NBER WORKING PAPER SERIES

REAL ANOMALIES

Jules H. van Binsbergen
Christian C. Opp

Working Paper 23238
http://www.nber.org/papers/w23238

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

We thank Andy Abel, Malcolm Baker (discussant), Markus Brunnermeier, Vincent Glode, João Gomes, Urban Jermann, Sydney Ludvigson, Stefan Nagel, Stijn van Nieuwerburgh, Dimitris Papanikolaou (discussant), Thomas Philippon (discussant), Rob Stambaugh, Jeremy Stein, Tony Whited, Amir Yaron, and seminar participants at the 7th Advances in Macro-Finance Tepper-LAEF Conference, Berkeley Haas, London Business School, Michigan Ross, NBER Summer Institute Asset Pricing, University of Oxford, Princeton University, and the Wharton Conference on Liquidity and Financial Crises for their helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Real Anomalies
Jules H. van Binsbergen and Christian C. Opp
NBER Working Paper No. 23238
March 2017
JEL No. D22,D24,D53,D92,E22,G2,G30

ABSTRACT

We examine the importance of asset pricing anomalies (alphas) for the real economy. We develop a novel quantitative model with lumpy investment that features such informational inefficiencies and yields closed-form solutions for cross-sectional distributions of firm dynamics. Our findings indicate that anomalies can cause material real inefficiencies, raising the possibility that agents that help eliminate them can provide significant value added to the economy. The framework reveals that alphas alone are poor indicators of real distortions, and that efficiency losses depend on the persistence of alphas, the amount of mispriced capital, and the Tobin's q of firms affected.

Jules H. van Binsbergen
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
julesv@wharton.upenn.edu

Christian C. Opp
The Wharton School
University of Pennsylvania
3620 Locust Walk, 2328 SH-DH
Philadelphia, PA 19104-6367
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 in 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, suggesting the presence of relative mispricings.[1] One important question that follows from these empirical findings is how large the real efficiency losses would be if these patterns in financial market returns were indeed reflective of informational inefficiencies. In this paper we address this question and find that the economy-wide real distortions can be large.

Our focus on real investment distortions connects our study to the literature in macroeconomics quantifying efficiency losses due to resource misallocations (see, e.g., Hsieh and Klenow, 2009).[2] Relative to this literature our study focuses on a novel friction — cross-sectional mispricings of financial assets — that is disciplined by direct measures (alphas) estimated in the existing finance literature. We propose a tractable dynamic framework that incorporates these alphas and evaluate the aggregate implications for the real economy.

An initial observation indicating the potential real impact of mispricings is the fact that cross-sectional variation in investment (asset growth) is related to future abnormal stock returns, what has been termed the investment anomaly.[3] One channel that could explain a relation between asset growth and mispricing is overvalued firms' opportunity to raise cash cheaply without investing in new capital. However, given that the relation between asset growth and alpha remains almost identical once cash holdings are excluded from assets, this

[1] Examples include anomalies based on Tobin's q (the value premium), investment, profitability, and past return performance (momentum). See, e.g., Rosenberg, Reid, and Lanstein (1985) for the value premium 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). For the profitability anomaly see Ball and Brown (1968), Bernard and Thomas (1990), and Novy-Marx (2013). For the investment anomaly see Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004), and Cooper, Gulen, and Schill (2008).

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

[3] As discussed below, the investment anomaly is robust to a variety of commonly used empirical asset pricing models.

channel does not appear to be the first-order effect. Instead, the investment-alpha relation suggests that firms with inflated (deflated) prices — that will be corrected in the future — overinvest (underinvest) in physical capital today.

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 mispricings and other firm variables, such as investment and capital. We develop a novel quantitative model with lumpy investment that yields closed-form solutions for the joint cross-sectional distribution of firm dynamics (for any given policy function). In the model, decision makers use (dis)information encoded in market prices when making investment decisions (Hayek, 1945).

We target and successfully match more than 40 moments describing the cross-sectional distributions of firm size, Tobin's q, the relation between Tobin's q and alpha (the value premium), and investment. More importantly, we match several key moments that were not targeted in the estimation. Namely, we replicate the well-known weak relation between investment and Tobin's q, as well as an investment-alpha relation. We then evaluate the counterfactual distributions of the variables of interest absent anomalies, allowing us to assess the magnitude of real inefficiencies. Our results reveal that informational inefficiencies can be associated with significant real effects despite a weak investment-q relationship.

The informational efficiency of prices per se is not measurable without some postulated asset pricing model (Fama, 1970). Hence, the existing finance literature estimates alphas based on deviations of average returns from a benchmark asset pricing model. Rather than taking a stance on which asset pricing model is the correct one, we provide a flexible methodology that allows the assessment of real distortions from alphas computed under a variety of standard asset pricing models. We use empirical alphas relative to the CAPM to estimate our model, but other asset pricing models can be accommodated.

While empirical alphas indicate informational inefficiencies, we find that they are generally poor measures of the real 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 price distortions. As a consequence, it is essential to account for the persistence of alphas, which is captured by our dynamic model. For the aggregate economy it is worse for firms to have a small persistent alpha instead of a short-lived large alpha. Whereas the investment management literature generally primarily cares about the magnitude of an alpha regardless of its persistence, real corporate investment decisions are little affected by short-lived mispricings. 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. Thirdly, and most importantly, it is not clear from studying alphas to what extent mispricings translate into real investment distortions and surplus losses. We show that firms with high Tobin's q are those that respond more to mispricings than firms with low Tobin's q and thus incur larger efficiency losses. Because firms with low Tobin's q wish to disinvest but face significant frictions when doing so (e.g., due to partly irreversible investment) mispricings have little effect on those firms' real decisions.

There is a large literature in macroeconomics and corporate finance studying external financing frictions that drive wedges between internal and external funds (Whited (1992), Kiyotaki and Moore (1997), Gomes, Yaron, and Zhang (2003), and Hennessy and Whited (2007)). For example, these wedges take the form of leverage constraints and issuance costs 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.

In contrast, the informational inefficiencies we study are measured with respect to *public* information. By definition, insiders and outsiders have symmetric access to this type of information. Cross-sectional anomalies are generally detected based on cross-sectional rankings of accounting and market data, such as Market-to-Book ratios (Tobin's q). Professional investors with access to large data bases can at least as easily determine whether a certain firm currently ranks in the 50th or 80th percentile of the cross-sectional distribution according to such a criterion as the firm itself. On the other hand, firms can obtain guidance on how to interpret market trends (risk premia etc.) from investment banks. As it is not obvious whether insiders or outsiders are better in processing this type of *public* information relevant for prices, we do not impose that either party is better at this task. The friction we study is therefore not driving a wedge between insiders and outsiders, but instead creates a wedge between the efficient and actual use of all public information, which is the very definition of an asset pricing anomaly. Yet similar predictions are obtained when managers can detect mispricings as long as they either have short horizons, are contractually incentivized to maximize current market values, or cannot raise material amounts of funds without investing them in actual capital (see, e.g., Stein, 1996). As argued above, the empirical finding that cash is not the main driver of the relation between asset growth and alpha indeed suggests that firms do not (fully) mitigate these distortions by merely adjusting cash holdings when they are mispriced.

To our knowledge we are the first to quantitatively assess the real value losses associated with cross-sectional financial market anomalies (alphas). Our paper relates to the important contributions of Baker, Stein, and Wurgler (2003), Gilchrist, Himmelberg, and Huberman (2005), and Warusawitharana and Whited (2016) who do evaluate the implications of an informational wedge between insiders and outsiders, where managers are better informed and therefore have different valuations. Baker, Stein, and Wurgler (2003) test the prediction in Stein (1996) that nonfundamental movements in stock prices have a stronger impact on the investment of equity-dependent firms, and find strong support for it.[4] Our findings are related to theirs in that our model also predicts that firms with high q and negative net payout, which the literature often uses as proxies for equity dependence, are those that respond most to mispricings.

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, effectively lowering the cost of capital and thereby increasing real investment. Using the variance of analysts' earnings forecasts to proxy for the dispersion of investor beliefs, the authors find that increases in dispersion are associated with increases in new equity issuance, Tobin's q, and real investment. Warusawitharana and Whited (2016) measure mispricings by the residuals of a hedonic regression and model a representative firm that (unlike investors) is informed

[4]Stein (1996) theoretically analyzes a firm's capital budgeting problem when market values are inefficient and explores how factors such as managerial time horizons and financial constraints affect the optimal hurdle rate.

about these mispricings. In contrast, we provide a novel methodology to estimate the aggregate real effects of misallocations associated with the cross-sectional asset pricing anomalies established by the finance literature.

The influence of market prices on real investment is already debated in an earlier influential literature. For example, Barro (1990) shows that changes in stock prices have substantial explanatory power for U.S. investment, especially for long-term samples, and in the presence of cash flow variables. The specification he employs outperforms standard Tobin's q regressions. On the other hand, Morck, Shleifer, and Vishny (1990) find that prices have little incremental explanatory power for firms' investment spending. Perhaps surprisingly, this debate does not affect the validity of our approach. For our research question, it is irrelevant if decision makers at firms learn about existing public information by using market prices as a summary statistic, or if they themselves process public information in the same (potentially inefficient) way as financial market participants do. A "real anomaly" only requires that real investment decisions end up maximizing an informationally inefficient market value. Even if firms independently process public information in the same flawed way as financial markets, the resulting investment distortion is the same. That said, financial markets can still be viewed as bearing responsibility for real inefficiencies — the *lack* of an efficient market price as a signal implies that agents' distorted beliefs are confirmed (rather than corrected) by prices.

In addition to the papers mentioned above, there are other empirical studies investigating the existence of a direct feedback effect of market prices on firm behavior. Chen, Goldstein, and Jiang (2007) show that two measures of the amount of new information in stock prices — price nonsynchronicity and the 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 financial markets have real effects by affecting managers' behavior.

In other work, 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 (i.e., firms with high Tobin's q). McLean and Pontiff (2016) further find that abnormal returns of anomaly strategies decline significantly after they are documented in academic publications. Their results suggest that the imperfect processing of public information is a relevant friction and that agents in the economy learn about mispricings from academic publications.

David, Hopenhayn, and Venkateswaran (2016) find that a firm's inability to predict its own productivity when making ex ante investment decisions causes sizable output losses. The authors estimate that firms learn more about their firm-specific productivity from private signals than from market prices. We are interested in a different question — the starting point for our analysis is the large literature documenting asset pricing anomalies. The counterfactual we evaluate is not a world where firms can predict their idiosyncratic productivity but one where expectations are informationally efficient with respect to public information.

An important starting point for our analysis is that data on prices in financial markets robustly indicate a cross-sectional relation between firms' real investment and standard measures of market mispricing, what has been termed the investment anomaly. For example, Hou, Xue, and Zhang (2015, 2016) show that sorting firms by their investment rates generates large spreads in alphas computed under commonly used benchmark asset pricing models. Our framework shows that significant losses may arise even when the investment-alpha relation is somewhat weaker than in the data. While our empirical application focuses on non-financial firms, similar empirical patterns apply for financial firms. Banks extend significantly more credit when they are overpriced relative to benchmark asset pricing models. Fahlenbrach, Prilmeier, and Stulz (2016) find that banks in the U.S. with the highest loan growth have predictably negative abnormal returns (alphas) going forward. Similar results hold at the country-level. By analyzing 20 developed countries between 1920 and 2012, Baron and Xiong (2016) find evidence that aggregate credit expansion is related to bank overpricing. Conditional on bank credit expansion of a country exceeding a 95th percentile threshold, the predicted excess return for the bank equity index in the subsequent three years is -37.3%.

