

Real Anomalies

JULES H. van BINSBERGEN CHRISTIAN C. OPP*

Journal of Finance, Forthcoming

ABSTRACT

We examine the importance of cross-sectional asset pricing anomalies (alphas) for the real economy. We develop a novel quantitative model of the cross-section of firms that features lumpy investment and informational inefficiencies, while yielding distributions in closed form. Our findings indicate that anomalies can cause material real inefficiencies, raising the possibility that agents that help to eliminate them add significant value to the economy. The framework reveals that the magnitude of alphas alone is a poor indicator of real implications, and highlights the importance of alpha persistence, the amount of mispriced capital, and the Tobin's q of firms affected.

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

*van Binsbergen is with the Wharton School of the University of Pennsylvania and the NBER and Opp is with the Wharton School of the University of Pennsylvania. The authors do not have any relevant or material financial interests that relate to the research described in this paper. We thank the editor Stefan Nagel and an anonymous referee for their helpful and constructive feedback. Moreover, we thank Andy Abel, Malcolm Baker (discussant), Markus Brunnermeier, John Campbell, Andrea Eisfeldt, Nicolae Gârleanu, Vincent Glode, Itay Goldstein, João Gomes, Urban Jermann, Tom Sargent, Sydney Ludvigson, Hanno Lustig, Stijn van Nieuwerburgh, Dimitris Papanikolaou (discussant), Thomas Philippon (discussant), Krishna Ramaswamy, Rob Stambaugh, Jeremy Stein, Jessica Wachter, Tony Whited, Amir Yaron, and seminar participants at the 7th Advances in Macro-Finance Tepper-LAEF Conference, Berkeley Haas, BI-SHoF Conference on Financial Econometrics, 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.

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. 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 the economy. In this paper we address this question and find that the economy-wide real distortions can be substantial.

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 — wedges in the form of cross-sectional asset mispricings — that is disciplined by alphas estimated in the finance literature. We propose a tractable dynamic framework that maps distortions in agents’ subjective beliefs to these alphas, and use the model to evaluate the aggregate implications for the real economy.

Any study aiming to quantify misallocations can do so only conditional on a model that is imposed by the econometrician. In particular, the stance on what constitutes the *efficient benchmark* determines what is ultimately identified as a wedge. In our context, informational inefficiencies encoded in prices can only be measured based on some postulated asset pricing model, which is known as the *joint-hypothesis problem* in the empirical asset pricing literature (Fama, 1970, 1991). Rather than solving the important problem of determining which model is the correct efficient benchmark, we aim to provide a flexible methodology that allows the assessment of real distortions associated with alphas computed under a variety of standard asset pricing models. As an example, we use the CAPM to identify alpha wedges, but other asset pricing models can be accommodated.

Moreover, our methodology can be used by economists believing that market prices are always informationally efficient. Conditional on this premise, our framework identifies how quantitatively important various risk factors are for real economic activity. This is relevant for the asset pricing literature's contribution to macroeconomics, as it allows gauging how important it is that a model's implied stochastic discount factor captures the variation in expected returns found in the data. The unifying insight of our paper, independent of one's view on the informational efficiency of prices, is that not all anomalies or risk factors are created equal; some are significantly more important than others from the perspective of their effects on the real economy. For expositional clarity, we will by default interpret our results conditional on the premise that alphas do measure informational inefficiencies, but we also provide an in-depth discussion of how our methodology can be used under the premise that they capture omitted risk factors. Ultimately, the purpose of this paper is to initiate a discussion on which documented return patterns are relevant for aggregate economic activity.

While alphas indicate mispricings conditional on the postulated asset pricing model, we find that the magnitude of alphas alone is generally a poor measure of the potential associated real economic distortions. First, alphas only represent expected changes in the level of mispricings over a given period. Mispricing is an inherently dynamic phenomenon — as alphas are realized over time, mispricings build up and/or resolve, leaving firms only temporarily affected by price distortions. It is therefore 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 relatively small but persistent alpha, instead of a very short-lived large alpha. Whereas the investment management literature primarily cares about the magnitude of an alpha (after trading costs) regardless of its persistence, real corporate investment decisions are little affected by short-lived mispricings. Second, as alphas measure percentage changes in mispricing, they also do not give an accurate representation of the economy-wide value of firms affected. 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 magnitudes of

alphas alone cannot be used to measure the economic importance of an anomaly. Therefore, information on the market capitalization of affected firms should also be taken into account. Thirdly, and most importantly, it is not clear from studying alphas in isolation to what extent mispricings translate into real investment distortions and surplus losses. We show for example that firms with high Tobin's q are those that respond more to mispricings than firms with low Tobin's q , implying that mispricings of growth firms tend to lead to 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), these firms' real responses to mispricings are also dampened significantly.

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.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 (see Internet Appendix A), this 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 capital today.

In order to assess real effects 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 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 potential real inefficiencies. Our results reveal that informational inefficiencies can be associated with significant real effects despite a weak investment- q relationship.

Our paper has several important parallels to the existing misallocation literature in macroeconomics. For example, in recent work, Hsieh and Klenow (2009) study wedges in effective input prices, as opposed to wedges in state prices. In Hsieh and Klenow (2009), the relevant prices — i.e., the effective “after-tax” input prices firms use when maximizing — are not directly observable to researchers. However, Hsieh and Klenow (2009) can use the identifying assumption of their static model that absent frictions, firms should face the same input costs. As a result, firms' marginal revenue products (MRPs) of a given input should also be equalized, implying that cross-sectional dispersion in those MRPs measures wedges.⁴ Instead, the wedges we are interested in can be estimated using the condition that expected returns on assets with the same riskiness (as defined by an asset pricing model) should be equalized.⁵ In summary, whereas Hsieh and Klenow (2009) use the difference between observed and model-implied MRPs to measure input price wedges, we use the difference between observed and model-implied security returns to measure state price wedges. As highlighted above, a common building block of studies quantifying misallocations is a model of the efficient benchmark, implying that researchers always test a joint hypothesis.⁶

There is also a large literature in macroeconomics and corporate finance studying external financing frictions that drive wedges between internal and external funds.⁷ 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 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 potential aggregate real value losses associated with commonly studied 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 et al. (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.⁸ 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 et al. (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. Warusawitharana and Whited (2016) measure equity mispricings by the residuals of a hedonic regression, and model a representative firm that (unlike investors) is informed about these mispricings. The authors find that misvaluation shocks help alleviate financial constraints, and market timing of equity issuances by managers increases equity value by up to 3%, relative to a model with no misvaluation shocks.

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. For mispricings to have real effects, it is not a necessary condition that prices have *incremental* explanatory power for investment, after controlling for a wide range of observables. In particular, there may be no effect when other endogenous observable variables included in the analysis already reflect disinformation associated with mispricings. Going beyond the question whether prices have incremental explanatory power, our model does not even predict that standard q regressions explain investment well in the first place (see Section III.E.1 for details). Nonetheless, prices in our model *are* essential for investment decisions, as they are used to determine the NPVs of investment plans.

Moreover, it is irrelevant for our analysis if decision makers at firms learn about existing public information by using market prices as a summary statistic, or if they themselves process all public information in the same (potentially inefficient) way as financial market participants do. A "real anomaly" only requires that real investment decisions end up max-

imizing an informationally inefficient market value. Even if firms independently process public information in the same flawed way as financial markets do, 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 measures of the amount of new information in stock prices have a strong positive effect on the sensitivity of corporate investment to stock prices. 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. 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. In contrast, 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 available information.

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. Cohen, Polk, and Vuolteenaho (2009) empirically study price levels, rather than abnormal returns. They

find that prices generally tend to be fairly close to those predicted by cash flow betas. In contrast, our dynamic model uses alpha *processes* as inputs and transforms them into *processes* of abnormal price levels. We show that for our analysis of real distortions the full *process* of price distortions is essential — an initial price level distortion is not a sufficient statistic.

An intriguing starting point for our analysis is that data on prices in financial markets indicate a cross-sectional relation between firm investment and alphas computed under commonly used benchmark asset pricing models, such as the CAPM and the Carhart (1997) four-factor model.⁹ While our empirical application focuses on non-financial firms, similar empirical patterns apply for financial firms; Fahlenbrach, Prilmeier, and Stulz (2016) find that banks extend significantly more credit when they are overpriced relative to benchmark asset pricing models.¹⁰ The fact that investment responds to discount rates is not surprising and does not by itself indicate any inefficiency. Yet, under the joint hypothesis of a given asset pricing model and market efficiency, no variable, including investment, should be related to average returns, after controlling for exposures to the model’s priced factors. Moreover, conditional on the view that alphas indeed measure informational inefficiencies, firm investment responding to the *abnormal* components of expected returns constitutes a *real* anomaly. Our framework shows that significant real distortions may arise even when the cross-sectional investment-alpha relation is somewhat weaker than it is in the data.

