative, we presume that agents have homogenous beliefs. The economy is arbitrage-free under these beliefs. Equation (14) can capture at least two types of distortions in expectations: (1) imperfect expectations about future state-contingent firm cash flows, or (2) imperfect expectations about future state-contingent marginal utilities

(SDF). For the first type of distortion, α can be specified as a function of both aggregate and firm-specific elements of the state vector s. In this case, the date-t market price of an Arrow-Debreu security paying at date τ in state s_τ is given by

$$q(s_\tau|s_t) = \Pr[s_\tau|s_t]\frac{m(s_\tau)}{m(s_t)}, \tag{15}$$

where $\Pr[s_\tau|s_t]$ is a state probability corresponding to the perfect Bayesian expectation operator mentioned above. Agents' expectations are, however, not informationally efficient in that agents believe that the firm's payout at time τ in state s_τ is:

$$\pi^{dis}(s_\tau|s_t) = \pi(s_\tau) \cdot \mathbb{E}[e^{-\int_t^\tau \alpha(s_k)dk}|s_\tau, s_t], \tag{16}$$

even though after perfectly processing all public information they should conclude that it is $\pi(s_\tau)$. Equation (16) implies that when approaching date τ and state s_τ agents' expectations converge to $\pi(s_\tau)$. As the distortion can vary across aggregate states, it generally affects the perceived exposures of a firm's cash flows to aggregate risks priced by the SDF, thus also capturing distortions in beta estimates. For example, if agents were excessively pessimistic about a firm's performance in a potential future recession, this would be captured by a persistent positive alpha that leads to excessive discounting of the firm's future payouts in those states. For the second type of distortion, α is specified a function of only the aggregate elements of the state vector s. In this case, (14) can be interpreted as capturing distortions in expectations about the marginal utilities across aggregate states. In particular, the date-t market price of an Arrow-Debreu security paying at date τ in state s_τ is given by:

$$q^{dis}(s_\tau|s_t) = \Pr[s_\tau|s_t]\frac{m(s_\tau)}{m(s_t)}\mathbb{E}[e^{-\int_t^\tau \alpha(s_k)dk}|s_\tau, s_t]. \tag{17}$$

We consider a partial equilibrium analysis in the sense that we quantify efficiency losses, taking as given a particular SDF. This partial equilibrium approach is motivated by two observations: first, due to data limitations we analyze only publicly traded firms, thus missing a significant part of output that would have to feature in a general equilibrium analysis. Second, while in general equilibrium the SDF would change under the counterfactual of efficient financial markets (since output and consumption change), we know that risk free rates tend to be not volatile and thus respond fairly little to increases in consumption growth,[11] suggesting that this indirect channel is of second-order

[11] Further, assuming that risk premia are unaffected would be conservative, if the undistorted risk

importance for our analysis.

We consider a flexible Markov-modulated jump diffusion process to describe the dynamics of m:

$$\frac{dm_t}{m_{t-}} = -r_f\left(Z_{t-}\right)dt - \nu(Z_{t-})dB_t + \sum_{Z' \neq Z_{t-}} (e^{\phi(Z_{t-},Z')} - 1)dM_t\left(Z_{t-}, Z'\right). \tag{18}$$

Here r_f denotes the risk free rate, ν is the price of risk for aggregate Brownian shocks, $\phi(Z, Z')$ is a jump risk premium, and $dM(Z, Z')$ is a compensated Poisson process capturing switches between the macroeconomic Markov states Z and Z'.[12] Let $\bar{\Lambda}_Z$ denote the generator under the risk neutral measure, that is, $\bar{\lambda}(Z, Z') = e^{\phi(Z,Z')}\lambda(Z, Z')$.[13]

4.4. Firm Objective

Firms take their market prices as given and choose the investment strategy that maximizes this market value at any point in time. Firms know how the market values all their potential investment plans. The view that firm managers maximize the firm's market value appears as a plausible benchmark for several reasons: first, it is not clear that firm managers know better than market participants how the market should value the firm's future cash flows — market participants might in fact have a better sense of what the firm's true risk exposures are and what the fair risk compensations (premia) should be. In addition, compensation contracts are often tightly linked to current market prices, creating incentives for managers to maximize the going market value at any point in time.

premia fell in the boom following the elimination of cross-sectional anomalies and improved capital allocation.

[12]Formally, $dM(Z, Z') = dN\left(Z, Z'\right) - \lambda(Z, Z')dt$, where $N(Z, Z')$ is a counting process that keeps track of the jumps from Markov state Z to state Z'.

[13]With this pricing kernel exposures to innovations in the state variables Y and Z can bear a risk premium. The cross-sectional distribution of firm-specific states z does not additionally affect the process m. More generally, in a general equilibrium setting, the current distribution of firm-specific states z in the cross-section of firms could constitute another state variable. The proposed setup is in principle sufficiently flexible to capture this aspect as well: the set Ω_Z can be defined in a way that allows the state Z to summarize the current state of the cross-sectional distribution of z as well.

5. Analysis

5.1. Firm Behavior

In this subsection we analyze firm behavior in the presence of pricing distortions. Let the firm's value function be denoted by $V(\kappa, z, Z, Y)$, where

$$V(\kappa, z, Z, Y) = \max_{\{i_+, i_-\} \geq 0} \mathbb{E}_t \int_t^{\infty} \frac{m_\tau}{m_t} e^{-\int_t^\tau \alpha_s ds} \pi(\tau) d\tau, \tag{19}$$

and where we define the conditional expected after-tax net-payout:

$$\begin{aligned}\pi(t) =& (1-\tau)(AK_t^\eta - (c_f + c_{1+}i_+ + c_{2+}i_+^2 - c_{1-}i_- + c_{2-}i_-^2)K_t) \\ &+ (i_- + \delta - i_+)c_{1+}\tau K_t,\end{aligned} \tag{20}$$

where we assume that firms obtain tax shields from depreciation, proportional cost of production, quadratic search cost, and selling capital below the purchase price c_{1+}.