Taken together, these empirical findings from various economic contexts suggest a robust relation between market mispricings and real economic activity, yet leaving open the question how large associated real inefficiency losses could be. This question motivates our investment model-based approach.

The paper proceeds as follows. In the next section, we introduce a simple one-period example to explain the main concepts underlying our analysis. In Section 3, we document reduced-form estimates of several moments related to our estimation. Section 4 introduces the model and solves it. Section 5 presents the estimation results. In Section 6 we quantify efficiency distortions by comparing the estimated model to a counterfactual economy without informational inefficiencies. Section 7 discusses the robustness of our analysis with respect to various modeling assumptions. Section 8 presents implications for future research. Section 9 concludes.

2. A Simple One-Period Example

To better understand why alpha measures by themselves are not informative about *real* economic inefficiencies, consider the following one-period example. Take a firm that only generates two cash flows, π_0 and π_1, at time 0 and time 1, respectively. The value that the market places on these cash flows at time 0 is a function of α_0 and given by:

$$V(\alpha_0) = \pi_0 + \frac{\mathbb{E}_0[\pi_1]}{1 + r_0 + \alpha_0}, \tag{1}$$

where r_0 is the appropriate discount rate and where α_0 is a price distortion. We can interpret α_0 as a distortion in either the discount rate or the expectation of the cash flow π_1. It is clear that for $\alpha_0 \neq 0$ the financial market is inefficient in that the market price of the firm V is affected by this distortion at time 0. However, as long as the firm's behavior and the corresponding cash flows are not affected by α_0, the misvaluation will resolve itself at time 1 through an abnormal excess return equal to α_0, and no real losses occur.

In this paper, we are interested in the distortions of real decisions. As firms maximize market value, real investment decisions are influenced by market prices at time 0, and thus, by α_0. Let the corresponding functional dependence between a firm's net payout and α_0 be denoted by $\pi_t(\alpha_0)$. To measure the present value of surplus losses, we compare the firm's true value under the distorted firm policies:

$$V^d = \pi_0(\alpha_0) + \frac{\mathbb{E}_0\left[\pi_1(\alpha_0)\right]}{1 + r_0}, \tag{2}$$

to the true value under the undistorted firm policies:

$$V^u = \pi_0(0) + \frac{\mathbb{E}_0\left[\pi_1(0)\right]}{1 + r_0}. \tag{3}$$

Note that in equations (2) and (3) we discount by r_0, the undistorted discount rate, not by $(r_0 + \alpha_0)$. In this one-period example mispricing was resolved within one period. In contrast, in reality mispricings resolve and build up dynamically over time, potentially affecting firms' long-term investment plans. We thus need a model that allows for such dynamic behavior.

Overall, the model we use for our quantitative analysis should fulfill three requirements. First, it should be dynamic. Second, it should be flexible and easy-to-solve. Third, conditional on firm policies we would like to obtain closed-form solutions such that estimation based on cross-sectional moments is computationally feasible. Before formally introducing such a model, we first explore several reduced-form characteristics of mispricings in the next section.

3. Reduced-Form Estimates

As highlighted in the previous section, the estimates of informational inefficiencies (alphas) the finance literature has documented are an important input for our model: alphas are distortions in the discount rate and/or expected growth rate that can affect firm investment. Interestingly, two of the most prominent anomalies are related to the two variables most often used in the investment literature. The anomaly related to Tobin's q has become known as the value anomaly and the anomaly related to asset growth as the investment anomaly. In this section, we start by replicating these two anomalies for the sample period 1975-2014 and find results that are consistent with the literature. What is different is that the existing literature treats the documented alphas as the end point of the analysis, whereas we use them as one of the inputs to our cross-sectional investment model. This allows us to evaluate the significance of associated real investment distortions.

Motivated by the model presented in the next section, we compute several moments that affect the importance of alphas for real distortions. First, what is the amount of capital that is affected by a particular alpha? If alphas are large for a large fraction of the market capitalization of firms, this will have larger aggregate investment implications compared to when only a small fraction is affected. Second, what is the persistence of the alpha? If the same firm is affected by alphas for a long period of time, this will lead to larger investment distortions compared to very transitory mispricings. In this section, we explore reduced-form estimates of these characteristics.

As standard in the literature, we sort firms into decile portfolios based on their Book-to-Market (BtM) ratio lagged by one month, and their investment as measured by asset growth over the past year. We form 10 value-weighted decile portfolios each month for both sorting variables. To compute alphas, we follow the literature and regress decile i's value-weighted excess return (denoted by $R_{i,t+1}$) minus the risk free rate ($R_{f,t}$) on the excess return of the market portfolio ($R_{m,t+1} - R_{f,t}$) as defined by all stocks traded on the New York Stock Exchange, the American Stock Exchange, and NASDAQ:

$$R_{i,t+1} - R_{f,t} = \alpha_i + \beta_i(R_{m,t+1} - R_{f,t}) + \varepsilon_{i,t+1}. \tag{4}$$

The results are summarized in Panels A and B of Table 1. The Panels confirm the findings in the literature that firms with high (low) Book-to-Market ratios and low (high) investment earn abnormally high (low) average returns relative to what the CAPM predicts. Our motivation to use the CAPM is that it is the benchmark asset pricing model most often used by both academics and practitioners. As mentioned in the introduction, our framework can easily accommodate alternative models. The so-called "value spread" is computed as the difference between the alpha in the 10th decile and the alpha in the 1st decile of the Book-to-Market sort, and equals 73bp per month, or about 9% per annum over this sample period.

As argued before, a downside of merely studying the alphas in Panel A is that they do

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
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
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
Panel D: Time Series Average of Decile's 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
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

Table 1

Anomalies. The table reports characteristics of the value and investment anomalies. We sort stocks into portfolios based on their lagged Book-to-Market ratio and their investment (percentage change in total assets). Panel A reports average monthly returns for each decile portfolio. Panel B reports monthly CAPM alphas. Panel C reports each decile's 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 number 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 each decile's diagonal element in the annual Markov transition matrix of decile assignments.

not properly account for size differences across deciles. To assess the importance of this issue, we compute the weights of each decile's market value of equity (E) as a fraction of the aggregate market value of equity. Define the weight of decile i for anomaly j as:

$$we_{i,j} = \frac{1}{T}\sum_{t=1}^{T}\frac{E_{i,j}}{\sum_{i=1}^{10} E_{i,j}}. \tag{5}$$

The weights $we_{i,j}$ are summarized in Panel C of Table 1. Interestingly, for both anomalies alphas tend to be larger in deciles with lower market capitalizations. That is, most of the market capitalization is concentrated around the middle deciles, and the extreme deciles are underrepresented in terms of the market value. Furthermore, even when comparing the two extreme portfolios (first versus tenth) do we see this pattern. The tenth decile portfolio of the Book-to-Market sort has an alpha of 59bp per month, which is much larger in absolute magnitude than the corresponding mispricing of the first decile (-14bp). Only 2% of the equity market capitalization of firms is concentrated in the tenth decile, whereas the first decile represents 14% of the equity market capitalization on average.

Another important question that naturally arises is whether only the equity portion of the balance sheet is mispriced or if the debt fraction of the firm is similarly mispriced. To gauge the potential importance of this issue, we recompute the weights including debt:

$$wm_{i,j} = \frac{1}{T}\sum_{t=1}^{T}\frac{E_{i,j}+L_{i,j}}{\sum_{i=1}^{10}E_{i,j}+L_{i,j}}, \tag{6}$$

where $L_{i,j}$ is the book value of the liabilities of the firms in decile i. The results are summarized in Panel D of Table 1 and show that debt largely offsets the differences in market capitalization across the BtM deciles. That is, even though the equity market capitalization of the first decile is much larger than the one of the tenth decile, once we include debt in the market capitalization measure, the two numbers are much closer. In contrast, for investment the numbers are essentially the same regardless of whether we include debt.

While these results indicate that accounting for debt can have a significant effect on the cross-sectional distribution of market capitalization across BtM deciles, the weights computed based on equation (6) are not informative about the extent to which the debt component of a firm's capital is mispriced. To be conservative, we will estimate alphas effectively assuming no mispricing on the debt portion (see Appendix A for further details).[5] When assessing the results of our model we will perform sensitivity analysis with respect to this assumption.

Finally, we explore one last important feature, which is the persistence of the mispricing. One way to get a sense of this is to assess the persistence of the sorting variable used to

[5]One impediment to estimating debt alphas for the whole cross-section of firms in our data base is that while the equity is (by definition) publicly traded, a substantial fraction of the debt is not.

identify over- and undervalued firms. To this end, we 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] The table shows that firms in a particular Book-to-Market decile have a substantial probability of still being in that decile after one year. The average diagonal element across all deciles is roughly 1/3. This persistence is lower for the investment sorts (about 1/5).

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 frictions in the form of costly search. The structural parameters of the model are governed by a set of exogenous state variables, which are described in detail in Section 4.2 below. For notational convenience, we will omit parameters' functional dependence on these states elsewhere in the model description.

4.1. Firm Technology

Each firm uses capital K to produce output at a flow rate $A_t K_t^{\eta}$, where $0 < \eta < 1$, and where A_t denotes a productivity process. Firms incur proportional costs of production at rate $c_f K_t$. The capital stock is affected by firm investment, disinvestment, and depreciation. We propose a novel specification of firm technology featuring search for lumpy investments. For any given firm policy function this technology yields closed-form solutions for conditional and stationary distributions of all quantities of interest, allowing us to side-step simulations when estimating the model. Capital takes values in a discrete set indexed by $\kappa_t \in \Omega_\kappa = \{1, 2, ..., N_\kappa\}$. The relation between the index κ_t and capital K_t is given by:

$$K_t = \underline{K} \cdot e^{(\kappa_t - 1)\cdot\Delta}, \qquad \text{with } \underline{K} > 0 \text{ and } \Delta > 0. \tag{7}$$

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

By choosing Δ small enough, this specification can approximate a continuous support for capital arbitrarily well. The discrete state space structure, however, increases the tractability of the model,[7] and allows capturing the lumpy nature of investment observed in the data.

Let N_t^+ and N_t^- denote Poisson processes keeping track of successful capital acquisitions and sales, and let N_t^δ denote a Poisson process for depreciation shocks. The corresponding capital evolution equation is given by:

$$d\log(K_t) = \Delta \cdot (dN_t^+ - dN_t^- - dN_t^\delta). \tag{8}$$

Firms actively search for opportunities to invest or divest capital. Specifically, they control the Poisson intensities of the processes N_t^+ and N_t^-. Each firm chooses its expected investment rate

$$i_t^+ \equiv \mathbb{E}_t\left[\frac{K_t(e^\Delta - 1)dN_t^+}{K_t}\right] \geq 0, \tag{9}$$

and stochastically succeeds in finding an opportunity to upgrade its capital to the next-higher level, that is, by an amount $K_t(e^\Delta - 1)$, with a Poisson intensity $\frac{i_t^+}{e^\Delta - 1}$. As a result of its search efforts, a firm incurs search costs at a rate that is quadratic in i_t^+:

$$\theta_t^+(i_t^+)^2 K_t, \qquad \text{with } \theta_t^+ > 0. \tag{10}$$

Once a purchase opportunity is found, the price of new capital is given by p_t^+ per unit of capital. When a firm reaches the upper bound ($\kappa_t = N_\kappa$), search is assumed to be ineffective in delivering additional investment opportunities. By choosing N_κ high enough, this restriction will be immaterial, as optimal investment will be zero above some endogenous threshold for capital.