I. A Simple One-Period Example

To better understand why alpha measures by themselves are not informative about the associated *real* economic distortions, 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. Let \mathbb{E} denote the expectations under the objective probability measure that efficiently incorporates all available information, and let \mathbb{E}^* denote expectations under agents’ homogenous subjective

beliefs, which may deviate from the objective beliefs. In addition, let m denote agents' marginal utility of consumption. The value that agents place on the firm's cash flows at time 0 is given by:

$$\begin{aligned} V_0 &= \pi_0 + \mathbb{E}_0^* \left[\frac{m_1}{m_0} \pi_1 \right] \\ &= \pi_0 + \mathbb{E}_0 \left[\frac{m_1}{m_0} e^{-\alpha_0} \pi_1 \right]. \end{aligned} \quad (1)$$

The second line makes two adjustments that offset each other. First, we change the probability measure of the expectation operator by moving from subjective expectations to objective expectations of an outside observer. Second, we adjust the effective discount rate by introducing α_0 as a log price distortion. This price wedge is thus due to belief distortions regarding the joint distribution of discount rates ($\frac{m_1}{m_0}$) and cash flows (π_1). For $\alpha_0 \neq 0$, the financial market is informationally inefficient in that the market price of the firm V_0 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 log 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 informationally efficient price given the distorted firm policies:

$$V_0^d = \pi_0(\alpha_0) + \mathbb{E}_0 \left[\frac{m_1}{m_0} \pi_1(\alpha_0) \right] \quad (2)$$

to the informationally efficient price given the undistorted firm policies:

$$V_0^u = \pi_0(0) + \mathbb{E}_0 \left[\frac{m_1}{m_0} \pi_1(0) \right]. \quad (3)$$

Note that in equations (2) and (3) we evaluate prices under the objective probability measure.

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. Moreover, the model should be flexible and easy-to-solve, facilitating estimation based on cross-sectional moments. Before formally introducing such a model, we first explore several reduced-form characteristics of alphas in the next section.

II. Reduced-Form Estimates

As highlighted in the previous section, the alpha estimates the finance literature has documented are an important input for our model: alphas reflect wedges in discount rates and/or expected growth rates 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 empirical anomalies literature generally treats the documented alphas as the end point of the analysis, whereas we use them as a key input to our framework.

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 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:

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

The results are summarized in Panels A and B of Table I. 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 tenth decile and the alpha in the first decile of the Book-to-Market sort, and equals 73bp per month, or about 9% per annum over this sample period.

[Place Table I about here]

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

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

The weights $we_{i,j}$ are summarized in Panel C of Table I. 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). Yet, 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:

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

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 I and show that debt largely offsets the differences in market capitalization across the Book-to-Market 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. These opposite results also indicate that sorts on investment and Book-to-Market (the inverse of Tobin's q) lead to significantly different firm rankings.

While accounting for debt can have a significant effect on the cross-sectional distribu-

tion of market capitalization across Book-to-Market 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. We will estimate alphas effectively assuming no mispricing on the debt portion (see Appendix A for further details).¹¹ This is a conservative assumption if mispricings of a firm's debt tend to be *directionally* consistent with the mispricings of its equity (for example, if beliefs about a firm's equity payoffs are too optimistic, then they are also too optimistic for the same firm's debt payoffs). When assessing the results of our model we will perform sensitivity analyses with respect to this assumption.

Finally, we explore one last important feature, which is the persistence of the mispricing. One rough way to assess the persistence of alphas is to evaluate the persistence of the sorting variable used.¹² Note, however, that the sorting variable is only a noisy proxy for firms' underlying alphas, implying that depending on the persistence of the noise, the results can be biased upwards or downwards. We compute annual Markov transition matrices that summarize for each anomaly how firms migrate across deciles. The diagonal elements of these matrices are summarized in Panel E of Table I.¹³ 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 measure is lower for the investment sorts (about 1/5).

III. 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. Structural parameters of the model are affected by a set of exogenous state variables, which are described in detail in Section III.B below. For notational convenience, we will at times omit parameters' functional dependence on these states elsewhere in the model description.

A. 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.

Log-capital $k_t = \log[K_t]$ takes values in a discrete set Ω_k , the n_k elements of which constitute an equidistant grid with lower bound $\min\{\Omega_k\} \in \mathbb{R}$ and grid increments of size $\Delta_k > 0$.¹⁴ By choosing Δ_k small enough, this specification can approximate a continuous support for capital arbitrarily well.¹⁵ Moreover, the discrete state space structure allows capturing the lumpy nature of investment observed in the data.

Let $N_{k,t}^+$ and $N_{k,t}^-$ denote Poisson processes keeping track of successful capital acquisitions and divestments, and let $N_{k,t}^\delta$ denote a Poisson process for capital depreciation shocks. The corresponding capital evolution equation is given by:

$$dk_t = \Delta_k \cdot (dN_{k,t}^+ - dN_{k,t}^- - dN_{k,t}^\delta). \quad (7)$$

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

$$i_t^+ \equiv (e^{\Delta_k} - 1)\mathbb{E}_t [dN_{k,t}^+] \geq 0, \quad (8)$$

and stochastically succeeds in finding an opportunity to upgrade its capital to the next-higher level, that is, by a log change of size Δ_k , with a Poisson intensity $i_t^+ / (e^{\Delta_k} - 1)$. Throughout,

\mathbb{E}_t denotes the expectation operator under the objective probability measure that efficiently incorporates all information available in the economy as of date t . As a result of its search efforts, a firm incurs search costs at a rate that is quadratic in i_t^+ :

$$\theta_t^+ (i_t^+)^2 K_t, \quad \text{with } \theta_t^+ > 0. \quad (9)$$

Once an investment opportunity is found, a firm requires p_t^+ units of output to install one unit of new capital, and is constrained to the scale of the given opportunity. When a firm's capital stock reaches the upper bound $e^{\max\{\Omega_k\}}$, search is assumed to be ineffective in delivering additional investment opportunities. By choosing n_k high enough, this restriction will be immaterial, as optimal investment will be zero above some endogenous threshold for capital.

Similarly, each firm chooses its expected disinvestment rate

$$i_t^- \equiv (1 - e^{-\Delta_k}) \mathbb{E}_t [dN_{k,t}^-] \geq 0, \quad (10)$$

and finds an opportunity to divest capital by a log change of size Δ_k with a Poisson intensity $i_t^- / (1 - e^{-\Delta_k})$. Searching for divestment opportunities causes a firm to incur search costs at a rate:

$$\theta_t^- (i_t^-)^2 K_t, \quad \text{with } \theta_t^- > 0. \quad (11)$$

Once an opportunity to divest is found, a firm can make this divestment and obtain p_t^- units of output per unit of capital deinstalled. Divestments are infeasible when capital reaches the lower bound $e^{\min\{\Omega_k\}}$. By specifying $\min\{\Omega_k\}$ 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 with a Poisson intensity $\delta / (1 -$

$e^{-\Delta}$), except at the lower bound $\min\{\Omega_k\}$ where the Poisson intensity is zero. Thus, for $k_t > \min\{\Omega_k\}$, the expected depreciation rate is δ , and the expected growth rate of capital is given by:

$$\frac{\mathbb{E}_t[dK_t]}{K_t} = (i_t^+ - i_t^- - \delta)dt. \quad (12)$$

We introduce the generator matrix Λ_k^i that collects the transition rates between all capital states Ω_k , where the superscript i indicates the dependence of this matrix on the investment controls $i = (i^+, i^-)$.

B. Exogenous Processes

In this subsection, we describe the dynamics of the exogenous processes in the economy.

Macroeconomic conditions. The state Z governs persistent, mean-reverting variation in 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 and has the generator matrix Λ_Z , which collects transition rates between states $Z \in \Omega_Z$. Let $\mathbf{N}_{Z,t}$ denote a matrix that collects counting processes that keep track of all jumps between the states $Z \in \Omega_Z$. To balance increases in the off-diagonal elements, the diagonal elements of this matrix count down by one each time a given state Z is left for another state Z' . We also define the matrix $\mathbf{M}_{Z,t}$ collecting the compensated processes:

$$d\mathbf{M}_{Z,t} = d\mathbf{N}_{Z,t} - \Lambda_Z dt. \quad (13)$$

Throughout the paper, for any given matrix \mathbb{M} , we will use the notation $\mathbb{M}(x)$ to indicate the x -th row of that matrix, and $\mathbb{M}(x, x')$ to represent the (x, x') -element.

Aggregate trend. The state Y captures an aggregate trend in the economy that follows a

geometric Brownian motion:

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

A firm's technology processes $\{A_t, c_{f,t}, p_t^+, p_t^-, \theta_t^+, \theta_t^-\}_{t=0}^\infty$ are all assumed to scale linearly with Y_t , capturing common growth in these variables. Going forward, a tilde indicates that a variable is scaled by Y , for example, $\tilde{A} = A/Y$. In addition, we collect the aggregate state variables in the vector $S = (Y, Z)$.