Since the parameters A, c_f, c_{1+}, c_{2+}, c_{1-}, and c_{2-} are assumed to be linear in the trend component Y, we can conjecture that the value function is linear in Y, that is, $V(\kappa, z, Z, Y) = Y \cdot \tilde{V}(\kappa, z, Z)$, where going forward, a tilde indicates that a variable is scaled by Y. The Hamilton-Jacobi-Bellman equation associated with problem (19) implies that $\tilde{V}(\kappa, z, Z)$ solves the following set of equations for all $(\kappa, z, Z) \in \Omega_\kappa \times \Omega_z \times \Omega_Z$:[14]

$$\begin{aligned}0 = \max_{i_+, i_- \geq 0} [&\tilde{\pi}(\kappa, z, Z) - (r_f(Z) + \sigma(Z)\nu(Z) + \alpha(z, Z) - \mu(Z))\tilde{V}(\kappa, z, Z) \\ &+ \frac{i_+}{(e^\Delta - 1)}(\tilde{V}(\kappa+1, z, Z) - \tilde{V}(\kappa, z, Z)) \\ &+ \frac{\delta + i_-}{(1 - e^{-\Delta})}(\tilde{V}(\kappa-1, z, Z) - \tilde{V}(\kappa, z, Z)) \\ &+ \bar{\Lambda}_Z(Z)\tilde{\mathbf{V}}^Z(z, \kappa) + \Lambda_z(Z)\tilde{\mathbf{V}}^z(Z, \kappa)]\end{aligned} \tag{21}$$

where $\mathbf{V}^Z$ and $\mathbf{V}^z$ are vectors that collect the values of the function V evaluated at all possible elements in the sets Ω_Z and Ω_z, respectively.

The first-order conditions of this problem yield a firm's optimal expected investment

[14] See Appendix B for details.

and disinvestment rates:

$$i_{+}^{*}(\kappa,z,Z)=\max\left[\frac{\left(\frac{\tilde{V}(\kappa+1,z,Z)-\tilde{V}(\kappa,z,Z)}{(e^{\Delta}-1)K(\kappa)}-c_{1+}(z,Z)\right)}{2(1-\tau)c_{2+}(z,Z)},0\right], \tag{22}$$

$$i_{-}^{*}(\kappa,z,Z)=\max\left[\frac{\left(\frac{\tilde{V}(\kappa-1,z,Z)-\tilde{V}(\kappa,z,Z)}{(1-e^{-\Delta})K(\kappa)}+(1-\tau)c_{1-}(z,Z)+\tau c_{1+}(z,Z)\right)}{2(1-\tau)c_{2-}(z,Z)},0\right]. \tag{23}$$

Note that conditional on these policy functions the system (21) is linear in $\tilde{V}(\kappa,z,Z)$, which will make determining an exact solution easy and fast.

Risk premium. The firm's risk premium under rational beliefs is given by:

$$rp(\kappa,z,Z)=\sigma(Z)\nu(Z)-\sum_{Z'\neq Z}\lambda\left(Z,Z'\right)\left(e^{\phi(Z,Z')}-1\right)\left(\frac{\tilde{V}\left(\kappa,z,Z'\right)}{\tilde{V}\left(\kappa,z,Z\right)}-1\right).$$

The risk premium features compensation for exposures to both innovations to the Markov state Z and Brownian innovations to the common trend Y. In the model both betas and risk prices are state-dependent.

5.2. Stationary Distribution

To measure efficiency losses it is essential to capture the stationary cross-sectional distribution of firm characteristics such as size and the Book-to-Market ratio. We show in Appendix C how, for any given policy function, we can compute the stationary distribution in closed-form, which greatly facilitates the estimation and evaluation of the model.

6. Estimating the Model

6.1. Specification of the Markov Processes

We consider a parsimonious specification of the model with three *independent* firm-specific processes: a process for productivity A, an α-process, and a technology process jointly governing depreciation δ and operating cost c_f. Corresponding to these three processes the firm-specific state z can be characterized by a tuple $(\tilde{A},\alpha,g)$, where g

denotes the technology state. In addition, the macro-state Z governs trend growth μ, trend volatility σ, as well as risk prices.

6.2. Calibration and Estimation

We calibrate the parameters of the macroeconomy based on the existing literature (e.g., Chen, Cui, He, and Milbradt, 2015) and estimate firm-specific parameters using a method of moments approach. We estimate 22 parameters by targeting 32 moments related to the cross-sectional distribution of firms: 9 book to market decile breakpoints, 6 book asset breakpoints, 7 book asset growth percentiles, and 10 CAPM alphas corresponding to the book-to-market deciles.[15]

The calibrated and estimated parameters are summarized in Table 3 and all fall in a range that is broadly consistent with the literature. The depreciation rates for the two technology states are 12% and 14% close to the standard values used in the literature. The proportional costs of production are 1% and 27% for the two technology states. The states themselves are driven by a persistent Markov process (conditional on being in a particular state, you are expected to stay there for roughly 10 years). The decreasing returns to scale parameter takes on a value of 0.95, which implies a technology close to an AK technology.[16] The mispricing variable (alpha), can take three values: -16.6%, 0% and 13.5%, and the Markov switching intensities vary by state. For example, if a firm is currently in a state $\alpha_3 = 13.5\%$, its expected abnormal return over the next year is significantly lower than 13.5%, due to strong mean reversion in the alpha process — the expected time until the alpha reverts back to zero is just 4 to 5 months. The discount on selling used capital $(1 - c_{1-})$ equals 30%. Finally, there are very high quadratic search

[15]Since we observe only publicly traded firms we also account for delistings that hit firms with Poisson arrival rates that are calibrated to the data as a function of a firm's sales-to-assets ratio. Specifically, to match delisting rates in the data we estimate historical (average) exit rates in each sales-to-book decile. A firm's exogenous delisting rate is then determined via interpolation as a function of its sales-to-book ratio. A firm that delists from public equity markets (e.g., because of an M&A transaction, a private equity deal, or a default that transfers assets to debt holders) is assumed to continue its operations, following the same policies as it would as a publicly traded firm. As a result, a delisting event by itself does not increase or destroy value, and the possibility of a delisting does not affect the firm's maximization problem analyzed in Section 4.4. Yet delistings do affect the distribution of various firm outcomes conditional on staying publicly traded. For example, delistings affect the distribution of annual book capital changes when the sample is restricted to firms that are publicly traded in years t and $(t+1)$. Finally, we presume that, in any state of the world, new firms enter the publicly traded universe at the same rate as existing firms delist.

[16]As a robustness we have also run our estimation setting this parameter to 0.65, leading to similar implications regarding welfare losses induced by financial market distortions.

costs for finding buyers for used capital. In contrast, upgrading capital is less subject to frictions.

The fitted moments generated by the parameters listed in Table 3 are summarized in Figure II. The figure summarizes the distributions of the moments in the data (the black solid line), the 95% confidence bounds generated from the data (the black dotted lines) as well as the model-implied distribution (red dashed line). The figure shows that the model has a reasonably good fit of the data moments when it comes to the firm-size distribution (the top panel), the book-to-market ratio distribution (the second panel), the investment (change in book value) distribution (the third panel), as well as the value premium alpha distribution (bottom panel). The model moments all fall within the 95% confidence bounds, with the exception of the extreme book value percentiles and lowest investment percentiles.