[7] Models with a continuous support in any case need to be approximated by a discrete support when solved numerically.

Similarly, each firm chooses its expected disinvestment rate

$$i_t^- \equiv \mathbb{E}_t\left[\frac{K_t(1-e^{-\Delta})dN_t^-}{K_t}\right] \geq 0, \tag{11}$$

and finds an opportunity to divest an amount of capital $K_t(1-e^{-\Delta})$ with a Poisson intensity $\frac{i_t^-}{1-e^{-\Delta}}$. Searching for divestment opportunities causes a firm to incur search costs at a rate:

$$\theta_t^-(i_t^-)^2 K_t, \quad \text{with } \theta_t^- > 0. \tag{12}$$

Once an opportunity to divest is found, the sales price per unit of capital is p_t^-. We assume that divestments are infeasible when capital reaches the lower bound $\underline{K}$. By choosing $\underline{K}$ low enough, we can ensure that a firm would never optimally attempt to divest at this lower bound, such that this restriction is also non-binding.

Capital depreciates stochastically to the next-lower level, that is, from K to $Ke^{-\Delta}$, with a Poisson intensity $\frac{\delta_t}{1-e^{-\Delta}}$, except at the lower bound $\underline{K}$. Thus, the expected depreciation rate is δ_t, except at the lower bound, where it is zero. Overall, the expected growth rate of capital is thus given by:

$$\frac{\mathbb{E}[dK_t]}{K_t} = (i_t^+ - i_t^- - \delta_t)dt, \text{ for } K_t > \underline{K}. \tag{13}$$

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 state (z). The firm-specific state z influences structural parameters affecting variables such as firm-level mispricing and productivity. 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 the 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.

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 between states Z, denoted by $\lambda(Z, Z')$, 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:

$$\frac{dY_t}{Y_t} = \mu(Z_t)dt + \sigma(Z_t)dB_t. \tag{14}$$

A firm's cost function parameters $(c_f, \theta^+, \theta^-)$, the purchase and sales prices of capital (p^+, p^-), and productivity A are all assumed to scale linearly with Y, capturing common growth in these variables.

4.3. Expectations and Market Valuations

Let $s_t = (\kappa_t, z_t, Z_t, Y_t)$ denote the vector of a firm's state variables. Agents value a firm's stream of future after-tax net payouts as follows:

$$V(s_t) = \mathbb{E}_t \int_t^\infty \frac{m(Y_\tau, Z_\tau)}{m(Y_t, Z_t)} e^{-\int_t^\tau \alpha(s_k)dk} \pi(s_\tau)d\tau, \tag{15}$$

where m represents the undistorted stochastic discount factor (SDF) reflecting agents' marginal utility, $\mathbb{E}$ represents an unbiased rational Bayesian expectation that incorporates all public information, and α captures distortions further discussed below. We use $\pi(s_\tau)$ in equation (15) to denote the undistorted *expected* net payout conditional on state s_τ. Note that after conditioning on s_τ, the actual net payout over the interval $[\tau, \tau+dt]$ is still stochastic because of the random nature of a firm's investment search technology.

The pricing equation (15) is flexible enough to capture multiple forces that could lead to informationally inefficient market prices. For the purposes of the exposition, we will consider deviations from perfect rational expectations that efficiently incorporate all public information, which may for example arise if agents incur costs when processing information, or are subject to rational inattention (Sims, 2003).[8] Agents have homogenous beliefs, and the economy is arbitrage-free under these beliefs.

To reduce the number of free parameters and for parsimony, we will specify α as a firm-specific process that is independent of all other exogenous parameters.[9] The corresponding lack of correlation with aggregate states implies that we are indeed studying a purely cross-sectional phenomenon. Beliefs encoded in market prices (15) are informationally inefficient in that agents' expectations at time t of a firm's net payout at time τ in state s_τ are:

$$\underbrace{\mathbb{E}[e^{-\int_t^\tau \alpha_k dk}|\alpha_\tau, \alpha_t]}_{\text{Mispricing Wedge}} \cdot \pi(s_\tau), \tag{16}$$

that is, agents' expectations are effectively discounting or inflating payments in state s_τ at time τ relative to the perfect information processing benchmark.[10] Yet (16) also highlights that, as the firm approaches date τ, expectations converge to the undistorted conditional expectations, $\pi(s_\tau)$, since the mispricing wedge converges to one.

The pricing equation (15) further implies that a firm in state α_t indeed generates an abnormal return equal to α_t over the next instant. In fact, the state α_t affects both the abnormal *return* over the next instant and, due to persistence in alpha, the current *level* of mispricing captured by the mispricing wedge in equation (16). For example, if a claim to a cash flow in one year is currently underpriced by -10% (the mispricing wedge equals 0.9) then abnormal returns on that claim must accumulate to $+10\%$ over this time period. The path of the resolution of mispricing is, however, stochastic.

[8] Empirical evidence suggests that many anomalies that have been uncovered by the finance literature were indeed caused by imperfect information processing. McLean and Pontiff (2016) show that asset pricing anomalies are attenuated substantially after they are documented in academic publications.

[9] In Appendix E we show that equation (15) is flexible in that it can capture a range of mispricing phenomena that are consistent with no arbitrage.

[10] Note that both the "mispricing wedge" in (16) and $\pi(s_\tau)$ condition on α_τ. After all, the state vector s_τ subsumes α_τ.

We consider a partial equilibrium analysis in that we quantify efficiency losses, taking as given a particular undistorted SDF m. In particular, we specify the dynamics of m as a flexible Markov-modulated jump diffusion process:

$$\frac{dm(Y_t, Z_t)}{m(Y_t, Z_t)} = -r_f(Z_t)\,dt - \nu(Z_t)dB_t + \sum_{Z' \neq Z_t} (e^{\phi(Z_t, Z')} - 1)dM_t(Z_t, Z'). \tag{17}$$

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'.[11] Let $\bar{\Lambda}_Z$ denote the generator matrix under the risk neutral measure, which collects the risk neutral transition rates $\bar{\lambda}(Z, Z') = e^{\phi(Z,Z')}\lambda(Z, Z')$. The specification of the SDF (17) is sufficiently flexible to capture a variety of benchmark asset pricing models, such as long run risk models and rare disaster models, making our methodology applicable to a variety of settings.

Our 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 informationally efficient markets (since output and consumption change), we know that risk free rates tend to be not volatile and thus, in any empirically plausible calibration, respond relatively little to changes in consumption growth. Further, if risk premia fell in a boom following the elimination of cross-sectional misallocations, this would lead to lower discount rates and correspondingly, even higher estimates for our efficiency gain measures. Overall, we therefore do not expect that our results would be materially affected after accounting for the general equilibrium effect of changes in consumption growth on the SDF.

4.4. Firm Objective

Firms take the market prices in (15) as given and choose the investment strategy $\{i^+, i^-\}$ that maximizes their market value at any point in time. As discussed in the introduction, the

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

view that agents have homogenous beliefs and firms maximize their market value is a useful benchmark given that we analyze anomalies that are based on *public* information. However, one could be concerned that our estimated model implies too strong a relation between mispricing and investment. What we will show is that, if anything, it underrepresents the empirical cross-sectional relation between firm investment and alphas. We discuss this issue as well as empirical support for our identifying assumptions in Sections 6.5 and 7.

4.5. Analysis

4.5.1. Firm Behavior

In this subsection we analyze firm behavior. Firms dynamically maximize their market value as specified in (15) by choosing their expected investment and disinvestment rates i^+ and i^-. Let V denote a firm's value function, where

$$V(s_t) = \max_{\{i^+,i^-\}\geq 0} \mathbb{E}_t \int_t^{\infty} \frac{m_\tau}{m_t} e^{-\int_t^\tau \alpha_k dk} \pi_\tau d\tau, \tag{18}$$

and where we define the undistorted expected after-tax net payout conditional on date-t information:

$$\begin{aligned}\pi_t =&(1-\tau)(A_t K_t^\eta - (c_{f,t} + p_t^+ i_t^+ + \theta_t^+ (i_t^+)^2 - p_t^- i_t^- + \theta_t^- (i_t^-)^2) K_t) \\ &+ (i_t^- + \delta_t - i_t^+) p_t^+ \tau K_t.\end{aligned} \tag{19}$$

Equation (19) reflects that firms obtain tax shields from depreciation, proportional cost of production, search cost, and selling capital below the purchase price p_t^+. As discussed above, π_t denotes an *expected* net payout conditional on date-t information, accounting for the stochastic arrival of (dis)investment opportunities over the interval $[t, t + dt]$.

Since A, c_f, p^+, p^-, θ^+, and θ^- are assumed to scale linearly with the trend factor Y, we can conjecture and verify that the value function is linear in Y, that is, $V(s) = Y \cdot \tilde{V}(\kappa, z, Z)$, where going forward, a tilde indicates that a variable is scaled by Y. The Hamilton-Jacobi-Bellman equation associated with the maximization problem in (18) implies that $\tilde{V}(\kappa, z, Z)$

solves the following set of equations for all $(\kappa, z, Z) \in \Omega_\kappa \times \Omega_z \times \Omega_Z$:[12]

$$\begin{aligned}
0 = \max_{\{i^+, i^- \geq 0\}} & [\tilde{\pi}(\kappa, z, Z) - (r_f(Z) + \sigma(Z)\nu(Z) + \alpha(z) - \mu(Z))\tilde{V}(\kappa, z, Z) \\
& + \frac{i^+}{(e^{\Delta} - 1)}(\tilde{V}(\kappa + 1, z, Z) - \tilde{V}(\kappa, z, Z)) \\
& + \frac{\delta + i^-}{(1 - e^{-\Delta})}(\tilde{V}(\kappa - 1, z, Z) - \tilde{V}(\kappa, z, Z)) \\
& + \bar{\Lambda}_Z(Z)\tilde{\mathbf{V}}^Z(z, \kappa) + \Lambda_z(Z)\tilde{\mathbf{V}}^z(Z, \kappa)],
\end{aligned} \tag{20}$$

where $\tilde{\mathbf{V}}^Z$ and $\tilde{\mathbf{V}}^z$ are vectors that collect the values of the function $\tilde{V}$ 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 and disinvestment rate:

$$i^+(\kappa, z, Z) = \max\left[\frac{\left(\frac{\tilde{V}(\kappa+1,z,Z) - \tilde{V}(\kappa,z,Z)}{(e^{\Delta}-1)K(\kappa)} - \tilde{p}^+(z, Z)\right)}{2(1-\tau)\tilde{\theta}^+(z, Z)}, 0\right], \tag{21}$$

$$i^-(\kappa, z, Z) = \max\left[\frac{\left(\frac{\tilde{V}(\kappa-1,z,Z) - \tilde{V}(\kappa,z,Z)}{(1-e^{-\Delta})K(\kappa)} + (1-\tau)\tilde{p}^-(z, Z) + \tau \cdot \tilde{p}^+(z, Z)\right)}{2(1-\tau)\tilde{\theta}^-(z, Z)}, 0\right]. \tag{22}$$

Conditional on these policy functions, the system (20) is linear in $\tilde{V}(\kappa, z, Z)$, which implies that obtaining exact solutions is straightforward and fast.