Firm-level productivity and subjective beliefs. Continuous-time Markov chains influence firm-level productivity and deviations in agents' subjective beliefs from the objective probability measure. The log of scaled productivity $a_t = \log[\tilde{A}_t]$ follows an n_a -state continuous-time Markov chain that takes values in the set Ω_a . The dynamics of a_t are given by:

$$da_t = \Delta_{a_t}^+ \cdot dN_{a,t}^+ - \Delta_{a_t}^- \cdot dN_{a,t}^- \quad (15)$$

where $\Delta_{a_t}^+ > 0$ and $\Delta_{a_t}^- > 0$ generally depend on the current value of a_t , and where $N_{a,t}^+$ and $N_{a,t}^-$ are Poisson processes with arrival intensities $h_a^+(a_t)$ and $h_a^-(a_t)$, respectively. We collect these objective transition rates in the generator matrix Λ_a , and impose the technical conditions that (1) the transition rates of moving to the boundaries of the set Ω_a (that is, $\min\{\Omega_a\}$ and $\max\{\Omega_a\}$) are zero, and (2) that at date 0, no firms are in these boundary states. In contrast, under agents' subjective beliefs, these boundaries can potentially be reached with positive probability. We denote by Λ_a^* the generator matrix under agents' subjective beliefs. This generator matrix is allowed to depend on other state variables, so that beliefs about productivity dynamics can be affected by variables other than the current a -state itself. However, for notational simplicity, we generally do not indicate the dependence of Λ_a^* on these other states. As will become clear later, it suffices for our purposes to restrict attention to the possibility of belief distortions regarding the dynamics of firm-

level productivity a . For this reason, we suppose that agents' beliefs regarding all other processes in the economy coincide with the objective beliefs.

As further described in the next section, an important variable imposing restrictions on agents' subjective beliefs is the firm-specific state α . This state follows an n_α -state continuous-time Markov chain taking values in the set Ω_α . 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 processes. The corresponding lack of correlation with the aggregate state variables (Y, Z) is consistent with our focus on a purely cross-sectional phenomenon. As a technical condition, we impose that $\alpha = 0$ when firm productivity reaches the boundaries of the set Ω_α , which occurs with probability zero under the objective probability measure.

C. Beliefs and Market Valuations

The before-mentioned joint-hypothesis problem implies that an econometrician may find cross-sectional alphas for two reasons: (1) prices are not informationally efficient, and/or (2) the econometrician's model for prices is misspecified. While our paper naturally cannot resolve the joint hypothesis problem, our methodology can be useful to researchers viewing either reason for the emergence of alphas as the relevant one. For the purposes of the exposition, we first suppose that the model of the econometrician is accurate, and that alphas indeed measure informational inefficiencies. In Section VI, we then address how our methodology's results can be used under the alternative view that prices efficiently incorporate public information, but the econometrician's model is misspecified.

Under the first view, the econometrician is viewed as an outside observer of the economy that efficiently processes public information and forms rational beliefs that correspond to the objective probability measure. In contrast, agents in the economy have homogenous, distorted beliefs that do not efficiently incorporate all public information, and prices are set based on these subjective beliefs. In the literature, belief distortions are common in models where agents are subject to biases and disagreement,¹⁶ or face costs or constraints when

processing information.¹⁷¹⁸ As introduced in Section I, we use \mathbb{E}^* to denote the mathematical expectation under agents' homogenous subjective beliefs.

As standard in the literature, agents in the economy value a firm's stream of future after-tax net payouts by discounting future payments:

$$\mathbb{E}_t^* \left[\int_t^\infty \frac{m_\tau}{m_t} d\Pi_\tau \right], \quad (16)$$

where m represents the stochastic discount factor (SDF) reflecting agents' marginal utility, and where Π_t denotes the cumulative net payout of the firm between dates 0 and t .

As our methodology uses cross-sectional alphas as a direct empirical input, we are interested in belief distortions that imply that a firm in state α_t indeed locally generates an expected abnormal return equal to α_t under the objective probability measure. This restriction on belief distortions implies that the following identity holds:

$$\mathbb{E}_t^* \left[\int_t^\infty \frac{m_\tau}{m_t} d\Pi_\tau \right] = \mathbb{E}_t \left[\int_t^\infty \frac{m_\tau}{m_t} \underbrace{e^{-\int_t^\tau \alpha_u du}}_{\text{Mispricing Wedge}} d\Pi_\tau \right], \quad (17)$$

which is the dynamic equivalent of equation (1). That is, the subjective expectations \mathbb{E}^* are effectively discounting or inflating payments by a “mispricing wedge” relative to the objective expectations \mathbb{E} . A firm's current α -state thus not only determines the abnormal *return* over the next instant, but also affects the current *level* of mispricing, due to persistence in the α -process. For example, if a claim to a stochastic cash flow in one year is currently underpriced by 10% — that is, the mispricing wedge equals 0.9 — then abnormal returns on that claim accumulate on average to +10% over this time period. The path of the resolution of mispricing is, however, stochastic.

Equation (17) also highlights that, as time approaches the payment date of a cash flow, agents' subjective expectations regarding this cash flow converge to the objective expectations — the mispricing wedge converges to one. Thus, agents' expectations of net payout

$d\Pi_t$ are undistorted at time t ,¹⁹ that is,

$$\frac{\mathbb{E}_t[d\Pi_t]}{dt} = \frac{\mathbb{E}_t^*[d\Pi_t]}{dt}. \quad (18)$$

For notational convenience, we define the conditional expected net payout rate $\pi_t \equiv \frac{\mathbb{E}_t[d\Pi_t]}{dt}$.²⁰

SDF dynamics. We specify a flexible process for m that can capture the pricing properties of a variety of benchmark asset pricing models, such as the CAPM, long run risk models, and rare disaster models. Given our focus on cross-sectional alphas, we assume that agents have undistorted beliefs about the evolution of the aggregate state variables $S = (Y, Z)$, which govern the dynamics of m . The SDF follows a general Markov-modulated jump diffusion process:

$$\frac{dm(S_t)}{m(S_t)} = -r_f(Z_t) dt - \nu(Z_t) dB_t + \sum_{Z' \neq Z_t} (e^{\phi(Z_t, Z')} - 1) d\mathbf{M}_{Z,t}(Z_t, Z'), \quad (19)$$

where r_f denotes the risk free rate, ν is the price of risk for aggregate Brownian shocks, and $\phi(Z, Z')$ determines the jumps in m conditional on a change in the state Z . Let $\bar{\Lambda}_Z$ denote the generator matrix under the objective risk neutral measure, which collects the risk neutral transition rates $\bar{\Lambda}_Z(Z, Z') = e^{\phi(Z, Z')} \Lambda_Z(Z, Z')$.

For one part of our analysis, general equilibrium effects could in principle be quantitatively relevant. When we use our methodology under the premise that alphas indeed represent informational inefficiencies, we consider a counterfactual economy to evaluate the impact of eliminating these informational inefficiencies. The associated adjustments in agents' real production decisions would, in general equilibrium, also create a feedback effect on consumption dynamics, thus simultaneously affecting the dynamics of marginal utility as well. In Section VII.A, we analyze the potential quantitative impact of such general equilibrium effects on our efficiency gain evaluations and find that the associated effects are of second-order importance. Moreover, under the alternative premise that alphas emerge due to the

omission of relevant risk factors in the SDF specification employed by the econometrician (see Section VI), agents in the actual economy already make decisions under the objective probability measure. Thus, under this alternative view, an efficiency gain does not need to be computed.

D. Firm Objective

Firms take market prices as determined by (16) as given and choose the investment controls $\{i_t^+, i_t^-\}_{t=0}^\infty$ that maximizes their market value at these prices. As discussed in the introduction, the 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 investment and alphas. 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 V.E and VII.

E. Analysis

E.1. Firm Behavior

In this subsection we analyze firm behavior. For a firm's decisions the relevant state is (s, S) , where $s = (k, a, \alpha)$ is a vector collecting the firm-level states, and where the vector of aggregate states $S = (Y, Z)$ was introduced above. Firms dynamically maximize their market value under agents' subjective expectations:

$$V(s_t, S_t) = \max_{\{i_\tau^+, i_\tau^-\}_{\tau=t}^\infty \geq 0} \mathbb{E}^* \left[\int_t^\infty \frac{m_\tau}{m_t} \pi_\tau d\tau \middle| s_t, S_t \right], \quad (20)$$

where the expected after-tax net payout conditional on date- t information is given by:

$$\begin{aligned} \pi_t = & (1 - \tau)(A_t K_t^\eta - (c_{f,t} + p_t^+ i_t^+ + \theta_t^+ (i_t^+)^2 - p_t^- i_t^- + \theta_t^- (i_t^-)^2) K_t) \\ & + (i_t^- - i_t^+ + \delta) p_t^+ \tau K_t. \end{aligned} \quad (21)$$

Equation (21) reflects that firms obtain tax shields from depreciation, proportional cost of production, search cost, and capital divestments below the installment cost p_t^+ . As defined above in Section III.C, π_t represents the *expected* net payout conditional on date- t information that accounts for the stochastic arrival of shocks to capital over the interval $[t, t + dt)$.