To interpret our results, it is important to note that in the model Tobin's q emerges as a *noisy* measure of a firm's underlying alpha state. Much of the variation in Tobin's q is still due to technology shocks, which are specified as independent of α.

Parameters of the Macroeconomy (Calibrated)			
Parameter	Variable	$Z = G$	$Z = B$
Transition rates for aggregate states	λ	0.100	0.500
Trend growth	μ	0.030	-0.010
Trend risk exposure	σ	0.160	0.160
Risk-free rate	r_f	0.020	0.020
Local risk price	ν	0.165	0.255
Jump in m upon leaving state Z	$e^{\phi} - 1$	1.000	-0.500
Tax rate (personal + corporate)	τ	0.450	

Constant Firm-specific Parameters		
Parameter	Variable	Estimated Values
Rate of moving to next-higher $\tilde{A}$	h_{A+}	3.476
Rate of moving to next-lower $\tilde{A}$	h_{A-}	3.989
Purchase price of capital	c_{1+}	1.034
Upward adjustment cost	c_{2+}	0.902
Sales price of capital	c_{1-}	0.708
Downward adjustment cost	c_{2-}	29.45
Decreasing returns to scale parameter	η	0.950

Firm-specific g-Process & Associated Technology Parameters			
Parameter	Variable	g_1	g_2
Rate of moving to next-higher g-state	h_{g+}	0.108	-
Rate of moving to next-lower g-state	h_{g-}	-	0.101
Depreciation rate	δ	0.144	0.120
Proportional cost of production	c_f	0.269	0.010

Firm-specific α-Process				
Parameter	Variable	α_1	α_2	α_3
Rate of moving to next-higher α-state	$h_{\alpha+}$	1.013	1.866	-
Rate of moving to next-lower α-state	$h_{\alpha-}$	-	0.155	2.699
Abnormal return	α	-0.166	0	0.135

Table 3

Parameters. The two tables list parameters of the macroeconomy and firm-specific parameters. The parameters of the macroeconomy are calibrated. We estimate firm-specific parameters via a method of moments approach. The set of factor productivity states are given by $\tilde{A}_i = A_1 e^{\sum_{j<i} a_j}$ with $A_1 = 0.163$ and $a_j \in \{0, 0.093, 0.051, 0.010, 0.189, 0.367, 0.323, 0.278, 0.389, 0.500\}$, where we estimate only every second a_j-value and determine the remaining values via interpolation. The capital grid is characterized by the lower bound $K_l = 0.00239$ (the value is scaled so that the median firm's capital is 1), the number of capital grid points $N_\kappa = 160$, and the log-change in capital between grid points, $\Delta = 0.1$.

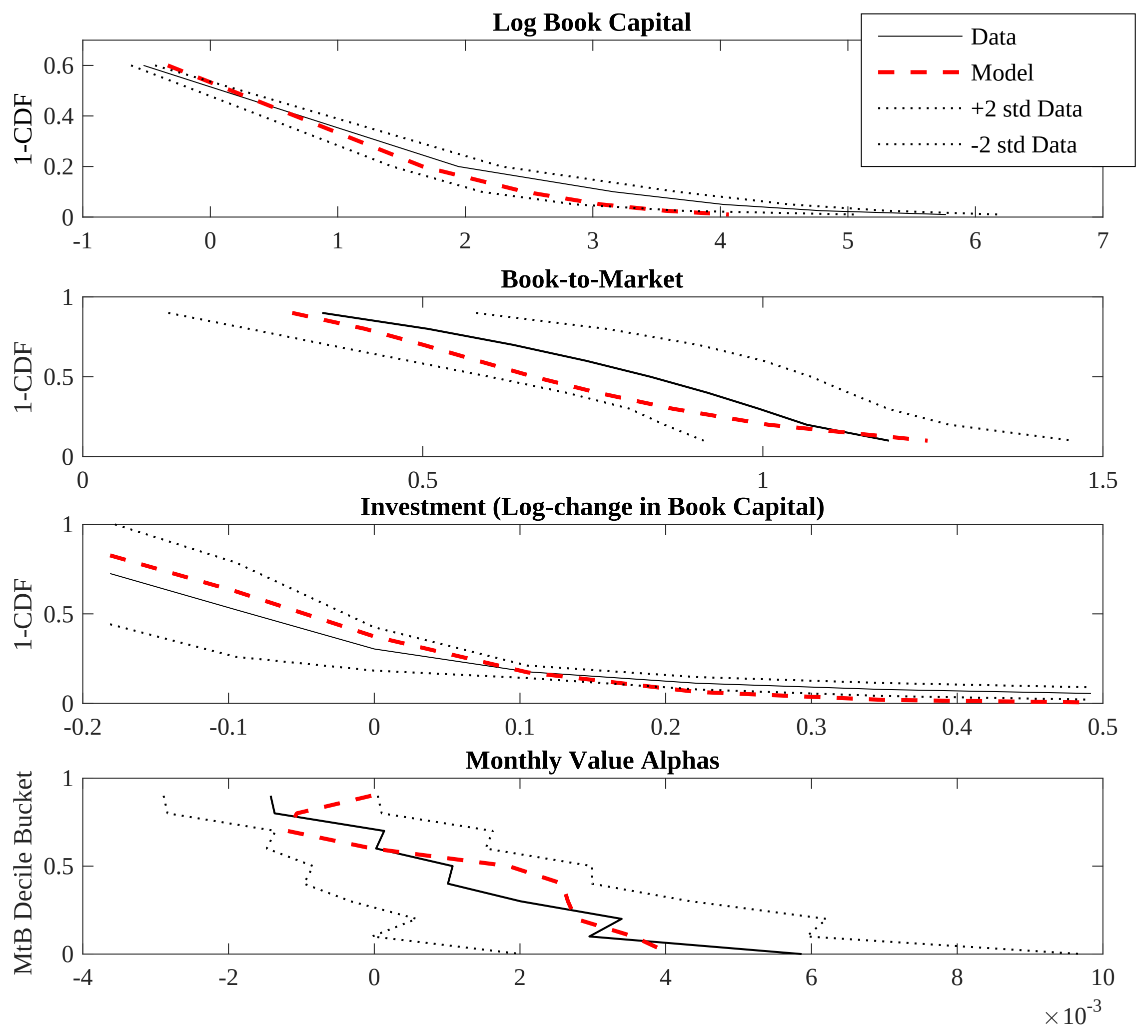

FIGURE II

Model Fit. This graph plots for each variable the model's values (dotted red line) and compares it with the data in black (including 2 standard error bounds).