Under agents' distorted beliefs, a firm's expected return in excess of the risk free rate is given by:

$$rp(\kappa, z, Z) = \sigma(Z)\nu(Z) - \sum_{Z' \neq Z} \lambda(Z, Z')\left(e^{\phi(Z,Z')} - 1\right)\left(\frac{\tilde{V}(\kappa, z, Z')}{\tilde{V}(\kappa, z, Z)} - 1\right). \tag{23}$$

In contrast, an observer outside of our economy that perfectly processes public information would expect an excess return of $(\alpha + rp)$. The risk premium rp features compensation for exposures to both innovations to the Markov state Z and Brownian innovations to the

[12] See Appendix C for details.

common trend Y.

4.5.2. Cross-sectional Distributions

To measure the aggregate magnitude of misallocations, it is important for the model to capture the cross-sectional distributions of firm characteristics such as size and Tobin's q. We show in Appendix D how, for any given policy function, we can compute the stationary and conditional cross-sectional distributions in closed-form, which greatly facilitates the estimation and evaluation of the model.

5. Estimating the Model

5.1. Specification of the Markov Processes

We consider a parsimonious specification of the model with two exogenous independent firm-level processes that are governed by the state z: a process for productivity $\tilde{A}$ and one for α. To facilitate matching the size distribution, we consider two sets of firms that differ with respect to the ranges of their productivity distributions. In particular, the set of possible productivity values for the two sets differ by some constant multiplier different from one. We estimate this multiplier and the relative masses of the two sets of firms. Conditional on belonging to one of the two sets, which is a constant firm characteristic, the firm-specific state z is thus characterized by the tuple $(\tilde{A}, \alpha)$. The macro-state Z governs expected trend growth μ, trend volatility σ, and the dynamics of the SDF (see equation (17)).

5.2. Calibration and Estimation

We calibrate the parameters of the macroeconomy based on the existing literature and estimate firm-specific parameters using a generalized method of moments (GMM) approach.[13]

[13] The literature used for the calibrated parameters includes Chen, Cui, He, and Milbradt (2015) and the references therein.

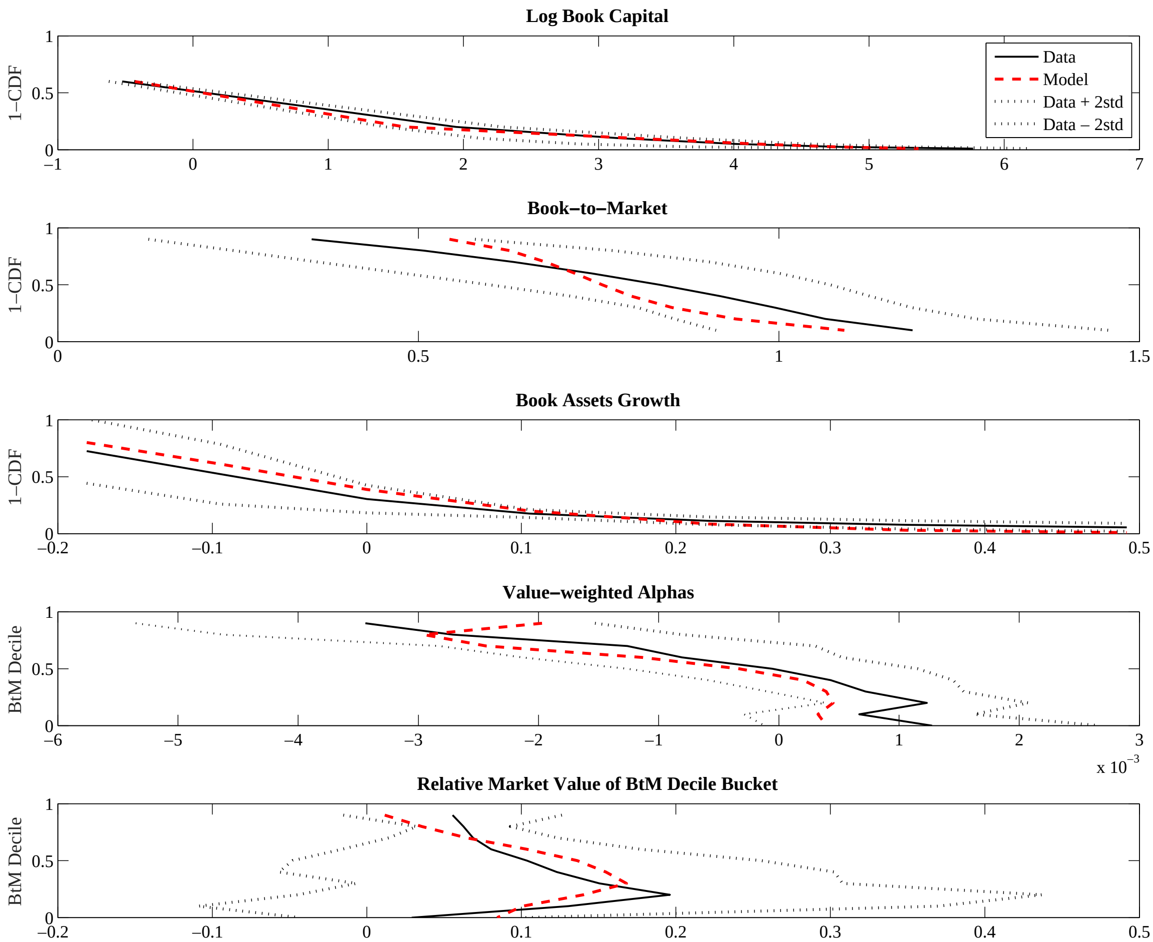

FIGURE I

Model fit. The panels plot for each variable the model-implied values (red dashed line) and compare them with the data (black solid line), including 95% confidence bounds (black dotted lines). The figure illustrates 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), the monthly value-weighted alpha of each Book-to-Market decile portfolio (the fourth panel), as well as the relative market value of each Book-to-Market decile portfolio (the bottom panel).

We estimate 23 parameters by targeting 42 moments related to the cross-sectional distribution of firms: 9 book to market decile breakpoints, 6 book asset breakpoints, 7 book asset growth percentiles, 10 CAPM alphas corresponding to the Book-to-Market deciles, and 10 market value weights of these Book-to-Market deciles.[14]

Consistent with the model's focus on total firm value as opposed to merely firm equity, we compute the book value of assets by the sum of book equity and book debt, and the market value of assets as the sum of market equity and book debt. The Book-to-Market ratio is the ratio of these two quantities. Correspondingly, we also compute alphas for the total return on assets, as detailed in Appendix A.

The fitted moments are summarized in Figure I, and the calibrated and estimated parameters are listed in Table 2. The figure plots 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), the monthly value-weighted alpha of each Book-to-Market decile portfolio (the fourth panel), and the relative market value of each Book-to-Market decile portfolio (the bottom panel). The model moments all fall within the 95% confidence bounds, with the exception of the highest investment percentiles.

The second and fourth panel ("Book-to-Market" and "Value-weighted Alphas") are worth discussing in more detail. In the data, about 30% of firms have a Tobin's q less than one. Further, the Tobin's q of those firms is quite persistent, as highlighted by Panel E of Table 1. The data therefore suggests that firms face substantial frictions when attempting to disinvest. After all, without such frictions, a firm's capital would adjust to ensure that Tobin's q does not stay below one. Further, it is worth discussing what drives the cross-sectional variation in alphas in relation to the Book-to-Market ratio in the model. Book-to-Market (the inverse of

[14]Since we observe only publicly traded firms, we also account for delistings and new listings. Specifically, to match delisting rates in the data we estimate historical (average) exit rates in each sales-to-book decile. A firm's exogenous Poisson 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.

Tobin's q) endogenously emerges in the model as a *noisy* measure of a firm's underlying alpha state, as market prices generally respond more quickly to mispricings than capital does.[15] Yet most of the variation in Tobin's q in the model is still due to variation in technology, such as productivity and depreciation shocks, which are independent of a firm's α process. In summary, the Book-to-Market ratio is only a proxy for mispricing, not a perfect measure.

The parameters in Table 2 all fall in a range that is broadly consistent with the literature. The depreciation rate is 16%, which is somewhat higher than the standard values used in the literature. The proportional costs of production are 2.7%.[16] The decreasing returns to scale parameter is 0.975, which implies a technology close to constant returns to scale, consistent with evidence in Hall (2003).[17] The discount on selling used capital $(1-p^-/p^+)$ equals 44%, and there are high quadratic search costs for finding opportunities to sell installed capital. These estimates emerge due to the before-mentioned empirical fact that about 30% of firms in the cross-section have a Tobin's q less than one. In contrast, search frictions for capital acquisitions are lower. Economically, this asymmetry is consistent with asset specificity, asymmetric information, and other forces causing investment to be at least partly irreversible (see Arrow, 1968, Pindyck, 1988).[18]

Firm-level alphas follow a mean-reverting three-state Markov process that is restricted to have an unconditional mean of zero, implying that the average alpha in the cross-section of firms is zero at any point in time.[19] The estimated transition rates of the alpha process imply substantial mean reversion, leading conditional alphas over a one year horizon to

[15] Related to this effect, Berk (1995) argues that sorts on lagged market values mechanically capture either omitted risk factors or mispricings.

[16] Given the estimated parameters, firms' market valuations in the model (measured empirically by debt plus equity) have the desirable feature that they are greater than zero in all states of the world, despite the presence of proportional costs of production.

[17] As a robustness check, we have also estimated our model setting this parameter to 0.65, leading to similar implications regarding efficiency distortions. However, we have found that imposing a lower decreasing returns to scale parameter leads to a deterioration of the model's ability to fit the Book-to-Market distribution.

[18] Pindyck (1988) argues that most major investment expenditures are at least partly irreversible "*because capital is industry- or firm-specific, that is, it cannot be used in a different industry or by a different firm.*" In addition, Pindyck highlights that partial irreversibility may result from lemons problems — "*Office equipment, cars, trucks, and computers are not industry-specific, but have resale value well below their purchase cost, even if new.*"

[19] More alpha states can be accommodated, but a three-state process turns out to be sufficient to capture the cross-sectional moments we target.

Parameters of the Macroeconomy (Calibrated)			
Parameter	Variable	Boom	Recession
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	0.031	0.031
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+}	2.828
Rate of moving to next-lower $\tilde{A}$	h_{A-}	2.238
Search costs for capital acquisitions	$\tilde{\theta}^{+}$	0.793
Search costs for capital sales	$\tilde{\theta}^{-}$	4.734
Sales price of capital	$\tilde{p}^{-}$	0.559
Decreasing returns to scale parameter	η	0.975
Expected depreciation rate	δ	0.164
Proportional cost of production	$\tilde{c}_f$	0.027

Firm-specific α-Process				
Parameter	Variable	α_1	α_2	α_3
Rate of moving to next-higher α-state	$h_{\alpha+}$	2.006	1.418	–
Rate of moving to next-lower α-state	$h_{\alpha-}$	–	0.354	0.415
Local alpha	α	-0.187	-0.037	0.020

Table 2

Parameters. The three panels list parameters of the macroeconomy and firm-specific parameters. The parameters of the macroeconomy are calibrated. There are two aggregate Markov states $Z \in$ {Boom, Recession}. We normalize the scaled purchase price of capital $\tilde{p}^{+}$ to one. We estimate firm-specific parameters via a GMM approach. For the first set of firms, the set of possible productivity states is given by $\tilde{A}_i = \tilde{A}_1 e^{\sum_{j\leq i} a_j}$ with $\tilde{A}_1 = 0.00512$ and $a_j \in \{0, 0.652, 0.597, 0.542, 0.452, 0.362, 0.550, 0.738, 0.534, 0.329\}$, where we estimate only every second a_j-value and determine the remaining values via interpolation. The possible productivity states for the second set of firms are given by $\{1.104 \cdot \tilde{A}_i\}_{\forall i}$, and the fraction of firms that have this upward-shifted productivity process is estimated to be 16.95%. The capital grid is characterized by the lower bound $\underline{K} = 0.00239$ (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 $\Delta = 0.1$.

be substantially closer to zero than instantaneous (local) alphas. The conditional *one-year* alphas in the three states are −9.5%, −2.3%, and +1.2%, whereas the corresponding *local* alphas reported in Table 2 are −18.7%, −3.7%, and +2.0%, respectively. The unconditional probabilities of the three alpha states are 4%, 22%, and 74%, indicating that firms are rarely in the state with the most negative alpha. At any point in time, 96% of firms in the cross-section face conditional one-year alphas ranging between −2.3% and +1.2%.