Since the processes $\{A_t, c_{f,t}, p_t^+, p_t^-, \theta_t^+, \theta_t^-\}_{t=0}^\infty$ scale linearly with the aggregate trend Y , the value function also scales linearly with Y , that is, $V(s, S) = Y \cdot \tilde{V}(s, Z)$. The Hamilton-Jacobi-Bellman (HJB) equation associated with the maximization problem in (20) implies that $\tilde{V}(s, Z)$ solves the following set of equations for all $(s, Z) \in \Omega_k \times \Omega_a \times \Omega_\alpha \times \Omega_Z$:²¹

$$\begin{aligned} 0 = \max_{(i^+, i^-) \geq 0} & \left\{ \tilde{\pi}(i, k, a) - (r_f(Z) + rp(s, Z) - \mu(Z)) \tilde{V}(s, Z) + \Lambda_k^i(k) \tilde{\mathbf{V}}_k(s, Z) \right. \\ & \left. + \Lambda_a^*(a) \tilde{\mathbf{V}}_a(s, Z) + \Lambda_Z(Z) \tilde{\mathbf{V}}_Z(s, Z) + \Lambda_\alpha(\alpha) \tilde{\mathbf{V}}_\alpha(s, Z) \right\}, \end{aligned} \quad (22)$$

where the notation $\tilde{\mathbf{V}}_x(s, Z)$ indicates a vector that collects the values of the function $\tilde{V}(s, Z)$ evaluated at all $x \in \Omega_x$ while keeping the other arguments fixed, and where rp is a firm's expected return in excess of the risk free rate under agents' subjective beliefs:

$$rp(s, Z) = \sigma(Z) \nu(Z) + (\Lambda_Z(Z) - \bar{\Lambda}_Z(Z)) \frac{\tilde{\mathbf{V}}_Z(s, Z)}{\tilde{V}(s, Z)}. \quad (23)$$

This risk premium compensates investors for exposures to innovations to the Markov state Z and Brownian innovations to the common trend Y . In contrast, an outside observer using the objective probability measure anticipates that the firm will yield an expected excess

return:

$$\alpha + rp(s, Z), \quad (24)$$

where α relates to the belief distortion $(\Lambda_a - \Lambda_a^*)$ as follows:

$$\alpha = (\Lambda_a(a) - \Lambda_a^*(a)) \frac{\tilde{\mathbf{V}}_a(s, Z)}{\tilde{V}(s, Z)}. \quad (25)$$

Under general technical conditions, there always exist subjective beliefs characterized by the generator matrices Λ_a^* that support a given α -process (see proof in Appendix C). Given the equality (25), the firm's optimization problem can be rewritten as follows:

$$0 = \max_{(i^+, i^-) \geq 0} \left\{ \tilde{\pi}(i, k, a) - (r_f(Z) + rp(s, Z) + \alpha - \mu(Z))\tilde{V}(s, Z) + \Lambda_k^i(k)\tilde{\mathbf{V}}_k(s, Z) \right. \\ \left. + \Lambda_a(a)\tilde{\mathbf{V}}_a(s, Z) + \Lambda_Z(Z)\tilde{\mathbf{V}}_Z(s, Z) + \Lambda_\alpha(\alpha)\tilde{\mathbf{V}}_\alpha(s, Z) \right\}. \quad (26)$$

This latter representation of a firm's problem is a more convenient starting point for our analysis, as it eliminates the subjective transition rates Λ_a^* and instead introduces objects for which we have empirical counterparts: the objective physical transition rates Λ_a and local abnormal returns α . While a given α generally does not uniquely pin down the subjective transition rates Λ_a^* , the stochastic process for α contains all information regarding the belief distortion $(\Lambda_a - \Lambda_a^*)$ that is relevant for a firm's real investment decisions and its valuations. That is, if two belief distortions generate the same α -process, they create identical real distortions.

The first-order conditions of the problems (22) and (26) yield a firm's expected investment

and disinvestment rates:

$$i^+(s, Z) = \max \left[\frac{\left(\frac{\tilde{V}(s^{k+}, Z) - \tilde{V}(s, Z)}{(e^\Delta - 1)e^k} - \tilde{p}^+ \right)}{2(1 - \tau)\tilde{\theta}^+}, 0 \right], \quad (27)$$

$$i^-(s, Z) = \max \left[\frac{\left(\frac{\tilde{V}(s^{k-}, Z) - \tilde{V}(s, Z)}{(1 - e^{-\Delta})e^k} + (1 - \tau)\tilde{p}^- + \tau\tilde{p}^+ \right)}{2(1 - \tau)\tilde{\theta}^-}, 0 \right], \quad (28)$$

where s^{k+} and s^{k-} denote vectors that are identical to the state vector s , except that the element corresponding to the state k is increased and decreased by one increment, respectively.²² Conditional on these policy functions, the systems (22) and (26) are linear in $\tilde{V}(s, Z)$, which implies that obtaining exact solutions is straightforward and fast.

Investment and q . Outside of an inaction region, the investment controls are chosen such that the marginal cost of an expected change in capital is equated to the marginal value. In the context of our lumpy investment model, a true marginal q cannot be computed. After all, changes in capital are always discrete. The optimal investment controls instead depend on the ratio of the change in firm value to the change in capital, conditional on a one-increment up- or downgrade in capital; the discrete analogue of a partial derivative of the value function with respect to capital. For ease of exposition, we will simply refer to this ratio as *marginal q* . In our model, like in many other models, *Tobin's q* (that is, average q) generally fails to fully explain investment. The reason it does so in our model is that firms' technology features: (1) stochastic, lumpy investment, (2) decreasing returns, (3) convex, asymmetric search cost for finding (dis)investment opportunities, and (4) differing capital installation and deinstallation efficiencies. Because the amount of output needed to install a unit of capital can be higher than the amount of output obtained from deinstalling a unit (that is, $p^+ > p^-$), there may be an inaction region, that is, a region of the state space where $i^+ = i^- = 0$. In that inaction region, *expected* investment is even unrelated to *marginal q* . In Section V.D, we quantitatively evaluate the relationship between investment and Tobin's

q in our estimated model and compare it to empirical estimates from the literature.

Real distortions and the dynamic nature of firms' decisions. The inter-temporal nature of firms' investment decisions has important implications for the magnitude of real distortions. In our model, firms face inter-temporal investment trade-offs because of frictions faced when adjusting capital. As a result, a firm has to consider the long-term implications of building up capital inside the firm when deciding on its current investment efforts. Biased expectations about *future* productivity then generally cause *current* investment to be distorted. For example, too optimistic expectations lead a firm to inefficiently build up capital in advance, causing real output losses due to overinvestment.

In contrast, if firms could frictionlessly adjust capital in continuous time and rent capital at an undistorted rate, upward-biased expectations about *future* productivity would *not* cause overinvestment. Rather, the firm would simply choose its capital stock date-by-date given the *contemporaneous* productivity A , which is known by the time of the decision and thus not subject to belief distortions. As a result, despite the presence of biases and abnormal financial returns, real inefficiencies would not occur — biased expectations about the future naturally do not distort deterministic short-term investments. At the other extreme, if adjustment costs were infinitely large, optimal firm (dis)investment would be zero, independently of expectations about future productivity, also implying that real decisions would be unaffected.

These insights highlight that the inter-temporal nature of firms' decisions — which require the forecasting and discounting of future cash flows — plays a central role in determining the magnitudes of real distortions. The data we use to estimate our model (see Section IV.A) suggest that firms do face significant frictions to adjusting capital, implying that their investment decisions crucially depend on expectations about the future.

E.2. Stationary 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 E how, for any given policy function, we can compute stationary and conditional distributions in closed-form, which greatly facilitates the estimation and evaluation of the model.

IV. Estimating the Model

A. Calibration and Estimation Procedure

We calibrate the parameters of the macroeconomy using values that are informed by the existing literature.²³ The aggregate Markov state Z can take two values {Boom, Recession}. We then estimate 22 firm-specific parameters using a generalized method of moments (GMM) approach. To facilitate matching the size distribution, we consider two sets of firms that permanently differ with respect to the ranges of their possible productivity values. In particular, the set of possible log-productivity values for the two types of firms differ by some constant shifter \underline{a} , that is, $\Omega_a^2 = \Omega_a^1 + \underline{a}$. The sets Ω_a^1 and Ω_a^2 each contain 11 productivity states that are reached with positive probability under the objective probability measure.²⁴ We estimate the lowest objectively possible log-productivity value for the first set of firms, $\min\{a \in \Omega_a^1 : \Pr[a] > 0\}$, the shifter \underline{a} , the relative masses of the two sets of firms, and firms' transition rates of moving to the next-higher and next-lower productivity state, (h_a^-, h_a^+) . The latter are not state dependent. To reduce the number of free parameters we determine every second value of the a -process increments (Δ_a^-, Δ_a^+) via interpolation and estimate the remaining ones. Firm-level alphas follow a mean-reverting three-state Markov process that is restricted to have an unconditional value-weighted mean of zero.²⁵ We estimate two alpha states (the third is implied by imposing a value-weighted mean alpha of zero), as well as

the transition rates of leaving a given alpha-state for the next-higher and next-lower state. We normalize the scaled capital installation costs \tilde{p}^+ to one. In addition, the model features the six technology parameters $(\tilde{c}_f, \tilde{\theta}^+, \tilde{\theta}^-, \tilde{p}^-, \eta, \delta)$. Overall, we estimate 22 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.²⁶

[Place Figure 1 about here]

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 1, and the calibrated and estimated parameters are listed in Table II. The figure plots the distributions of the moments in the data (the black solid line), the 95% confidence bounds generated from the time-series dimension of 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 I. 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 Tobin's q) endogenously emerges in the model as a *noisy* proxy of a firm's underlying alpha state, as market prices generally respond more quickly to mispricings than capital does.²⁷ Yet most of the cross-sectional variation in Tobin's q in the model is still due to technology shocks that are independent of a firm's α -process (e.g., productivity and depreciation shocks). This also implies that the Book-to-Market ratio and investment, which have a relatively low correlation with each other, can both arise as useful proxies for the underlying firm-level alphas.