7. Results

7.1. Under- and Overinvestment

First we assess the influence of cross-sectional distortions on investment. In Figure III we plot the probability distribution function (PDF) of the log ratio of actual over optimal investment (where defined). To purely focus on the cross-sectional effects, we demean the alpha process by its unconditional average before solving the equilibrium. The plot shows substantial deviations from the optimal investment policy induced by alpha, with both substantial over- and underinvestment.

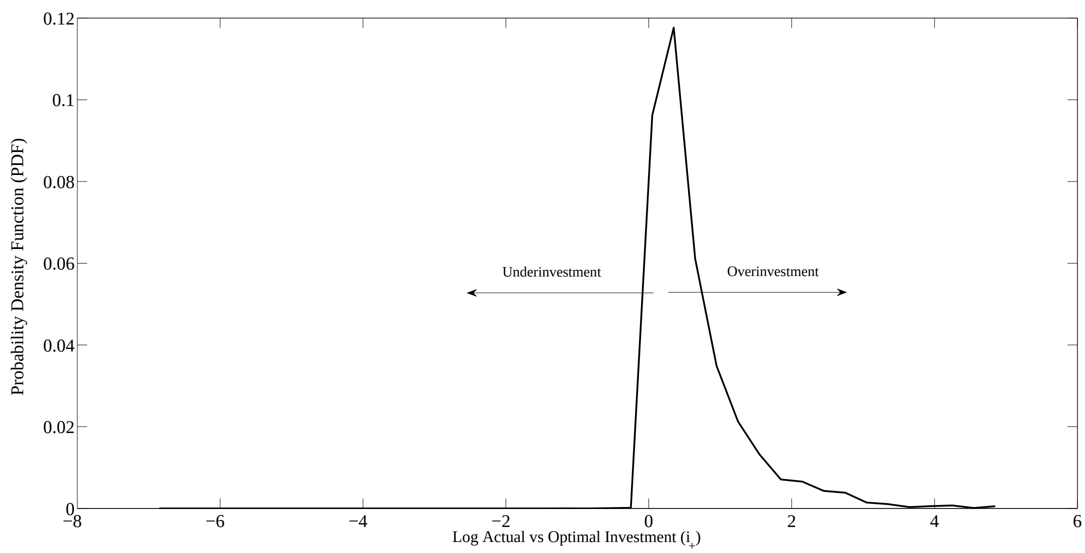

FIGURE III

Investment Distortions. The graph plots the probability distribution function (PDF) of the log ratio of actual over optimal investment (where defined). To purely focus on the cross-sectional effects, we demean the alpha process by its the unconditional average before solving the equilibrium.

Even though the plot shows that investment distortions can be large, it is not clear how important these effects are on aggregate value creation, which we explore in the next section.

7.2. Measuring Potential Efficiency Gains

To assess the influence of cross-sectional distortion on value, we first compute the stationary distribution of all states for the estimated model under the distorted policies. Let p^{act} denote the vector of probabilities for all states under this stationary distribution. Further, let $V^{act}(\alpha = 0)$ denote the corresponding vector of firm values if firms follow the actual (suboptimal) policies and prices are determined under the undistorted SDF $(\alpha = 0)$. The unconditional true value of the cross-section of all firms is then given by $p^{act} \cdot V^{act}(\alpha = 0)$. Finally, let $V^{opt}(\alpha = 0)$ denote the vector of firm values if firms do follow socially optimal policies and prices are determined under the undistorted SDF $(\alpha = 0)$.

If, starting from the actual stationary distribution of all states (in particular capital), firms switch from following suboptimal policies to following optimal policies the present value of surplus rises in expectation by:

$$gain = \frac{\mathbb{E}[\int \frac{m_\tau}{m_t} \pi_\tau^{opt} d\tau | p^{act}]}{\mathbb{E}[\int \frac{m_\tau}{m_t} \pi_\tau^{act} d\tau | p^{act}]} - 1 = \frac{p^{act} \cdot V^{opt}(\alpha = 0)}{p^{act} \cdot V^{act}(\alpha = 0)} - 1 = 10.6\% \tag{24}$$

As before, to purely focus on the cross-sectional effects, we demean the alpha process by its unconditional average before solving the equilibrium.

Interpreting the *gain* estimate. The *gain* estimate can be interpreted as society's willingness to pay as a perpetual percentage fee of total firm net payout for perpetually eliminating the alpha process under consideration. *gain* thus can be viewed as the magnitude of potential compensation of financial intermediaries, provided these intermediaries completely eliminate alpha. It is important to note that 10.6% of public firm net-payout is significantly less than 10.6% of GDP. Further, trading activity that simply reduces the value effect until it is statistically insignificant would not yield the full effect, as value-sorts are merely noisy measures of underlying alphas — variation in book-to-market ratios is largely due to technology shocks, which are *orthogonal* to the alpha shocks in our model. Finally, this estimate says nothing about the fair compensation that the financial sector should have received historically. Instead, it evaluates how large compensation could be if the financial sector eliminated existing informational inefficiencies in the future.[17]

[17] See, e.g., Philippon (2010), Philippon and Reshef (2012), and Philippon (2015) for papers analyzing financial sector compensation.

7.3. Value Gain and Tobin's q

In this subsection we assess whether the value gain from moving to optimal investment policies (i.e. removing alpha) differs for firms with different book-to-market ratios. Figure IV plots the value gain as a function of Tobin's q. That is, conditional on having a particular value of Tobin's q, the picture shows how much value can be gained. The graph shows that most of the value gain from removing alpha distortions (i.e. moving to optimal investment policies) is achieved for growth firms, not value firms.

This asymmetric result is directly related to the asymmetric adjustment cost in our estimated model. To see why, consider the case of firms with a Tobin's q lower than 1. Both in the data as well as in our model about 30% of firms have this characteristic, which illustrates that the costs for disinvesting are high. Such firms, if anything, would like to disinvest, as their capital could be used more efficiently outside of the firm. Yet, because the frictions to disinvesting are so large, they refrain from doing so. When the alpha distortion for such a firm is removed, the firm still does not have an incentive to invest, nor can it disinvest easily. As a consequence, the alpha distortion does not materially affect the firm's investment behavior. This is different for growth firms. Growth firms invest heavily and their investment rate is highly sensitive to their valuation, leading to large deviations from optimal investment, and thus value destruction. In light of these results the findings by Bai, Philippon, and Savov (2016) are particularly encouraging: according to their analysis, price informativeness has risen much more for growth firms than for value firms since the 1960s. Note that even though high Tobin's q firms are more likely to overinvest in our model, a significant fraction of them underinvests. This is because q is merely a noisy measure of alpha — much of the variation in q is determined by technology shocks that are independent of mispricing, implying that a high-q firm also has a significant chance of being undervalued.