6. Results

6.1. Under- and Overinvestment

As a first step, we assess the influence of cross-sectional alphas on investment. In Figure II we plot the probability distribution function (PDF) of the log difference between distorted and undistorted (expected) investment rates, where defined.[20] The plot shows substantial deviations from undistorted investment policies, with both sizable over- and underinvestment. Yet the graph also reveals that the distortions are asymmetric. Because of the substantial frictions associated with disinvesting capital, firms respond little to mispricing distortions when wishing to disinvest, making underinvestment less prevalent.

Even though the plot shows that investment distortions can be large, it is not clear how important these effects are for aggregate surplus creation. We explore this key question in the next section.

6.2. Measuring Potential Value Gains

To quantitatively assess the influence of cross-sectional distortions on value creation, we first compute the stationary firm distribution for the estimated model under the distorted policies. We denote this cross-sectional firm distribution by F^d. Next we consider the counterfactual where, starting from this cross-sectional distribution, firms switch from following

[20] Due to differences in the sales and purchase prices of capital, there are substantial inaction regions where firms neither invest nor disinvest, implying that $\log(i_+ - i_-)$ is not defined for a subset of states.

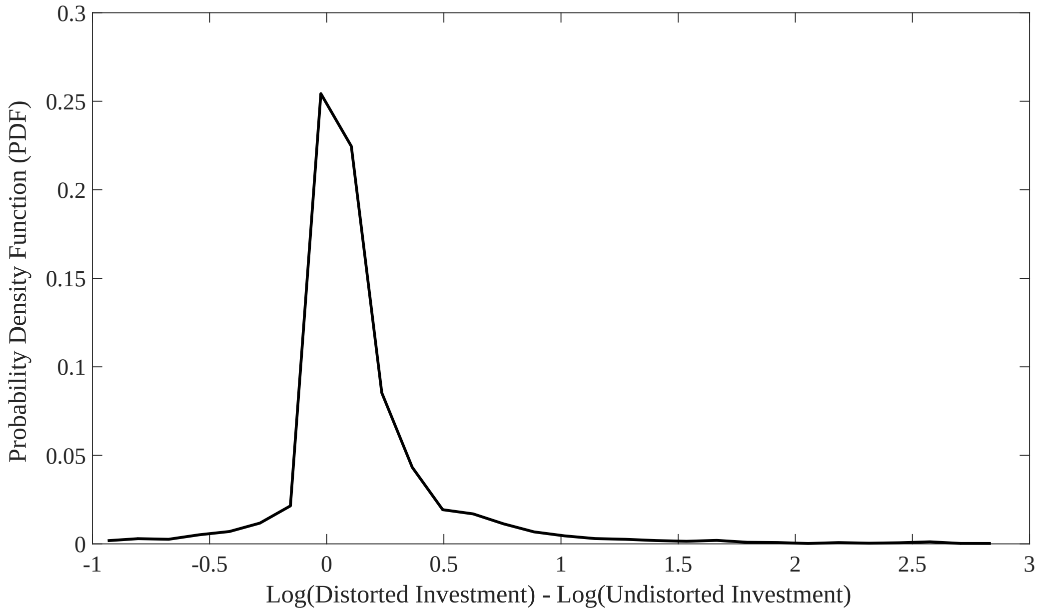

FIGURE II

Investment distortions. The graph plots the probability distribution function (PDF) of the log difference between distorted and undistorted expected investment rates in each state, where defined. Under the estimated parameterization there are substantial inaction regions where $(i_+ - i_-) = 0$, implying that $\log(i_+ - i_-)$ is not defined for a subset of states. The PDF of all states is computed for the distorted economy, the parameters of which are detailed in Table 2. The undistorted expected investment rates in each state are computed by solving the model with the same parameters, but setting $\alpha_i = 0 \ \forall i$.

distorted investment policies to following undistorted policies going forward. Since we have closed-form solutions, it is straightforward to account for transition dynamics when computing the associated present value of surplus gains. Specifically, we introduce the aggregate value gain measure:

$$gain = \frac{\mathbb{E}[\int_0^\infty \frac{m_\tau}{m_0} \pi_\tau^u d\tau | F^d]}{\mathbb{E}[\int_0^\infty \frac{m_\tau}{m_0} \pi_\tau^d d\tau | F^d]} - 1, \tag{24}$$

where $\{\pi_t^u\}_0^\infty$ denotes the stochastic stream of *aggregate* net payout associated with undistorted firm policies and $\{\pi_t^d\}_0^\infty$ is the one associated with distorted policies. As in the one-period example in Section 2, we value these cash flow streams using the undistorted beliefs and the undistorted SDF.

The *gain* estimate can be interpreted as the perpetual percentage of total firm net payout that society is willing to pay for eliminating the alpha process under consideration. The *gain* measure thus can be viewed as the magnitude of potential compensation (fees) for financial intermediaries, provided these intermediaries completely eliminate alpha.

The independent alpha process is primarily identified and estimated through the data on the Book-to-Market anomaly (the value spread). As Tobin's q is only a noisy measure of the underlying alpha state (see the discussion in Section 5.2), eliminating the alpha process implies the elimination of the value spread, while the opposite implication does not hold. Put differently, even if financial intermediaries were to completely eliminate the value spread, this would still not yield the full efficiency gain we compute.

Figure III plots the aggregate value gain defined in equation (24) for different alpha process parameterizations. The parameterizations only differ in the range of alpha, resulting in different levels of model-implied value spreads in a (non-linear) one-to-one mapping. Our baseline estimation, as presented in Table 2 and Figure I, corresponds to an annual value spread in the model of about 3%, which is more conservative than the 5.4% found in the data (see the fourth Panel in Figure I which plots monthly alphas in the model and in the data).

Further, we chose not to make the assumption that cross-sectional debt mispricing is the same as the corresponding cross-sectional equity mispricing. As a consequence, our fitted value spread that applies to total firm value is substantially smaller than the empirical estimates of the value spread for equity reported in the literature as well as in Table 2. If one is willing to assume that the alpha for debt is as large as it is for equity, the relevant range for the value spread to consider is 6% to 9%, which results in substantially larger value gains due to a convex relation. In summary, the value gains from eliminating alpha are several percentage points, with estimates ranging between 2% and 16% depending on the targeted value spread.

When interpreting the magnitudes of our gain measure, it is important to keep in mind that our present value computations refer to the net payout of public firms, which is not the same as GDP. In fact, these payouts are typically less than half of GDP. In addition, the estimates do not directly speak to the fair compensation that the financial sector should have

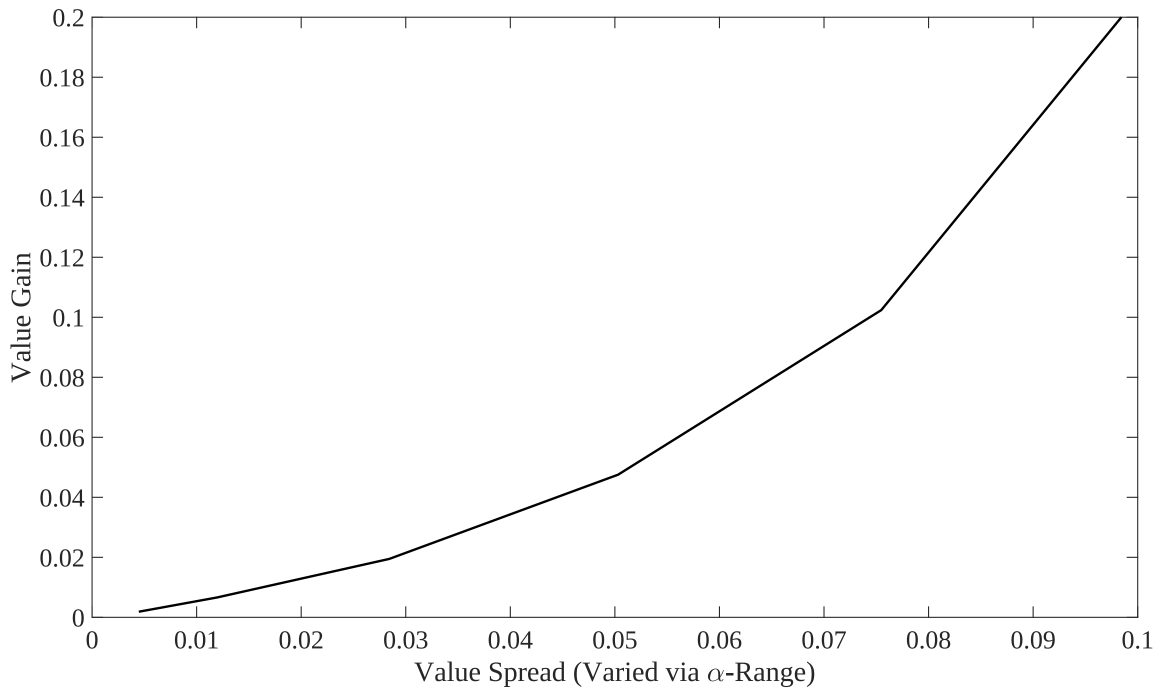

FIGURE III

Aggregate value gain as a function of the value spread. This graph plots the aggregate *gain* measure in equation (24) for different economies that are indexed by their value spread. The value spread is measured by the difference between the alpha of the top decile portfolio and the alpha of the bottom decile portfolio. We vary the value spread by introducing a mean-preserving spread in the alpha process specified in Table 2. In particular, we multiply the value of α_3 by a constant between 0.6 and 1.6, and then subtract the unconditional mean of this adjusted process from each of the three alpha states to maintain an alpha process that has an unconditional mean of zero.

received historically. Instead, it evaluates how large compensation could be if the financial sector eliminated existing informational inefficiencies in the future.[21]

Real distortions and the dynamic nature of firms' decisions. A key ingredient generating the real distortions that we quantify is the inherently dynamic nature of firms' investment decisions. As the time-horizon of the investment decision approaches zero, the cumulative belief distortion defined in equation (16) also goes to zero, implying that real inefficiencies vanish despite the presence of financial inefficiencies. The effective time-horizon

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

in models like ours is in turn affected by adjustment frictions. Thus, as the search frictions in our model go to zero and firms' problems become essentially static, real inefficiencies disappear because beliefs about *current*-period cash flows are not distorted by alphas. Similarly, if search costs become infinitely large then optimal firm (dis)investment is zero, independently of alpha, also implying that real decisions are unaffected by mispricings. These insights highlight that the dynamic nature of firms' decisions — involving the forecasting and discounting of future cash flows — is essential for informational inefficiencies as measured by alpha to cause real distortions.