[Place Table II about here]

The parameters in Table II 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%.²⁸ 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). The discount on deinstalling used capital ($1 - p^-/p^+$) equals 38.5%, and there are high quadratic search costs for finding opportunities to divest 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).²⁹

The estimated transition rates of the alpha process imply substantial mean reversion, leading conditional alphas over a one year horizon to be substantially closer to zero than

instantaneous (local) alphas. The conditional *one-year* alphas in the three states are -9.2% , -2.0% , and $+1.4\%$, whereas the corresponding *local* alphas reported in Table II are -18.4% , -3.4% , and $+2.3\%$, 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 -2.0% and $+1.4\%$.

V. Counterfactual Analysis

To evaluate the real impact of informational inefficiencies measured by alphas, we analyze how firms' behavior and outcomes adjust once alpha wedges are eliminated. That is, we evaluate the differences between two economies that differ with respect to their informational efficiency, but have the same initial distribution across technology and capital states. In this counterfactual analysis, firms consistently maximize their market values under the prevailing market prices and beliefs of the economy in which they reside. In the initial distorted economy we estimate, firms maximize their value under the then-prevailing distorted subjective expectations \mathbb{E}^* , leading to what we call the *distorted* firm policies. In the counterfactual economy, firms maximize their value using the same technology but under the objective expectations \mathbb{E} , leading to what we call the *undistorted* firm policies. The undistorted firm policies are not chosen in the initial economy, since they are not value maximizing under the then prevailing expectations \mathbb{E}^* and the associated state prices.³⁰

A. Under- and Overinvestment

As a first step, we assess the influence of cross-sectional alphas on investment. In Figure 2 we plot a histogram of the difference between distorted and undistorted (expected) investment rates. The plot shows substantial deviations from undistorted investment policies, with both sizable over- and underinvestment. The very right tail of the distribution is populated

by firms that are in the state α_1 , where overvaluations are most severe. As a result, these firms exhibit the largest overinvestment.

[Place Figure 2 about here]

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

B. Measuring Potential Efficiency Gains

To quantitatively assess the influence of cross-sectional distortions on aggregate efficiency, we first compute the stationary distribution of firms across Markov states for the estimated model where firms maximize under the distorted subjective expectations \mathbb{E}^* . We denote the associated vector of stationary probabilities by $\widehat{\mathbf{p}}^d$. Next, we consider a counterfactual economy where, starting from this stationary distribution, firms permanently switch from maximizing under the distorted probability measure to maximizing under the objective measure. Since for any given firm policy function, we have closed-form solutions for conditional distributions, it is straightforward to account for transition dynamics in our framework.

We introduce the aggregate efficiency gain measure:

$$gain = \frac{\mathbb{E}[\int_0^\infty \frac{m_\tau}{m_0} \pi_\tau^u d\tau | \widehat{\mathbf{p}}^d]}{\mathbb{E}[\int_0^\infty \frac{m_\tau}{m_0} \pi_\tau^d d\tau | \widehat{\mathbf{p}}^d]} - 1, \quad (29)$$

where $\{\pi_t^u\}_0^\infty$ denotes the stochastic stream of *aggregate* net payout when firms maximize under the undistorted expectation operator \mathbb{E} , and $\{\pi_t^d\}_0^\infty$ the stream when firms maximize under \mathbb{E}^* . As in the one-period example in Section I, we evaluate cash flow streams under \mathbb{E} . We will discuss in Section VII.A and Internet Appendix B how this *gain* measure is related to a classic Lucas (1987) type welfare calculation. In these sections, we will also quantitatively assess how general equilibrium effects might affect our efficiency gain estimates.

The *gain* measure in (29) can be interpreted as the perpetual percentage of total firm net payout that society should be 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 the alphas considered.

The alpha process in our analysis is primarily identified by the empirical Book-to-Market alphas and the value-weights of the associated portfolios. As Tobin's q is only a noisy measure of a firm's underlying alpha state (see the discussion in Section IV.A), eliminating the alpha process implies the elimination of the value anomaly in the model, while the opposite implication does not hold. Put differently, even if financial intermediaries were to eliminate the statistical significance of the so-called value spread, this would generally not yield the full efficiency gain we compute.

Figure 3 plots the aggregate *gain* measure defined in equation (29) for different alpha process parameterizations. The parameterizations vary the range of the alpha process, resulting in different levels of model-implied value spreads in a (non-linear) one-to-one mapping. Our baseline estimation, as presented in Table II and Figure 1, corresponds to an annual value spread in the model of about 3.1%, which is more conservative than the 5.4% found in the data (see the fourth Panel in Figure 1 which plots monthly alphas in the model and in the data).

[Place Figure 3 about here]

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 II. Generally, we would expect debt to be less mispriced than equity. But, if one is nevertheless 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 *gain* measures. As a result, we find that the gains from eliminating the considered alpha processes are several percentage points, with estimates ranging between 1.4% and 4.3% depending on the targeted value spread. Moreover, considering value spreads greater than 9% could be informative in the context of other countries, some of which appear to have larger value spreads (Fama and French, 1998) and if one is interested in the potential value financial intermediaries in the U.S. are currently generating — the value spread would likely be larger absent financial intermediaries’ seeking (and thereby partially eliminating) alphas.

When interpreting the magnitudes of our *gain* estimates, 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 addition, the estimates do not directly speak to the fair compensation that the financial sector should have received historically. Instead, it evaluates how large compensation could be if the financial sector eliminated the estimated informational inefficiencies in the future.³¹

C. Efficiency Gains and Tobin’s q

In this subsection we assess whether the gains from moving to undistorted investment policies vary across firms with different Tobin’s q . The solid line in Figure 4 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 the largest gains can be achieved for high Tobin’s q firms (“growth firms”).

This asymmetric result is 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 less than one, it still does not want to invest, nor can it disinvest, once again leaving the real investment behavior undistorted. Instead, consider a growth firm, which by definition has a Tobin's q larger than one. Such a firm's investment rates are positive and sensitive to (mis)valuations. As a consequence, alphas can cause substantial investment distortions for this type of firm.

In light of these results, the findings by Bai et al. (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.³²

[Place Figure 4 about here]

While the solid line in Figure 4 reflects a typical firm's 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 how the economy's aggregate value *gain* is distributed across firms with different Tobin's q (the orange squares). This distribution shows that most of the gain in the economy can be generated by adjusting the investment policies of firms with a Tobin's q between 1 and 2. The reason for this is that two conditions have to be met to generate large efficiency gains. First, conditional on a particular Tobin's q the efficiency gain (i.e., the solid line in Figure 4) needs to be substantial.

Second, a significant amount of market capitalization needs to be concentrated at that level of Tobin's q .

D. 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 0.11 with an R^2 value of 0.078.³³ These results indicate that our model features a relatively weak investment- q relationship, broadly in line with results established in the empirical literature.³⁴ 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, and are thus responding to disinformation in market prices.

E. Investment-Alpha Relationship

The cross-sectional relation between Book-to-Market decile portfolios and value-weighted alphas is targeted in the estimation of our model. As can be seen from the lower panel of Figure 1, the model fits this relation 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 5 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.

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 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 (see Internet Appendix A for details). In Section VII, we further discuss supplementary empirical evidence in the existing literature indicating that prices do affect firms' real investment decisions. In addition, while we measured alphas relative to the CAPM, investment alphas survive even when considering other often used benchmark asset pricing models,³⁵ 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 et al., 2016).

[Place Figure 5 about here]

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 necessarily 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 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 VII.D below).

VI. Efficient Prices and Model Misspecification

Researchers may have differing beliefs on which asset pricing model provides the correct efficient benchmark, and thus, whether empirical alphas indeed measure informational inefficiencies. Thus far, we have interpreted our results under the premise that alphas estimated conditional on an asset pricing model do measure informational inefficiencies. Yet our methodology can also be used under the view that market prices are simply always informationally efficient. Under this alternative, alphas are by construction the result of model misspecification, in particular the omission of priced risk factors in the econometrician's model. An economist may then wonder whether these omitted factors are of first-order relevance for real economic activity. For the finance literature this perspective is informative, as it allows ranking asset pricing model failures by their repercussions for a model's real implications. As shown below, the answers to these questions are intimately linked to our results so far — the same types of alphas that are important for real distortions under the premise that alphas represent mispricing also indicate material model misspecification problems under the premise that market prices are efficient.