In Figure V we plot the same relationship as in Figure IV with the difference that this graph incorporates the amount of value that is concentrated at each level of Tobin's q. It shows that most of the value gain in the economy can be generated by adjusting the investment policies of firms with a Tobin's q between 1 and 3, partly because there are so many firms that have values of Tobin's q in this region.

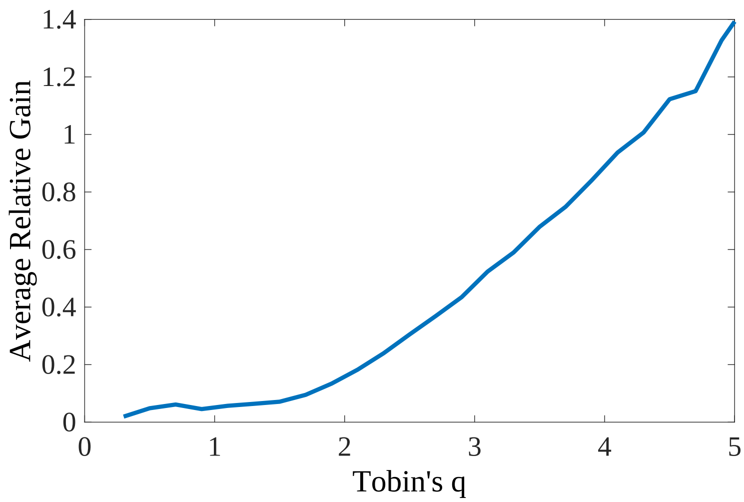

FIGURE IV

Value Gains vs Tobin's q. This graph plots the average value gain for an individual firm from moving to optimal investment policies as a function of Tobin's q (horizontal axis).

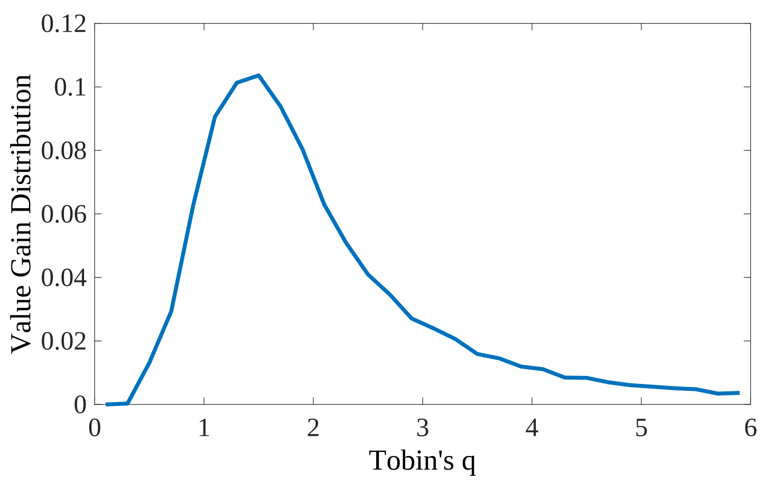

FIGURE V

Value Gains vs Tobin's q. This graph plots the value gain distribution from moving to the undistorted investment policies, as a function of Tobin's q (horizontal axis). The value gain is scaled by the total.

7.4. The Investment-q Relationship

As is well known in the investment literature, the relationship between Tobin's q and investment is generally weak. One may wonder to what extent we replicate this weak relationship in our model. When considering the exact relation between expected investment rates, $(i_+ - i_-)$ and q under the ergotic distribution we find an investment-q

slope of 0.07 with an R^2 value of 0.16. Further, simulations indicate that the slope drops about by half when considering realized investment rates over a one-year horizon rather than local expected investment rates, a difference that matters particularly due to the lumpy nature of investment in our model. These results indicate that our model features a relatively weak investment-q relationship, broadly in line with results established in the empirical literature (see, e.g., Peters and Taylor, 2016). This is important. It shows that we cannot conclude from weak investment-q regression results that firm managers are not responding to prices (and mispricing). After all, in our model, managers are by construction responding to market prices.

7.5. The Investment-α Relationship

We chose to estimate our alpha process by matching the empirical relation between Market-to-Book deciles and alphas. As can be seen from the lower panel of Figure II we fit the alphas generated by the book-to-market distribution quite well. We now evaluate whether the estimated alpha process also generates investment alphas that are consistent with the data. In Figure VI we plot the CAPM alpha generated by the model and compare it to the data. The graph illustrates that our model also generates investment alphas. These results suggest that Book-to-Market ratios and investment are both noisy measures of alpha. As illustrated in the previous section, Tobin's q (i.e. inverted Book-to-Market) and investment are in fact not very highly correlated, indicating that firms in corresponding Book-to-Market and investment deciles are not the same. Yet, the alpha process that we estimated does generate both anomalies in the model.

The empirical relation between alpha and firm investment is an important fact for our analysis, as it is consistent with the notion that managers adjust firm investment in response to the *abnormal* components of discount rates encoded in market prices. High investment predicts abnormally low returns, and low investment predicts abnormally high returns, consistent with the view that firms over- or underinvest when markets discount their cash flows too little or too much.[18] Put differently, it does not appear to be the case that managers decouple their investment policies from informational inefficiencies measured relative to the CAPM. Moreover, Hou, Xue, and Zhang (2016) show that investment alphas survive even when considering the Fama-French 3 factor model, the Carhart model, and the Pastor-Stambaugh model, highlighting the robustness of this

[18] Alphas associated with sorts on total asset growth are almost identical to alphas associated with the growth of total assets without cash, suggesting that this effect is not coming from managers that raise cash when they are overvalued.

finding with respect to various benchmark asset pricing models.

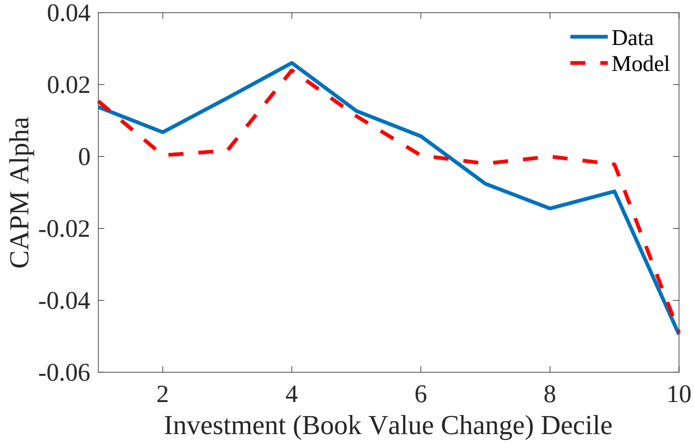

FIGURE VI

Investment Alphas. The graph plots the CAPM alpha of investment-sorted decile portfolios in the model and compares them to the data. As before, to purely focus on cross-sectional effects, both series are demeaned by their unconditional average.