6.3. Value Gains and Tobin's q

In this subsection we assess whether the value gain from moving to undistorted investment policies differs for firms with different Tobin's q ratios. The solid line in Figure IV plots how much value an individual firm can gain on average conditional on having a particular value of Tobin's q. The graph shows that most of the gain is achieved for high Tobin's q firms ("growth 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 one ("value firms"). Both in the data as well as in our model about 30% of firms have this characteristic, which as argued before, suggests that the frictions for disinvesting are high. Such firms would like to disinvest, as their capital could be used more efficiently outside of the firm. Yet, because of these frictions, they hardly do so. For ease of exposition, suppose that investment is completely irreversible. Now take a firm with Tobin's q less than one that becomes undervalued due to a positive alpha shock. The firm then wishes to disinvest even more, but is still unable to do so because of the irreversibility of investment. As a consequence the real investment behavior with or without the alpha is the same. Similarly, if the firm becomes overvalued but still has a Tobin's q lower than one, it still does not invest or disinvest, once again leaving the real investment behavior undistorted. Now consider growth firms, which by definition have a Tobin's q larger than one. These firms' investment rates are positive and sensitive to (mis)valuations. As a consequence, alphas can cause substantial investment distortions for growth firms.

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 despite having a high-Tobin's q, a firm can still be undervalued.

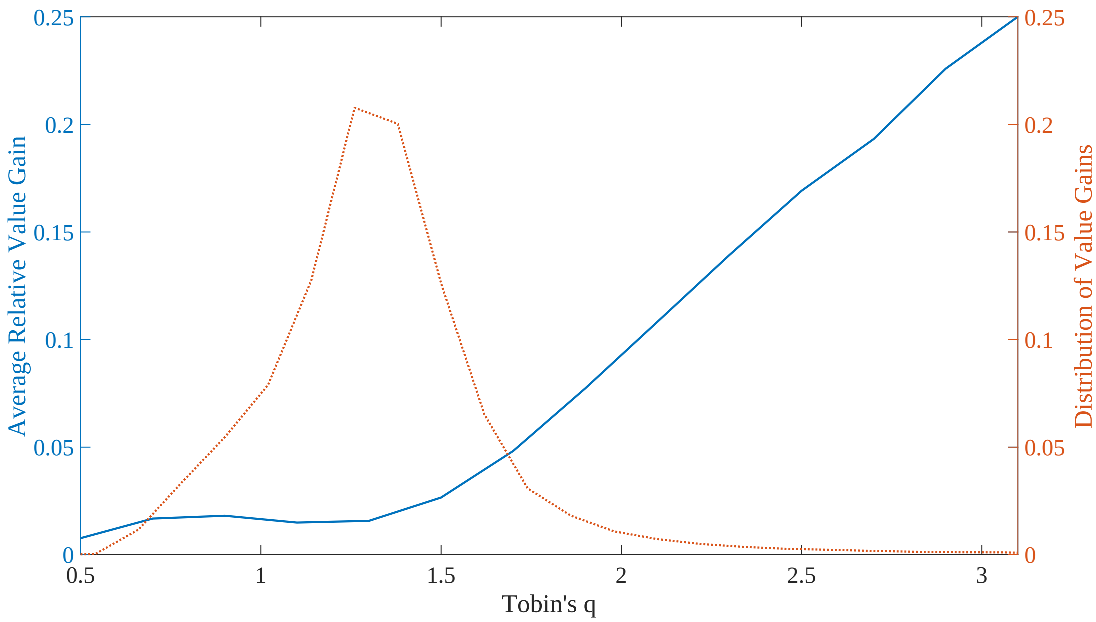

FIGURE IV

Value gains and Tobin's q. The blue solid line plots the average relative value gain (left vertical axis) for an individual firm from moving to undistorted investment policies as a function of the firm's Tobin's q (horizontal axis). The orange dotted line plots the value gain distribution (right vertical axis) from moving to the undistorted investment policies as a function of Tobin's q. The latter is scaled by the economy-wide gain. The parameters of the economy are detailed in Table 2.

While the solid line in Figure IV reflects a firm's value gain conditional on having a particular Tobin's q, it does not reveal the amount of aggregate market value that is concentrated at that level of Tobin's q. For example, little market value is associated with a Tobin's q larger than 2.5. For this reason, we also plot in the same figure the distribution of the total value gain (the orange dotted line). This distribution 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 2. After all, to generate welfare gains two conditions have to be met. First, conditional on a particular Tobin's q the welfare gain (i.e., the solid line in Figure IV) needs to be substantial. Second, a significant amount of value needs to be concentrated at that level of Tobin's q.

6.4. The Investment-q Relationship

As is well-known in the investment literature, the relationship between Tobin's q and investment is weak. One may wonder to what extent we replicate this weak relationship in our model. In the estimated model the investment-q slope is 0.079 with an R^2 value of 0.064.[22] 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 firms are not responding to (dis)information that is reflected in market prices. After all, in our model, firms are by construction maximizing market values.

6.5. Investment-α Relationship

We chose to estimate our alpha process by matching the cross-sectional relation between Book-to-Market decile portfolios and alphas. As can be seen from the lower panel of Figure I, we fit the alphas associated with the Book-to-Market distribution quite well. We now evaluate whether given these parameter estimates, our model also generates the investment anomaly, that is, the cross-sectional relationship between investment decile portfolios and alphas. In Figure V we plot the monthly CAPM alphas generated by the model and compare them to the data. The graph illustrates that our model indeed generates an investment anomaly despite the fact that it was not targeted in the estimation.

[22] These numbers were obtained by simulating 500,000 firms from the stationary distribution for one year. We regress asset growth over one year (from t to $t+1$) on an intercept and Tobin's q at time t. Given that we use 500,000 firm years, the remaining uncertainty about the model-implied slope is very small (the standard error is 0.000445).

The figure also shows that the estimated model is not overstating the extent to which alphas vary across investment decile portfolios. Just as our estimated model underrepresents the magnitude of the value premium, it also underrepresents the size of the investment-alpha relationship, suggesting that our estimates of the aggregate gain measure are conservative. Further, as discussed in the previous subsection, Tobin's q (i.e., the inverse of Book-to-Market) and investment are not highly correlated. Therefore, rank-ordering firms based on these two measures does not lead to identical decile portfolios. Yet, the single alpha process that we estimated endogenously generates both anomalies in the model.

The empirical cross-sectional relation between firm investment and alpha is an intriguing fact, as it is consistent with the notion that firms adjust investment in response to (dis)information encoded in market prices. High investment predicts abnormally low returns, and low investment predicts abnormally high returns, an empirical pattern that naturally tends to arise when firms over- or underinvest when markets over- or undervalue them. As discussed before, 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 firms merely adjusting cash holdings when they are mispriced. In Section 7 we further discuss supplementary empirical evidence in the existing literature indicating that prices do affect firms' real investment decisions. In addition, while we measured informational inefficiencies relative to the CAPM, Hou, Xue, and Zhang (2016) show that investment alphas survive even when considering the other most often used benchmark asset pricing models,[23] highlighting the robustness of this finding. Finally, as discussed in the introduction, similar empirical patterns apply for banks, indicating the economic significance of the friction we analyze (Baron and Xiong, 2016, Fahlenbrach, Prilmeier, and Stulz, 2016).

It is also useful to keep in mind that, even when investment policies are distorted in the way our model posits, rank-ordering firms by investment rates does not have to yield significant spreads in abnormal returns across decile portfolios. Consider the following extreme scenario. Take a stylized deterministic version of our model where firms are exposed to perfectly persistent alphas of different magnitudes. In the steady state of the cross-sectional distribution of firms, all firms' investment rates are equal to each other (and equal to the

[23]These models include the Fama-French 3 factor model, the Carhart model, and the Pastor-Stambaugh factor.

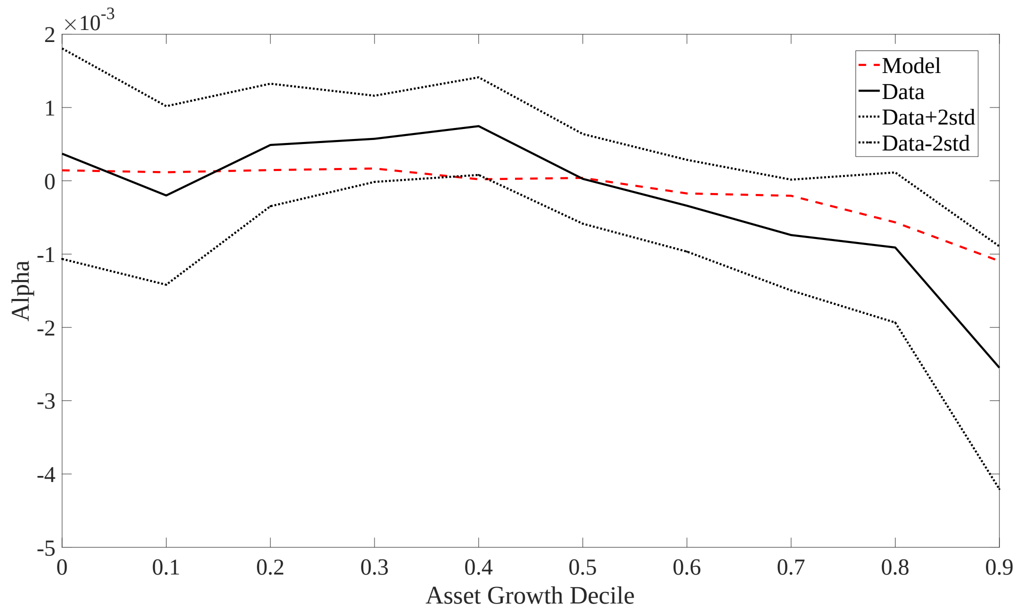

FIGURE V

Investment-α relation. The graph plots the CAPM alphas of investment-sorted decile portfolios in the model (red dashed line) and compares them with the data (black solid line), including 95% confidence bounds (black dotted lines). The parameters of the economy are detailed in Table 2.

depreciation rate), implying that investment rates yield no information about cross-sectional variation in the underlying alphas. Yet this economy can feature large efficiency losses, since alphas are perfectly persistent (see also Section 7.3 below).

7. Robustness Considerations

In this section we discuss the robustness of our quantitative analysis with respect to various modeling assumptions.

7.1. Existing Evidence Supporting Identifying Assumptions

Arguably, two of the most important identifying assumptions for our quantitative analysis are that (1) mispricings measured by CAPM alphas indeed represent informational inefficiencies, and (2) firms decide on investment plans based on (dis)information encoded in market prices. We now discuss these two assumptions in more detail.

First, while we have used CAPM alphas — the main benchmark in the literature — to estimate our model, other asset pricing models can be accommodated by researchers using our methodology. In addition, the gain estimates we present as a function of the value spread in Figure III (see also Figure VI below) allow readers to gauge magnitudes given their individual views about the size of the value anomaly.