In the following, we detail how our methodology can be used under the premise that market prices are informationally efficient and the econometrician's model is misspecified.³⁶ Suppose S^* denotes a vector containing all aggregate variables that are potentially of relevance for agents' marginal utility. Further, let $m^*(S^*)$ denote the true SDF agents in the economy use, but which is unknown to the econometrician, causing the model misspecification problem.

The vector S^* can possibly be very large and might contain objects such as the whole

cross-sectional distribution of various firm characteristics. In contrast, the econometrician considers a parsimonious candidate model for the SDF denoted by $m(S)$, where S is the vector of variables that the econometrician allows to affect the model SDF. An example consistent with the presented results would be that the econometrician uses the CAPM, such that the only variable relevant for risk premia is the market portfolio, but in fact asset prices also depend on comovement with other factors, which are contained only in the larger vector S^* .

To focus on the problem of omitted risk factors, suppose the econometrician efficiently uses available information to forecast expected cash flows, just like market participants do, and also employs accurate risk free rates r_f . An Arrow-Debreu claim paying one dollar in state (s_τ, S_τ^*) has the following prevailing efficient market price:

$$V_t = \mathbb{E}_t \left[\frac{m^*(S_\tau^*)}{m^*(S_t^*)} \mathbb{1}(s_\tau, S_\tau^*) \right]. \quad (30)$$

Let rp_t and rp_t^* denote the expected excess returns on this Arrow-Debreu claim under the econometrician's model and under the true model, respectively, that is:

$$rp_t dt = - \mathbb{E}_t \left[\frac{dm(S_t)}{m(S_t)} \frac{dV_t}{V_t} \right], \quad (31)$$

$$rp_t^* dt = - \mathbb{E}_t \left[\frac{dm^*(S_t^*)}{m^*(S_t^*)} \frac{dV_t}{V_t} \right]. \quad (32)$$

As is standard in the finance literature, the econometrician can now empirically measure alphas as the deviations between the average excess returns of traded assets and her model's predictions:

$$\alpha_t = rp_t^* - rp_t. \quad (33)$$

Whereas the true SDF is not known to the econometrician, these alphas and the stochastic process they follow are amenable to estimation. A researcher can then analyze how firm

investment is affected in her model economy if prices are instead determined based on the amended SDF

$$\frac{m(S_\tau)}{m(S_t)} e^{-\int_t^\tau \alpha_u du}, \quad (34)$$

which mirrors the discount factor specified in our baseline analysis in Section III.C. This amended SDF converts an alpha *process*, which characterizes misspecification errors affecting *expected returns*, to corrections in *price levels*, without requiring knowledge of the true SDF m^* . Imposing the discount factor (34) implies that prices in the model economy replicate observed market prices, that is:

$$\mathbb{E}_t \left[\frac{m(S_\tau)}{m(S_t)} e^{-\int_t^\tau \alpha_u du} \mathbb{1}(S_\tau, S_\tau^*) \right] = V_t. \quad (35)$$

A researcher can now verify whether the pricing errors under the candidate SDF m are a first-order concern for quantitative analyses of firms' real investment decisions. For instance, one may use an analysis as shown in Figure 2 to evaluate the impact on firms' investment rates.

Clearly, under the premise that market prices are always informationally efficient, the model economy where alphas are set to zero is not a desirable counterfactual; it is simply a misspecified economy. Thus, this premise does not lend itself to determining a *gain* measure as considered previously. Yet, the unifying insight of our analysis — independent of one's view regarding market efficiency — is that not all alphas are created equal, and our methodology helps identify the types of alphas that are quantitatively relevant.

This discussion also highlights another relevant insight for the misallocation literature following Hsieh and Klenow (2009). As cross-sectional dispersion in risk factor exposures creates dispersion in investment, it also implies dispersion in firms' marginal revenue products of capital. This channel thus provides another reason why interpreting dispersion in marginal revenue products as evidence of misallocations can be problematic.

VII. Robustness Considerations

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

A. *Efficiency Gains in General Equilibrium*

Due to data limitations we have thus far analyzed only publicly traded firms, thus missing a significant part of output that affects aggregate consumption. As a result, we chose a partial equilibrium approach when quantifying gains from improving informational efficiency. In particular, when calculating our *gain* measure introduced in Section V.B, we assumed that the process for agents' marginal utility of consumption m stays unaltered in the counterfactual economy. This partial equilibrium approach is naturally a good approximation if the counterfactual economy is associated with "small" adjustments in consumption dynamics. Yet, since public firms' net payout accounts for a non-negligible fraction of aggregate consumption, one might wonder to which extent this partial equilibrium approach biases our efficiency gain evaluations. In this section, we address this question quantitatively and find that the effects are of second-order importance.

In particular, our analysis builds on a classic Lucas (1987) type welfare calculation, which expresses agents' attitudes towards alternative consumption processes in terms of lifetime consumption equivalent gains. Naturally, such a calculation requires taking a stance on agents' preferences and on the process of net payout generated by all other parts of the economy — in general equilibrium, consumption equals the sum of net payout from public firms and from these other parts of the economy. We consider standard Epstein Zin preferences, and, to ensure feasibility, suppose that the other parts of the economy do not yield different net payout under the counterfactual (e.g., private firms obtain the same investment and generate the same output as in the initial economy). As a result, all differences in net payout by public firms between the initial and the counterfactual economy

directly imply changes in households' consumption — if public firms in aggregate require more resources/investment under the counterfactual, consumption has to drop; if public firms generate more output under the counterfactual, consumption increases accordingly.

Contrary to our previous partial equilibrium approach, this analysis does not assume that agents evaluate changes in state-contingent payoffs at *fixed* state-contingent marginal utilities. Instead, due to utility function curvature, there are *decreasing* utility gains from increases in consumption in a given aggregate state. To ensure comparability, we express the resulting Lucas (1987) gain measure as a fraction of public firm net payout, and denote it as $gain_{GE}$. This measure converges to our previous partial equilibrium *gain* measure (equation (29)) when public firms' contribution to consumption becomes small (see Internet Appendix B for details). Our original *gain* measure and the Lucas (1987) type calculation are thus closely related.

To economize on space, we relegate additional details on the modeling and parameter choices to Internet Appendix B, and now discuss the key results of this analysis. We find that the $gain_{GE}$ measure is only slightly smaller than our original *gain* estimate: for our baseline analysis we find that $\frac{gain_{GE}}{gain} = 0.95$, that is, 95% of our original estimate survives. Accounting for the general equilibrium implications of changes in state-contingent marginal utility thus has limited (indeed second-order) effects on our gain estimates. Moreover, we explain in Internet Appendix B why this estimate is even conservative given the specific approach we took to calculate it. We note however, that general equilibrium effects would likely become more relevant in analyses that also consider the elimination of informational inefficiencies in non-publicly traded sectors. More work in this direction would be very valuable and is left for future research (see also Section IX).

B. Existing Evidence Supporting Identifying Assumptions

Arguably, two of the most important identifying assumptions for our quantitative analysis of efficiency gains are that (1) CAPM alphas indeed measure 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. Note that if our methodology is used under the premise that market prices are efficient (as discussed in Section VI) these identifying assumptions do not apply.

First, while we have used the CAPM — the standard benchmark model in the literature — to identify alpha wedges, other asset pricing models can be accommodated by researchers using our methodology. In addition, the *gain* estimates we presented as a function of the value spread in Figure 3 (see also Figures 6a and 6b 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.³⁷ The results from this literature support the view that (dis)information encoded in market prices affects 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 for example be accommodated by multiplying α in equation (26) 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 3 provides a preliminary assessment of the effects of such shading.

C. (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 would happen if such a mechanical relation was introduced and firms still maximized their going market values. For example, one could consider 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 assumption 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 upwards, where firms face negative alphas, and are thus overvalued. Given the magnitude of the value spread, the model then likely has additional difficulty matching the empirical cross-sectional distribution of Tobin's q in the data. In addition to this type of alpha specification, there are many alternative ways to specify alpha processes as being correlated with various technology parameters and firm characteristics, which is an interesting avenue for future research.

D. 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 6a plots the aggregate efficiency *gain* as computed in equation (29) for different levels of persistence. To vary persistence we multiply the transition rates $(h_{\alpha}^{+}, h_{\alpha}^{-})$ of the baseline parameterization by factors ranging between 0.5 and 2. Afterwards, the values of all alpha states are shifted by an equal amount to ensure that the value-weighted average alpha is still zero in each economy. These adjustments do not change the unconditional probabilities of the alpha states. What they do change are the magnitudes of the mispricing wedge (see equation

17), the persistence of deviations from informationally efficient price levels, and 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. Figure 6b shows this result by plotting the corresponding value spread in each of these economies. The results in Figure 6a indicate that the efficiency 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 impact on real efficiency. On the other hand, if anomalies are more persistent, they can create material real inefficiencies.