It is also useful to note that, even when investment policies are distorted in the way our model posits, investment does not need to yield significant spreads in abnormal returns across decile portfolios. In alternative model parameterizations with a more persistent alpha process and a lower decreasing returns to scale parameter, we have found that investment sorts can become poor indicators of alphas, whereas profitability can become a more informative sorting variable. Since investment is affected by several state variables, cross-sectional sorts yield only noisy measures of alphas, even when investment responds to mispricing. The fact that the estimated model generates a similar alpha spread as observed in the data further suggests that our results do not overstate the responsiveness of firm investment to mispricings.

7.6. Sensitivity Analysis

Persistence of the Alpha process. In this section we consider the sensitivity to varying the persistence of the α-process. Changing the persistence of this Markov process allows us to gauge how important the persistence of an anomaly is for the aggregate value losses. Figure VII plots the change in value as computed in Equation 24 for different levels of persistence. The x-axis is the multiplier on the transition rates $(h_{\alpha+}, h_{\alpha-})$ of the baseline parameterization by a factor that ranges between 0.7 and 2. When the multiplier is 1, we obtain the baseline value loss of 10.6%. The graph shows that the value losses are highly sensitive to the persistence of the anomaly, thereby confirming the intuition that non-persistent anomalies, such as the momentum effect, are unlikely to have a large effect on value added. On the other hand, if anomalies are more persistent than the value premium effect, they can create very large real inefficiencies.

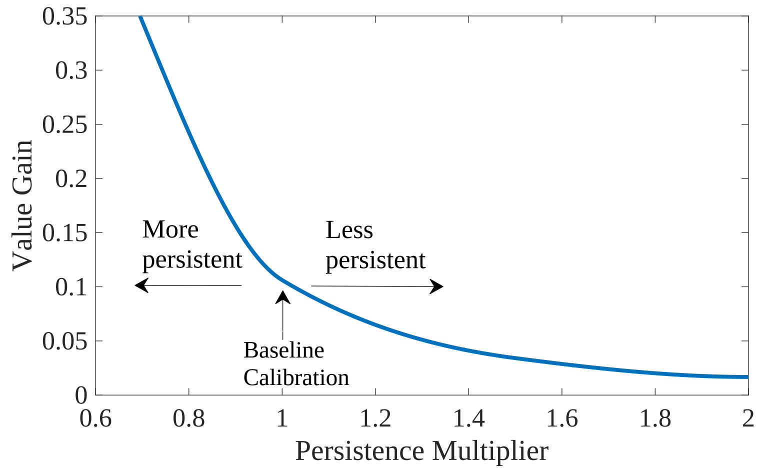

FIGURE VII

Changing α-state persistence. The graphs illustrate the effects of changing the persistence of the α-state by multiplying the transition rates $(h_{\alpha+}, h_{\alpha-})$ of the baseline parameterization by a factor $[0.7, 2]$. The first graph plots the present value of gains as measured by equation (24).

Debt mispricing. In our current analysis we have assumed that the debt-portion of the firm is equally mispriced as the equity portion. One may wonder whether our results would change if we assumed that the debt portion was less affected by mispricing. Our previous results suggest that value distortions are concentrated among high Tobin's q firms, which empirically also tend to have lower debt-to-value ratios as illustrated by Panel D of Table 1. This suggests that our results may not be particularly sensitive to assuming that debt is as mispriced as equity. We will address this issue explicitly in the next version of this paper by targeting equity mispricing alone.

8. Conclusion

Cross-sectional stock pricing anomalies are a widely studied topic. A large fraction of this literature exclusively focuses on the financial markets aspect of such anomalies, that is, the implications of these anomalies for investors and price informativeness. Another important fraction studies whether or not the first order conditions of the firm are consistent with the observed financial returns and argues that even if the first order conditions of consumers are not able to price assets appropriately, at least the investment behavior of corporate managers seems more consistent with the observed return patterns.

Instead, this paper quantitatively evaluates the potential real economic implications of the documented pricing effects assuming that these effects represent financial market imperfections. Taking as given that managers maximize the value of their firm as assessed by the market, we estimate the joint dynamic distribution of firm characteristics that have been linked to financial imperfections and other firm variables, such as investment, capital, and output. Based on a model that matches these joint dynamics we then evaluate the counterfactual dynamic distribution of capital, investment, and value added absent financial market imperfections and find that they can cause large and persistent deviations. This implies that financial intermediaries that can reduce and/or eliminate such market imperfections can provide large value added to the economy. As such, our paper contributes to the debate on the role and optimal size of the financial sector.

Even though we find that financial intermediaries can potentially add significant value to the economy by resolving anomalies, we do not show that they are currently engaged in that activity. In particular, we have shown that alphas are a poor measure of real inefficiencies, implying that simply chasing high alphas, for example with momentum strategies, may not be as important for real allocations. Further, it is unclear how large cross-sectional anomalies would be absent these financial intermediaries, even though time series changes and secular trends suggest that financial intermediation may have a material impact (McLean and Pontiff, 2016, Bai, Philippon, and Savov, 2016). Finally, there may also be large aggregate (as opposed to cross-sectional) mispricings, that would further enhance the value that financial intermediaries could add to the economy.

A. Markov Matrices

Book-to-Market Sorted Portfolios

From/To	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10
Dec 1	0.577	0.215	0.071	0.033	0.015	0.008	0.006	0.004	0.006	0.009
Dec 2	0.160	0.352	0.216	0.100	0.044	0.025	0.012	0.007	0.010	0.013
Dec 3	0.044	0.182	0.289	0.198	0.100	0.045	0.025	0.015	0.016	0.017
Dec 4	0.018	0.067	0.177	0.258	0.195	0.099	0.041	0.026	0.024	0.023
Dec 5	0.009	0.028	0.074	0.172	0.246	0.197	0.089	0.045	0.036	0.030
Dec 6	0.005	0.014	0.035	0.079	0.171	0.244	0.188	0.094	0.061	0.038
Dec 7	0.003	0.008	0.016	0.036	0.077	0.168	0.264	0.207	0.099	0.048
Dec 8	0.003	0.005	0.013	0.021	0.039	0.079	0.198	0.302	0.194	0.065
Dec 9	0.003	0.006	0.011	0.019	0.031	0.056	0.088	0.182	0.338	0.178
Dec 10	0.006	0.008	0.011	0.018	0.024	0.031	0.039	0.059	0.152	0.527