Second, the assumption that firm investment responds to (dis)information encoded in market prices is supported by substantial empirical evidence in the existing literature. To identify a causal effect, several existing studies focus on subsets of firms and/or events with plausibly exogenous shocks to prices.[24] The results from this literature go against the alternative view that (dis)information encoded in market prices is a sideshow and does not affect firm investment. While these studies do find a significant feedback effect of prices on investment in specific situations where clean identification is possible, they do not aim to estimate the economy-wide real efficiency losses associated with cross-sectional alphas, which is our goal. As highlighted above, our estimated model endogenously generates the cross-sectional relations between investment and alpha, and Tobin's q and alpha, that is, we do not hardwire such relations through ad hoc assumptions. In addition, if anything, our estimated model understates the cross-sectional relation between investment decile portfolios and alphas. Yet, if researchers believe that firm investment should respond less to disinformation encoded in market prices than implied by our benchmark setup, such a view can be accommodated by multiplying $\alpha(z)$ in equation (20) by a shading parameter that can take values between zero and one. In this case, firms effectively maximize based on their own valuations that are closer to being informationally efficient than prices in financial markets. Once again, Figure III provides a preliminary assessment of the effects of such shading.

[24]See, e.g., Edmans, Goldstein, and Jiang (2012) and the references therein. See also the recent papers by Foucault and Fresard (2014) and Dessaint, Foucault, Fresard, and Matray (2016).

7.2. (In)dependence of the Alpha Process

Our specification of the alpha process does not mechanically assume a cross-sectional relation between Tobin's q and alpha, as pointed out above. That said, readers might still be curious what happens if such a mechanical relation was introduced and firms still maximized their going market values. We evaluated this possibility by specifying alphas directly as a monotonically increasing function of a firm's Book-to-Market ratio instead of as an independent process. As firms' past investment is a determinant of their current book capital, this specification implies that alpha responds to a firm's investment behavior. Now firms maximizing their market value have strong incentives to reduce their Tobin's q. That is, firms' investment policies respond such that the whole cross-sectional distribution of Tobin's q shifts significantly upwards, where firms face negative alphas, and are thus overvalued. Yet, given the magnitude of the value spread, we find that it is then impossible for the model to get even close to matching the empirical cross-sectional distribution of Tobin's q in the data. Further, this specification does not endogenously generate an investment alpha relation as suggested by the data.

7.3. Persistence of the Alpha Process

In this section we analyze the sensitivity of our aggregate gain estimate to 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 aggregate value losses. Figure VI plots the aggregate change in value as computed in equation (24) for different levels of persistence. To vary persistence we multiply transition rates $(h_{\alpha+}, h_{\alpha-})$ of the baseline parameterization by factors ranging between 0.5 and 2. Changing persistence in this manner does not alter the mean of the alpha process, which is equal to zero. Neither does it change the unconditional probabilities of the alpha states. What it does change however is the cross-sectional relation between Book-to-Market and alpha. As the alpha process becomes more persistent, the Book-to-Market ratio becomes a better proxy for the underlying alpha state, leading to higher value spreads (which are measured by the difference between the alpha in the top decile and the alpha in the bottom decile). There is a non-linear one-to-one mapping between the persistence multiplier and the value spread under these parameterizations. Figure VI plots our gain measure for different levels of persistence (and thus, value spreads). The value spreads on the horizontal axis vary between 0.01 (a persistence multiplier of 2) and 0.075 (a persistence multiplier of 0.5). The graph shows that the gains 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 anomaly, they can create large real inefficiencies.

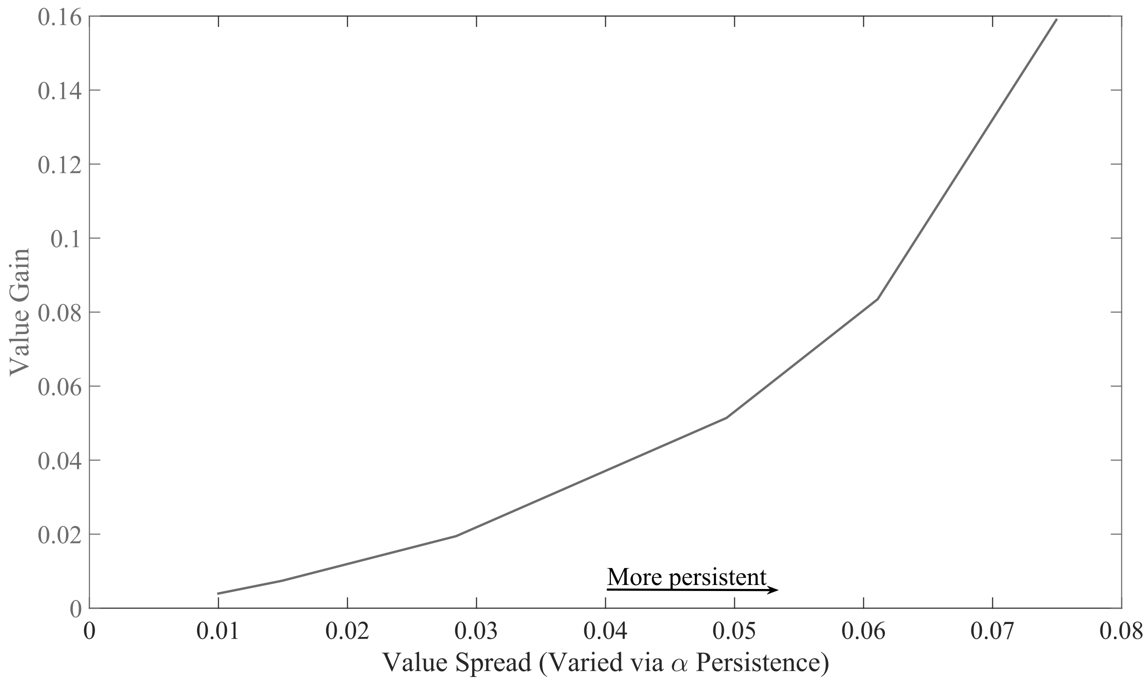

FIGURE VI

Changing α-process persistence. The graph illustrates the effects of changing the persistence of the α-process by multiplying the transition rates $(h_{\alpha+}, h_{\alpha-})$ of the baseline parameterization by a factor $[0.5, 2]$. Because there is a (non-linear) one-to-one mapping between the persistence multiplier and the value spread, we plot the value spread on the horizontal axis. The value spread varies between 0.01 (a persistence multiplier of 2) and 0.075 (a persistence multiplier of 0.5). The other parameters of the economy are detailed in Table 2.

7.4. The Role of Other Frictions

As argued in Section 6.2, real distortions are absent with either infinite or zero adjustment frictions, which implies that there is a non-monotonic relation between the magnitude of adjustment frictions and the real distortions we study. If there are no adjustment frictions, firms maximize current-period net payout, rendering market prices and alphas irrelevant for investment decisions. On the other hand, with extremely high adjustment costs, investment responds so sluggishly that alphas mean revert before investment is significantly distorted. We highlighted previously that approximately 30% of firms have a Tobin's q less than one, suggesting that a significant fraction of firms do face severe frictions when attempting to disinvest. Consequently, any investment model aiming to fit this characteristic in the data will necessarily have to feature significant downward adjustment constraints of some form, be it standard adjustment costs or other frictions. While we model adjustment frictions in the form of costly search, alternative frictions considered in many micro and macro models would act similarly in limiting the responsiveness of firms' investment to disinformation encoded in market prices. Sensitivity analysis with respect to our two search cost parameters (θ^+ and θ^-) indicates in fact that lowering adjustment frictions relative to our benchmark estimation increases the magnitude of real distortions.

Given that our model does not explicitly feature various other adjustment frictions, the estimation assigns all impediments to adjusting capital suggested by the data to the search frictions in our model. As such, these estimates should be interpreted as a catch all — estimates of the search frictions would likely be smaller if other frictions were introduced. However, the overall adjustment frictions would likely have to be of similar magnitude for them to be consistent with the data. The prediction that firms with high q are more affected by mispricings in turn may also be consistent with models where managers are aware of the fact that they face irrational markets, as considered in Stein (1996). In particular Baker, Stein, and Wurgler (2003) find strong support for Stein's (1996) prediction that equity dependent firms — proxied by characteristics such as a high Tobin's q — are more responsive to nonfundamental movements in stock prices. Additionally, Gomes (2001) shows that Tobin's q should capture most of investment dynamics even when there are credit constraints, suggesting that similar results would apply if we extended the model to feature such constraints.

8. Implications and Future Directions

In addition to contributing to the literature that quantifies the economic magnitude of various forms of misallocations, our paper yields insights for several other parts of the literature. First, our framework provides guidance to the return anomalies literature. In the past few decades, a vast number of potential anomalies have been uncovered (see Harvey, Liu, and Zhu (2016) for an overview). In the next subsection we discuss how our framework can help rank these candidate anomalies by their potential economic importance rather than merely the statistical significance of alphas. Second, we have shown that in our framework a single underlying alpha process can reproduce several anomalies, which provides guidance on what information portfolio sorts exactly contain. Finally, our framework contributes to the debate on active versus passive investment management. We discuss how in our model the popular arithmetic proposed by Sharpe fails to capture the real economic value that active managers provide.

8.1. Ranking Anomalies

The finance literature has spent considerable effort documenting potential return anomalies. Once a candidate anomaly is proposed, two important questions tend to arise. First, how robust is the phenomenon? That is, does the return pattern survive in out-of-sample tests? Second, is it possible to provide a risk-based theoretical explanation of the observed return pattern, thereby raising the possibility that the return pattern under consideration is in fact not anomalous? Before spending effort answering these undoubtedly relevant questions, one may wonder if it really matters from a real economic point of view whether or not the observed return pattern is truly present and/or anomalous. Our framework is designed to answer this question and can thus help classify candidate anomalies into those with substantial real economic consequences and those without. The most robust way of addressing this question is to estimate our model using the data related to the anomaly at hand. Our framework also suggests that a newly discovered candidate anomaly can be quickly evaluated based on reduced-form measures such as the persistence of the anomaly, the amount of capital it affects, and whether it applies to high Tobin's q firms.

8.2. Anomaly Factors

Our model reveals that two of the most prominent asset pricing anomalies uncovered in the empirical finance literature — the value anomaly and the investment anomaly — emerge endogenously when firms face idiosyncratic firm-specific alpha processes. Whereas part of the literature interprets these anomalies as distinct phenomena, we show that cross-sectional sorts on investment and Tobin's q naturally yield noisy measures of a single firm-specific source of mispricing. These sorting variables further exhibit low correlations with each other (see Section 6.4) and different degrees of persistence. The Markov matrix of an anomaly's sorting variable (see Table 1) thus also does not directly reveal the persistence of the underlying alpha process. Instead, the persistence of sorting variables is jointly determined by firm technology and mispricings, which is captured by our dynamic model.

8.3. Sharpe's Critique

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.[25] This argument has been further extrapolated to imply that active managers therefore cannot add value. 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. Absent alphas, firms' investment decisions 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. There is a free-riding problem that allows passive investors to benefit from the price corrections induced by active investors

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

(similar to the free-rider problem in Grossman and Hart, 1980). By simply comparing the performance of active and passive investors (the financial arithmetic), the gains from altering real economic outcomes are not taken into account. This raises the possibility that active management should not be discouraged.

9. Conclusion

We quantify the magnitude of allocational distortions associated with widely studied cross-sectional asset pricing anomalies, taking the view that firms use (dis)information encoded in market prices when making real investment decisions (Hayek, 1945). As a methodological contribution, we introduce a novel lumpy investment model that can incorporate the inherent dynamic nature of such anomalies. The advantage of our framework is that it yields closed-form solutions for the entire cross-sectional distribution of firm dynamics, a necessary ingredient to estimating the aggregate distortions associated with cross-sectional mispricings. As such, our objective is markedly different from the literature that exclusively focuses on the financial market aspects of such anomalies.