[Place Figures 6a and 6b about here]

E. The Role of Adjustment Frictions

As argued at the end of Section III.E.1, the real distortions we study 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 our *gain* measure. If there are no adjustment frictions, a firm effectively maximizes current-period net payout date-by-date, rendering distorted expectations about a firm's future productivity irrelevant for its 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 substantial 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 and differing capital installation and deinstallation efficiencies, 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 analyses with respect to our two search cost parameters ($\tilde{\theta}^+$ and $\tilde{\theta}^-$) indicate in fact that lowering adjustment frictions relative to our benchmark estimation increases the magnitude of real distortions. For example, reducing the search cost parameter $\tilde{\theta}^+$ in our baseline parameterization by 10% increases the aggregate *gain* estimate by 8.5%, while reducing the model's fit of the targeted moments only slightly.

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 ones featured in our model. Thus, these estimates should be interpreted as a catch all — they would likely be smaller if other types of frictions were introduced as well.³⁸ 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, can also be consistent with models where managers are aware of the fact that they face irrational markets, as considered in Stein (1996). Consistent with our predictions and those of Stein (1996), Baker et al. (2003) find that equity dependent firms — proxied by characteristics such as a high Tobin's q — are more responsive to nonfundamental movements in stock prices.

VIII. 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 by 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 how well a particular portfolio sort identifies the true underlying alpha process. 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 (1991) fails to capture the real economic value that active managers provide.

A. 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.

B. 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 V.D) and different degrees of persistence. The Markov matrix of an anomaly's sorting variable (see Table I) 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.

C. 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, 1991). 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.³⁹ 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, they thus can help increase the true value of the 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 (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.

IX. 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). Thus, our objective is markedly different from the literature that has exclusively focused on the financial market aspects of such anomalies. As a methodological contribution, we introduce a novel lumpy investment model that can incorporate the dynamic nature of alphas. The advantage of our framework is that it yields closed-form solutions for the distributions of firm dynamics, facilitating the estimation of aggregate distortions associated with cross-sectional mispricings. We find that the aggregate real distortions 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, it is unclear to what extent they are actually engaged in that activity, raising the question how large cross-sectional anomalies would be absent financial intermediaries. Moreover, 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. On the other hand, reducing anomalies with (potentially) even statistically insignificant alphas that are very persistent may be very valuable for society. One potential example of such an anomaly is the size effect. That said,

time series changes and secular trends suggest that financial intermediation may have had a material impact historically (see McLean and Pontiff, 2016; Bai et al., 2016).

While our paper has provided a first analysis of the aggregate real implications of cross-sectional financial market mispricings, it is by no means intended to provide the final answer. Rather, it is intended to start a discussion on which financial market anomalies matter for real (mis)allocations. Going forward, it would be valuable to extend our framework and to use it to quantify the gains from eliminating other types of mispricings. Examples of possible extensions include a full-fledged general equilibrium analysis that also features private firms, labor, and the government sector. Examples of other types of mispricings include aggregate stock market misvaluations, 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 respect to 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 I. 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_{it}^A denote the return on assets of decile portfolio i at time t . We compute this return as follows:

$$R_{it}^A = we_{it-1}R_{it} + (1 - we_{it-1})R_t^D, \quad (\text{A1})$$

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_t^D 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 Figure 3 and in Figures 6a and 6b.

Appendix B. Markov Matrices

[Place Table III about here]

Appendix C. Beliefs Supporting Alpha Processes

The purpose of this Appendix is to show why any α -process that implies that the value function associated with equation (26) is strictly monotone in a can be micro-founded by subjective beliefs, which are characterized by the matrices Λ_a^* (recall that these matrices are allowed to depend on various state variables). Suppose that there indeed exist generator matrices Λ_a^* that imply a given α -process that follows a continuous-time Markov chain, as introduced in Section III.B. Given such distorted generator matrices, the value functions \tilde{V} and the controls $i = (i^+, i^-)$ solve both the original HJB equation (22) and the HJB equation under the objective measure (26), the latter of which substitutes in the definition of α in (25). The HJB equation (26) yields \tilde{V} for all states, except for those where a reaches the boundaries $\min\{\Omega_a\}$ and $\max\{\Omega_a\}$, which are zero probability events under the objective measure. For simplicity, suppose that the boundaries $\min\{\Omega_a\}$ and $\max\{\Omega_a\}$ are absorbing states under the subjective measure, in order to solve for the valuations \tilde{V} in these boundary states (the assumption that the boundaries are absorbing states is not necessary, but our objective here is merely to establish existence). Now one can use the valuations \tilde{V} in all states to verify that the value function is strictly increasing in log-productivity a , that is,

whether for all states (s, Z) the following conditions hold:

$$r_a^+(s, Z) = \frac{\tilde{V}(s^{a+}, Z)}{\tilde{V}(s, Z)} - 1 > 0, \text{ for } a < \max\{\Omega_a\} \quad (\text{C1})$$

$$r_a^-(s, Z) = \frac{\tilde{V}(s^{a-}, Z)}{\tilde{V}(s, Z)} - 1 < 0, \text{ for } a > \min\{\Omega_a\}, \quad (\text{C2})$$

where s^{a+} and s^{a-} denote vectors that are identical to the state vector s , except that the element corresponding to the state a is increased and decreased by one increment, respectively. Then there always exist subjective transition rates $h_a^{+*}(s, Z) > 0$ and $h_a^{-*}(s, Z) > 0$ such that in all states (s, Z) that are reached with strictly positive probability under the objective measure, a given $\alpha \in (-\infty, +\infty)$ is indeed supported, that is, the following equation holds:

$$\alpha = (h_a^+(a) - h_a^{+*}(s, Z)) \cdot r_a^+(s, Z) + ((h_a^-(a) - h_a^{-*}(s, Z)) \cdot r_a^-(s, Z)). \quad (\text{C3})$$

Finally, for the boundary states of a , which are not reached under the objective probability measure, we imposed the technical condition that $\alpha = 0$.

Appendix D. Hamilton-Jacobi-Bellman Equation

The Hamilton-Jacobi-Bellman equation associated with the maximization problem in (20) is given by:

$$\begin{aligned} 0 = \max_{(i^+, i^-) \geq 0} & \left\{ \pi(i, k, a, Y) - r_f(Z)V(s, Z, Y) + \Lambda_k^i(k)\mathbf{V}_k(s, Z, Y) \right. \\ & + V_Y(s, Z, Y)Y\mu(Z) + \frac{1}{2}V_{YY}(s, Z, Y)Y^2\sigma(Z)^2 - V_Y(s, Z, Y)Y\sigma(Z)\nu(Z) \\ & \left. + \Lambda_a^*(a)\mathbf{V}_a(s, Z, Y) + \bar{\Lambda}_Z(Z)\mathbf{V}_Z(s, Z, Y) + \Lambda_\alpha(\alpha)\mathbf{V}_\alpha(s, Z, Y) \right\}, \quad (\text{D1}) \end{aligned}$$

where the notation $\mathbf{V}_x(s, Z, Y)$ indicates a vector that collects the values of the function $V(s, Z, Y)$ evaluated at all $x \in \Omega_x$ while keeping the other arguments fixed. Given the

conjecture that $V(s, Z, Y) = Y \cdot \tilde{V}(s, Z)$, it can be verified that the HJB equation scales with Y . Dividing by Y , rearranging and using the definition for rp in equation (23) yields equation (22).

Appendix E. Distributions

Let \mathbf{n}_t denote the vector of length $n = n_k \cdot n_a \cdot n_\alpha \cdot n_Z$ containing the mass of firms in each Markov state at time t , and let $\mathbf{pr}_t = \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}$ denote the corresponding vector of probabilities of the distribution of firms across these n states. The law of motion for \mathbf{pr}_t is given by:

$$d\mathbf{pr}_t = \frac{d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} - \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} \frac{\mathbf{1}'d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t}. \quad (\text{E1})$$

The expected change in this distribution is thus given by:

$$\mathbb{E}_t [d\mathbf{pr}_t] = \mathbb{E}_t \left[\frac{d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} - \frac{\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} \frac{\mathbf{1}'d\mathbf{n}_t}{\mathbf{1}'\mathbf{n}_t} \right] = (\mathbf{I}_n - \mathbf{pr}_t \mathbf{1}') \Lambda' \mathbf{pr}_t dt, \quad (\text{E2})$$

where Λ denotes an $n \times n$ generator matrix for all n states (see more details below), and 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 distribution is defined as the vector of probabilities $\widehat{\mathbf{pr}}$ starting from which there is no expected change in the distribution, that is,

$$\mathbb{E}_t [d\mathbf{pr}_t | \mathbf{pr}_t = \widehat{\mathbf{pr}}] = \mathbf{0}. \quad (\text{E3})$$

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

$$(\Lambda' - \mathbf{I}_n \mathbf{1}' \Lambda' \widehat{\mathbf{pr}}) \widehat{\mathbf{pr}} = \mathbf{0}. \quad (\text{E4})$$

The off-diagonal elements of the matrix Λ contain the (endogenous) rates with which firms transition between the n states. The diagonal elements of the matrix Λ contain the

sum of all flow rates of leaving a given state and net-growth associated with firms entering and exiting the system (see footnote 26), that is:

$$\Lambda(j, j) = - \sum_{j' \neq j} \Lambda(j, j') + h_{entry}(j) - h_{exit}(j) \quad \text{for } j = 1, \dots, n. \quad (\text{E5})$$