Profitability Sorted Portfolios

From/To	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10
Dec 1	0.657	0.118	0.039	0.023	0.014	0.007	0.005	0.004	0.004	0.005
Dec 2	0.112	0.613	0.148	0.026	0.008	0.005	0.002	0.001	0.001	0.002
Dec 3	0.041	0.130	0.538	0.161	0.033	0.011	0.006	0.003	0.003	0.002
Dec 4	0.024	0.025	0.139	0.456	0.188	0.049	0.020	0.010	0.006	0.005
Dec 5	0.015	0.009	0.032	0.164	0.400	0.201	0.062	0.023	0.010	0.006
Dec 6	0.008	0.004	0.016	0.052	0.175	0.376	0.202	0.060	0.020	0.008
Dec 7	0.005	0.002	0.008	0.022	0.059	0.183	0.375	0.201	0.055	0.015
Dec 8	0.004	0.002	0.004	0.012	0.028	0.064	0.180	0.405	0.193	0.035
Dec 9	0.003	0.001	0.003	0.007	0.014	0.028	0.061	0.180	0.470	0.158
Dec 10	0.003	0.001	0.002	0.007	0.009	0.013	0.022	0.045	0.165	0.664

Investment Sorted Portfolios

From/To	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10
Dec 1	0.275	0.138	0.075	0.055	0.040	0.037	0.038	0.039	0.046	0.085
Dec 2	0.156	0.179	0.135	0.095	0.072	0.060	0.054	0.051	0.045	0.048
Dec 3	0.084	0.135	0.163	0.137	0.103	0.085	0.070	0.056	0.047	0.041
Dec 4	0.057	0.098	0.130	0.151	0.129	0.107	0.084	0.070	0.057	0.043
Dec 5	0.047	0.073	0.102	0.132	0.153	0.133	0.109	0.078	0.068	0.045
Dec 6	0.044	0.063	0.085	0.103	0.131	0.150	0.130	0.101	0.076	0.053
Dec 7	0.040	0.062	0.072	0.085	0.107	0.132	0.150	0.131	0.099	0.064
Dec 8	0.044	0.056	0.059	0.073	0.092	0.105	0.136	0.160	0.128	0.086
Dec 9	0.056	0.058	0.061	0.063	0.068	0.080	0.099	0.140	0.185	0.132
Dec 10	0.098	0.073	0.063	0.058	0.055	0.058	0.068	0.091	0.141	0.227

Momentum Sorted Portfolios

From/To	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10
Dec 1	0.176	0.110	0.082	0.056	0.049	0.047	0.045	0.048	0.066	0.129
Dec 2	0.122	0.115	0.099	0.089	0.076	0.074	0.069	0.075	0.088	0.107
Dec 3	0.084	0.100	0.104	0.097	0.093	0.088	0.089	0.093	0.090	0.092
Dec 4	0.063	0.085	0.092	0.105	0.108	0.109	0.107	0.099	0.095	0.078
Dec 5	0.056	0.075	0.090	0.105	0.108	0.117	0.113	0.108	0.096	0.074
Dec 6	0.049	0.071	0.088	0.099	0.111	0.122	0.119	0.114	0.099	0.071
Dec 7	0.046	0.072	0.087	0.103	0.113	0.112	0.114	0.115	0.104	0.073
Dec 8	0.054	0.074	0.088	0.101	0.108	0.106	0.116	0.108	0.099	0.079
Dec 9	0.069	0.087	0.093	0.095	0.099	0.097	0.098	0.098	0.097	0.093
Dec 10	0.123	0.114	0.099	0.083	0.077	0.068	0.069	0.075	0.093	0.116

Table 4

Annual Markov Matrices of Decile Portfolios Sorted by Book-to-Market, Investment, Profitability and Momentum.

B. Proof: HJB Equation

The corresponding Hamilton-Jacobi-Bellman equation is given by:

$$\begin{aligned} 0 = \max_{i_+, i_- \geq 0} [&\pi(\kappa, z, Z, Y) - (r_f(Z) + \alpha(z, Z))V(\kappa, z, Z, Y) \\ &+ \frac{i_+}{(e^{\Delta} - 1)}(V(\kappa + 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\ &+ \frac{\delta + i_-}{(1 - e^{-\Delta})}(V(\kappa - 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\ &+ \Lambda_Z(Z)\mathbf{V}^Z(z, \kappa, Y) + \Lambda_z(Z)\mathbf{V}^z(Z, \kappa) \\ &+ V_A A\mu(Z) + \frac{1}{2}V_{AA}A^2\sigma(Z)^2 - V_A A\sigma(Z)\nu(Z)], \end{aligned} \tag{25}$$

where $\mathbf{V}^Z$ and $\mathbf{V}^z$ are vectors that collect the values of the function V evaluated at all possible elements in the set Ω_Z and Ω_z, respectively.

C. Stationary Distribution

Let m_s denote the mass of firms in state $s = (\kappa, z, Z)$ and let m denote the corresponding $N_s \times 1$ vector, where $N_s = N_\kappa \cdot N_Z \cdot N_z$. The vector that contains the fraction of firms in each state s evolves according to:

$$d\left(\frac{m}{\mathbf{1}'m}\right) = \frac{dm}{\mathbf{1}'m} - \frac{m}{\mathbf{1}'m}\frac{\mathbf{1}'dm}{\mathbf{1}'m}$$

We know that $\mathbb{E}[\frac{m}{\mathbf{1}'m}] = p$, where p is the vector unconditional probabilities $p_s = \Pr[s]$.

Stationarity implies that if we were to initialize the system with a vector $\tilde{m}$ such that $\frac{\tilde{m}}{\mathbf{1}'\tilde{m}} = p$ then there would be no expected change in the distribution, that is:

$$\mathbb{E}\left[d\left(\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\right)\right] = 0 \tag{26}$$