We find that the aggregate real distortions associated with cross-sectional anomalies can be substantial, raising the possibility that financial intermediaries that can reduce and/or eliminate such market imperfections could provide substantial value added to the economy. In that respect, our paper contributes to the debate on the role and optimal size of the financial sector, and sheds light on the potential benefits of price discovery and the relevance of policies affecting it (e.g., short-sale constraints).

Even though we find that financial intermediaries could potentially add significant value to the economy by resolving anomalies, we do not show that they are currently engaged in that activity. We have provided multiple reasons why alphas alone are poor measures of real inefficiencies. Financial intermediaries that simply chase high alpha opportunities with low persistence (e.g., momentum) may therefore not be particularly important for improving real allocations. Further, it is unclear how large cross-sectional anomalies would be absent financial intermediaries that trade on alphas. That said, time series changes and secular trends suggest that financial intermediation may have had a material impact historically

(McLean and Pontiff, 2016, Bai, Philippon, and Savov, 2016). Finally, our framework can be used going forward to quantify the gains from eliminating other types of mispricings. Examples include aggregate stock market mispricings and the remainder of the large set of proposed cross-sectional anomalies, including those affecting debt prices.

Appendix

A. Data Description

The data sources that we use are standard in the anomalies literature. We use the CRSP Compustat merged database for the accounting variables and the CRSP data for returns. The main difference with the existing literature is that we focus on total firm value instead of merely the market value of equity. We thus define the Book-to-Market ratio as the book value of total assets over the market value of assets as proxied by the sum of the market value for equity (i.e., the product of shares outstanding and the stock price) and the book value of debt. We use this valuation ratio to sort stocks into portfolios as reported in Table 1. The equity mispricing measures reported in that table still need to be adjusted though. Given that we focus on the total market capitalization of firms, we need to factor debt into our analysis and thus adjust the mispricing measures. Let R^A_{it} denote the return on assets of decile portfolio i at time t. We compute this return as follows:

$$R^A_{it} = we_{it-1} R_{it} + (1 - we_{it-1}) R^D_t, \tag{25}$$

where we_{it-1} is the sum of the market capitalization of equity of all firms in decile i at time $t-1$, as a fraction of the total market value (equity plus book value of debt) of decile i at time $t-1$, and where R^D_t is the return on debt as proxied by the return on the Barclays Baa intermediate debt portfolio. We then compute the market portfolio as the weighted average over all deciles of these returns on assets, where the weights are the total market values (equity plus book value of debt) of each decile portfolio as a fraction of the total market value of all deciles. We then recompute CAPM alphas as before using these decile portfolio returns and this market portfolio return. Because we are not assuming that the debt is equally mispriced as the equity, the practical implication of these computations is that alphas on assets are scaled-down versions of the alphas on equity. The value spread (the difference between the alphas of the tenth decile and the first decile) for equity over our sample period is 9%, whereas it is just 5.4% when using the debt-adjusted returns. We explore the implications of the magnitude of the value spread in Figures III and VI.

B. 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
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

Table 3

Full Markov matrices. The two panels report the annual Markov matrices of decile assignments when rank-ordering firms by their Book-to-Market ratio and investment (percentage change in total assets).

C. Hamilton-Jacobi-Bellman Equation

The Hamilton-Jacobi-Bellman equation associated with the maximization problem in (18) is given by:

$$
\begin{aligned}
0 = \max_{\{i^+,i^-\}\geq 0} [&\pi(\kappa, z, Z, Y) - (r_f(Z) + \alpha(z, Z))V(\kappa, z, Z, Y) \\
&+ \frac{i^+}{(e^{\Delta} - 1)}(V(\kappa + 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\
&+ \frac{\delta + i^-}{(1 - e^{-\Delta})}(V(\kappa - 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\
&+ \Lambda_Z(Z)\mathbf{V}^Z(z, \kappa, Y) + \Lambda_z(Z)\mathbf{V}^z(Z, \kappa) \\
&+ V_A A\mu(Z) + \frac{1}{2}V_{AA}A^2\sigma(Z)^2 - V_A A\sigma(Z)\nu(Z)],
\end{aligned} \tag{26}
$$

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. Given the conjecture that $V(\kappa, z, Z, Y) = Y \cdot \tilde{V}(\kappa, z, Z)$, it can be verified that the equation scales with Y. Note that α can be a function of both the aggregate and firm-specific states, although in our estimation we specify alpha as a function of a firm-specific state only, allowing us to focus on a purely cross-sectional phenomenon.

D. Cross-sectional Distributions

Let $\mathbf{n}_t$ denote the vector of length $N = N_\kappa \cdot N_Z \cdot N_z$ containing the mass of firms in each Markov state at time t. Let $\mathbf{pr}_t = \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}$ denote the corresponding vector of probabilities of the cross-sectional distribution of firms across all N states. The law of motion for $\mathbf{pr}_t$ is given by:

$$
d\mathbf{pr}_t = \frac{d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} - \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}\frac{\mathbf{1}'d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}. \tag{27}
$$

The expected change in the cross-sectional distribution is thus given by:

$$\mathbb{E}_t\left[d\mathbf{pr}_t\right] = \mathbb{E}_t\left[\frac{d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} - \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}\frac{\mathbf{1}'d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}\right] = (\mathbf{I}_N - \mathbf{pr}_t\mathbf{1}')\Lambda'\mathbf{pr}_t dt, \tag{28}$$

where we use the fact that $\mathbb{E}[d\mathbf{n}_t] = \Lambda'\mathbf{n}_t dt$, and where $\mathbf{I}_N$ is an identity matrix of size $N \times N$. The stationary cross-sectional distribution is defined as the vector of probabilities $\widehat{\mathbf{pr}}$ starting from which there is no expected change in the distribution, that is,

$$\mathbb{E}_t\left[d\mathbf{pr}_t | \mathbf{pr}_t = \widehat{\mathbf{pr}}\right] = \mathbf{0}. \tag{29}$$

Thus, $\widehat{\mathbf{pr}}$ solves the following system of equations (in addition to $\mathbf{1}'\widehat{\mathbf{pr}} = 1$):

$$(\Lambda' - \mathbf{I}_N\mathbf{1}'\Lambda'\widehat{\mathbf{pr}})\widehat{\mathbf{pr}} = \mathbf{0}. \tag{30}$$

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 firms entering and exiting the system (see footnote 14), that is:

$$\Lambda(s, s) = -\sum_{s' \neq s} \Lambda(s, s') + h_{entry}(s) - h_{exit}(s). \tag{31}$$

The condition

$$\frac{\mathbf{1}'\mathbb{E}[d\mathbf{n}_t]}{dt} = \mathbf{1}'\Lambda'\mathbf{n}_t = 0 \tag{32}$$

implies that there is no expected change in the total mass of firms and that $\mathbf{1}'\Lambda'\widehat{\mathbf{pr}} = 0$. As discussed in footnote 14, we assume that entry and exit rates are identical in all states s (that is, $h_{entry}(s) = h_{exit}(s)$), such that this condition is satisfied. Equation (30) then simplifies and $\widehat{\mathbf{pr}}$ is the solution to the linear system:

$$\begin{pmatrix} \Lambda' \\ \mathbf{1}' \end{pmatrix} \widehat{\mathbf{pr}} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}. \tag{33}$$

That is, the vector of stationary probabilities, $\widehat{\mathbf{pr}}$, is the left normalized eigenvector of Λ associated with the eigenvalue 0.

The conditional distributions can also be characterized easily. If the current cross-sectional distribution of firms across states is given by $\mathbf{pr}_0$ then the conditional distribution of the cross-section after a time period of length τ is obtained based on the matrix exponential:

$$\mathbf{pr}_\tau = \mathbf{pr}_0 e^{\Lambda \cdot \tau}. \tag{34}$$

E. Market Valuations and No Arbitrage

In this section we discuss the generality of the types of distortions that the pricing equation (15) introduced in Section 4.3 can accommodate. Specifically, it can capture at least two types of distortions in agents' 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

$$\Pr[s_\tau | s_t] \frac{m(Y_\tau, Z_\tau)}{m(Y_t, Z_t)}, \tag{35}$$

where $\Pr[s_\tau | s_t]$ is a state probability corresponding to the perfect Bayesian expectation operator mentioned above. Beliefs encoded in market prices (15) are informationally inefficient in that agents' expectations of a firm's net payout at time τ in state s_τ are given by:

$$\mathbb{E}[e^{-\int_t^\tau \alpha(s_k) dk} | s_\tau, s_t] \cdot \pi(s_\tau). \tag{36}$$

In contrast, an observer outside of this economy that efficiently processes all public information would expect $\pi(s_\tau)$. 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, that is, (Y, Z). In this case, (15) 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:

$$\mathbb{E}[e^{-\int_t^\tau \alpha(Z_k)dk}|Z_\tau, Z_t] \cdot \Pr[s_\tau|s_t] \cdot \frac{m(Y_\tau, Z_\tau)}{m(Y_t, Z_t)}. \tag{37}$$

References

Arrow, K. J. (1968): *Optimal Capital Policy with Irreversible Investment.* James N. Wolfe, ed., Value, Capital and Growth, Essays in Honor of Sir John Hicks, Edinburgh: Edinburgh University Press.

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.

Ball, R., and P. Brown (1968): "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research*, 6(2), 159–178.

Baron, M., and W. Xiong (2016): "Credit Expansion and Neglected Crash Risk," *Quarterly Journal of Economics*, forthcoming.

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.

Berk, J. B. (1995): "A Critique of Size-Related Anomalies," *Review of Financial Studies*, 8(2), 275–286.

Bernard, V. L., and J. K. Thomas (1990): "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings," *Journal of Accounting and Economics*, 13(4), 305 – 340.

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.

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *The Journal of Finance*, 63(4), 1609–1651.

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

Dessaint, O., T. Foucault, L. Fresard, and A. Matray (2016): "Ripple Effects of Noise on Corporate Investment," CEPR Discussion Papers 11081, C.E.P.R. Discussion Papers.

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.

Fahlenbrach, R., R. Prilmeier, and R. M. Stulz (2016): "Why Does Fast Loan Growth Predict Poor Performance for Banks?," Working Paper 22089, National Bureau of Economic Research.

Fairfield, P. M., J. S. Whisenant, and T. L. Yohn (2003): "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing," *The Accounting Review*, 78(1), 353–371.

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

Foucault, T., and L. Fresard (2014): "Learning from peers' stock prices and corporate investment," *Journal of Financial Economics*, 111(3), 554–577.

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

Gomes, J. F. (2001): "Financing Investment," *American Economic Review*, 91(5), 1263–1285.

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.

Grossman, S. J., and O. D. Hart (1980): "Takeover Bids, the Free-Rider Problem, and the Theory of the Corporation," *Bell Journal of Economics*, 11(1), 42–64.

Hall, R. E. (2003): "Corporate Earnings Track the Competitive Benchmark," Working Paper 10150, National Bureau of Economic Research.

Harvey, C., Y. Liu, and H. Zhu (2016): "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29(1), 5–68.

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.

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.

Novy-Marx, R. (2013): "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108(1), 1–28.

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.

Pindyck, R. S. (1988): "Irreversible Investment, Capacity Choice, and the Value of the Firm," *The American Economic Review*, 78(5), 969–985.

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.

Stein, J. C. (1996): "Rational Capital Budgeting in an Irrational World," *The Journal of Business*, 69(4), 429–455.

Titman, S., K. C. J. Wei, and F. Xie (2004): "Capital Investments and Stock Returns," *Journal of Financial and Quantitative Analysis*, 39(04), 677–700.

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.