We can write:

$$\mathbb{E}\left[d\left(\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\right)\right] = \mathbb{E}\left[\frac{dm}{\mathbf{1}'\tilde{m}} - \frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\frac{\mathbf{1}'dm}{\mathbf{1}'\tilde{m}}\right] \tag{27}$$

$$= (\mathbf{I}_{N_s} - p\mathbf{1}')\frac{\mathbb{E}\left[dm\right]}{\mathbf{1}'\tilde{m}} \tag{28}$$

$$= (\mathbf{I}_{N_s} - p\mathbf{1}')\Lambda'\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}dt \tag{29}$$

$$= (\mathbf{I}_{N_s} - p\mathbf{1}')\Lambda' p dt \tag{30}$$

where $\mathbb{E}[dm] = \Lambda'\tilde{m}dt$ and where $\mathbf{I}_{N_s}$ is an identity matrix of size $N_s \times N_s$. Overall the vector of probabilities p thus solves:

$$(\Lambda' - \mathbf{I}_{N_s}\mathbf{1}'\Lambda' p)p = \mathbf{0} \tag{31}$$

The off-diagonal elements of the matrix Λ contain the (endogenous) rates $\Lambda(s, s')$ with which firms transition from state s to s'. The diagonal element s of the matrix Λ contains the sum of all flow rates of leaving state s and net-growth associated with entering and exiting the system, that is:

$$\Lambda(s, s) = -\sum_{s' \neq s}\Lambda(s, s') + h_{entry}(s) - h_{exit}(s).$$

The condition $\frac{\mathbf{1}'\mathbb{E}[dm]}{dt} = \mathbf{1}'\Lambda' m$ implies that, in expectation, there is no change in the total mass of firms. If we impose that expected growth in the total mass of firm is always zero, that is $\mathbf{1}'\Lambda' m = 0$ for all m, for example, if we set $h_{entry}(s) - h_{exit}(s) = 0$ for all s, then we obtain the stationary cross-sectional distribution of firms by solving the linear system:

$$\begin{pmatrix} \Lambda' \\ \mathbf{1}' \end{pmatrix} p = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}. \tag{32}$$

The vector of probabilities, p, is the left normalized eigenvector of Λ associated with the eigenvalue 0.

References

Asness, C., T. J. Moskowitz, and L. H. Pedersen (2013): "Value and Momentum Everywhere," *Journal of Finance*, 68(3), 929–985.

Bai, J., T. Philippon, and A. Savov (2016): "Have Financial Markets Become More Informative?," *Journal of Financial Economics*.

Baker, M., J. Stein, and J. Wurgler (2003): "When Does the Market Matter? Stock Prices and the Investment of Equity-Dependent Firms," *Quarterly Journal of Economics*, 118, 969–1005.

Barro, R. (1990): "The Stock Market and Investment," *Review of Financial Studies*, 3(1), 115–31.

Berk, J., and J. van Binsbergen (2014): "Assessing Asset Pricing Models Using Revealed Preference," Working Paper, The Wharton School.

——— (2015): "Measuring Skill in Mutual Fund Industry," *Journal of Financial Economics*, 118(1), 1–20.

Chen, H., R. Cui, Z. He, and K. Milbradt (2015): "Quantifying Liquidity and Default Risks of Corporate Bonds over the Business Cycle," Working Paper, MIT Sloan.

Chen, Q., I. Goldstein, and W. Jiang (2007): "Price Informativeness and Investment Sensitivity to Stock Price," *Review of Financial Studies*, 20, 619–650.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016): "Information, Misallocation and Aggregate Productivity*," *The Quarterly Journal of Economics*.

Dow, J., and G. Gorton (1997): "Stock Market Efficiency and Economic Efficiency: Is There a Connection?," *Journal of Finance*, 52, 1087–1129.

Edmans, A., I. Goldstein, and W. Jiang (2012): "The Real Effects of Financial Markets: The Impact of Prices on Takeovers," *The Journal of Finance*, 67(3), 933–971.

Eisfeldt, A. L., and A. A. Rampini (2006): "Capital reallocation and liquidity," *Journal of Monetary Economics*, 53(3), 369–399.

Fama, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25(2), 383–417.

Gilchrist, S., C. Himmelberg, and G. Huberman (2005): "Do Stock Price Bubbles Influence Corporate Investment?," *Journal of Monetary Economics*, 52, 805–827.

Goldstein, I., E. Ozdenoren, and K. Yuan (2013): "Trading frenzies and their impact on real investment," *Journal of Financial Economics*, 109, 566582.

Gomes, J. F., A. Yaron, and L. Zhang (2003): "Asset Prices and Business Cycles with Costly External Finance," *Review of Economic Dynamics*, 6(4), 767–788.

Hayek, F. (1945): "The Use of Knowledge in Society," *American Economic Review*, 35, 519–530.

Hennessy, C. A., and T. M. Whited (2007): "How Costly Is External Financing? Evidence from a Structural Estimation," *Journal of Finance*, 62(4), 1705–1745.

Hou, K., C. Xue, and L. Zhang (2015): "Digesting anomalies: An investment approach," *Review of Financial Studies*, 28(3), 650–705.

——— (2016): "A Comparison of New Factor Models," Working paper.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Jegadeesh, N., and S. Titman (1993): "Returns to Buying Winners and Losers: Implications for Stock Market Effiency," *Journal of Finance*, 48(1), 65–91.

Kiyotaki, N., and J. Moore (1997): "Credit Cycles," *Journal of Political Economy*, 105(2), 211–48.

Leland, H. (1992): "Insider Trading: Should It Be Prohibited?," *Journal of Political Economy*, 100(4), 859–87.

McLean, R. D., and J. Pontiff (2016): "Does Academic Research Destroy Stock Return Predictability?," *The Journal of Finance*, 71(1), 5–32.

Morck, R., A. Shleifer, and R. Vishny (1990): "The Stock Market and Investment: Is the Market a Sideshow?," *Brookings Papers on Economic Activity*, 2.

Peters, R. H., and L. A. Taylor (2016): "Intangible Capital and the Investment-q Relation," *Journal of Financial Economics*.

Philippon, T. (2010): "Financiers versus Engineers: Should the Financial Sector Be Taxed or Subsidized?," *American Economic Journal: Macroeconomics*, 2(3), 158–82.

——— (2015): "Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation," *American Economic Review*, 105(4), 1408–38.

Philippon, T., and A. Reshef (2012): "Wages and Human Capital in the U.S. Finance Industry: 1909–2006," *The Quarterly Journal of Economics*, 127(4), 1551–1609.

Polk, C., and P. Sapienza (2009): "The Stock Market and Corporate Investment: A Test of Catering Theory," *Review of Financial Studies*, 22, 187–217.

Rosenberg, B., K. Reid, and R. Lanstein (1985): "Persuasive evidence of market inefficiency," *Journal of Portfolio Management*, 11, 9–16.

Shiller, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71, 421–436.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–690.

Subrahmanyam, A., and S. Titman (2001): "Feedback from Stock Prices to Cash Flows," *Journal of Finance*, 56(6), 2389–2413.

Warusawitharana, M., and T. Whited (2016): "Equity Market Misvaluation, Investment, and Finance," *Review of Financial Studies*, forthcoming.

Whited, T. M. (1992): " Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data," *Journal of Finance*, 47(4), 1425–60.