The condition

$$\frac{\mathbf{1}' \mathbb{E}[d\mathbf{n}_t]}{dt} = \mathbf{1}' \Lambda' \mathbf{n}_t = 0 \quad (\text{E6})$$

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 26, we assume that entry and exit rates are identical in all states (that is, $h_{entry}(j) = h_{exit}(j)$), such that this condition is satisfied. Equation (E4) then simplifies and $\widehat{\mathbf{pr}}$ is the solution to the linear system:

$$\begin{pmatrix} \Lambda' \\ \mathbf{1}' \end{pmatrix} \widehat{\mathbf{pr}} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}. \quad (\text{E7})$$

That is, the vector of stationary probabilities, $\widehat{\mathbf{pr}}$, is the left normalized eigenvector of Λ associated with the eigenvalue 0. Moreover, if the current cross-sectional distribution of firms across states is given by \mathbf{pr}_0 then the conditional distribution of firms across all states after a time period of length τ is obtained by computing the matrix exponential:

$$\mathbf{pr}_\tau = \mathbf{pr}_0 e^{\Lambda \cdot \tau}. \quad (\text{E8})$$

REFERENCES

- Abel, Andrew B., 1989, Asset Prices under Heterogenous Beliefs: Implications for the Equity Premium, Working paper, The Wharton School.
- Arrow, Kenneth 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).
- Bai, Jennie, Thomas Philippon, and Alexi Savov, 2016, Have Financial Markets Become More Informative?, *Journal of Financial Economics* .
- Baker, Malcolm, Jeremy Stein, and Jeff Wurgler, 2003, When does the market matter? stock prices and the investment of equity-dependent firms, *Quarterly Journal of Economics* 118, 969–1005.
- Ball, Ray, and Philip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research* 6, 159–178.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Baron, Matthew, and Wei Xiong, 2016, Credit expansion and neglected crash risk, *Quarterly Journal of Economics* forthcoming.
- Barro, Robert, 1990, The stock market and investment, *Review of Financial Studies* 3, 115–31.
- Berk, Jonathan, and Jules van Binsbergen, 2014, Assessing asset pricing models using revealed preference, Working Paper, The Wharton School.
- Berk, Jonathan B, 1995, A Critique of Size-Related Anomalies, *Review of Financial Studies* 8, 275–286.

- Bernard, Victor L., and Jacob K. Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305 – 340.
- Carhart, Mark M, 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57–82.
- Chen, Hui, 2010, Macroeconomic conditions and the puzzles of credit spreads and capital structure, *Journal of Finance* 65, 2171–2212.
- Chen, Qi, Itay Goldstein, and Wei Jiang, 2007, Price informativeness and investment sensitivity to stock price, *Review of Financial Studies* 20, 619–650.
- Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2009, The Price Is (Almost) Right, *Journal of Finance* 64, 2739–2782.
- Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *The Journal of Finance* 63, 1609–1651.
- David, Alexander, 2008, Heterogeneous Beliefs, Speculation, and the Equity Premium, *Journal of Finance* 63, 41–83.
- David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran, 2016, Information, misallocation and aggregate productivity*, *The Quarterly Journal of Economics* .
- Dessaint, Olivier, Thierry Foucault, Laurent Fresard, and Adrien Matray, 2016, Ripple Effects of Noise on Corporate Investment, CEPR Discussion Papers 11081, C.E.P.R. Discussion Papers.
- Edmans, Alex, Itay Goldstein, and Wei Jiang, 2012, The real effects of financial markets: The impact of prices on takeovers, *The Journal of Finance* 67, 933–971.
- Eisfeldt, Andrea L., and Adriano A. Rampini, 2006, Capital reallocation and liquidity, *Journal of Monetary Economics* 53, 369–399.

- Fahlenbrach, Rdiger, Robert Prilmeier, and Ren M. Stulz, 2016, Why does fast loan growth predict poor performance for banks?, Working Paper 22089, National Bureau of Economic Research.
- Fairfield, Patricia M., J. Scott Whisenant, and Teri Lombardi Yohn, 2003, Accrued earnings and growth: Implications for future profitability and market mispricing, *The Accounting Review* 78, 353–371.
- Fama, Eugene F, 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25, 383–417.
- Fama, Eugene F, 1991, Efficient Capital Markets: II, *Journal of Finance* 46, 1575–1617.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 1998, Value versus Growth: The International Evidence, *Journal of Finance* 53, 1975–1999.
- Foucault, Thierry, and Laurent Fresard, 2014, Learning from peers' stock prices and corporate investment, *Journal of Financial Economics* 111, 554–577.
- Gilchrist, Simon, Charles Himmelberg, and Gur Huberman, 2005, Do stock price bubbles influence corporate investment?, *Journal of Monetary Economics* 52, 805–827.
- Gomes, Joao F., 2001, Financing Investment, *American Economic Review* 91, 1263–1285.
- Gomes, Joao F., Amir Yaron, and Lu Zhang, 2003, Asset Prices and Business Cycles with Costly External Finance, *Review of Economic Dynamics* 6, 767–788.
- Grossman, Sanford J., and Oliver D. Hart, 1980, Takeover Bids, the Free-Rider Problem, and the Theory of the Corporation, *Bell Journal of Economics* 11, 42–64.

- Hall, Robert E., 2003, Corporate earnings track the competitive benchmark, Working Paper 10150, National Bureau of Economic Research.
- Hansen, Lars Peter, and Thomas J. Sargent, 2001, Acknowledging Misspecification in Macroeconomic Theory, *Review of Economic Dynamics* 4, 519–535.
- Hansen, Lars Peter, and Thomas J. Sargent, 2008, *Robustness* (Princeton University Press).
- Harris, Milton, and Artur Raviv, 1993, Differences of Opinion Make a Horse Race, *Review of Financial Studies* 6, 473–506.
- Harvey, Campbell, Yan Liu, and Heqing Zhu, 2016, ... and the Cross-Section of Expected Returns, *Review of Financial Studies* 29, 5–68.
- Hayek, F.A., 1945, The use of knowledge in society, *American Economic Review* 35, 519–530.
- Hennessy, Christopher A., and Toni M. Whited, 2007, How Costly Is External Financing? Evidence from a Structural Estimation, *Journal of Finance* 62, 1705–1745.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2016, A Comparison of New Factor Models, Working paper.
- Hsieh, Chang-Tai, and Peter J. Klenow, 2009, Misallocation and manufacturing tfp in china and india, *The Quarterly Journal of Economics* 124, 1403–1448.
- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.
- Kiyotaki, Nobuhiro, and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105, 211–48.
- Lintner, John, 1969, The aggregation of investor's diverse judgments and preferences in purely competitive security markets, *Journal of Financial and Quantitative Analysis* 4, 347400.

- Lucas, Robert E., 1987, *Models of business cycles* (New York: Basil Blackwell).
- McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.
- Morck, Randall, Andrei Shleifer, and Robert Vishny, 1990, The stock market and investment: Is the market a sideshow?, *Brookings Papers on Economic Activity* 2.
- Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.
- Opp, Christian, 2017, Learning, optimal default, and the pricing of distress risk, Working paper, The Wharton School.
- Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642–685.
- Peters, Ryan H., and Lucian A. Taylor, 2016, Intangible capital and the investment-q relation, *Journal of Financial Economics* .
- Philippon, Thomas, 2010, Financiers versus engineers: Should the financial sector be taxed or subsidized?, *American Economic Journal: Macroeconomics* 2, 158–82.
- Philippon, Thomas, 2015, Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation, *American Economic Review* 105, 1408–38.
- Philippon, Thomas, and Ariell Reshef, 2012, Wages and Human Capital in the U.S. Finance Industry: 1909–2006, *The Quarterly Journal of Economics* 127, 1551–1609.
- Pindyck, Robert S., 1988, Irreversible investment, capacity choice, and the value of the firm, *The American Economic Review* 78, 969–985.

- Polk, Christopher, and Paola Sapienza, 2009, The stock market and corporate investment: A test of catering theory, *Review of Financial Studies* 22, 187–217.
- Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–16.
- Sharpe, William, 1991, The arithmetic of active management, *Financial Analysts Journal* 47, 7–9.
- Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.
- Sims, Christopher A., 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.
- Stein, Jeremy C, 1996, Rational Capital Budgeting in an Irrational World, *The Journal of Business* 69, 429–455.
- Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2004, Capital Investments and Stock Returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.
- Warusawitharana, Missaka, and Toni Whited, 2016, Equity market misvaluation, investment, and finance, *Review of Financial Studies* forthcoming.
- Whited, Toni M, 1992, Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data, *Journal of Finance* 47, 1425–60.
- Williams, Joseph T., 1977, Capital asset prices with heterogeneous beliefs, *Journal of Financial Economics* 5, 219–239.

Notes

¹Examples 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 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).

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

³As discussed below, investment alphas arise under a variety of commonly used empirical asset pricing models.

⁴In dynamic models, standard features of technology, such as adjustment costs, also generate dispersion in MRPs.

⁵In fact, cross-sectional variation in MRPs alone cannot be used to identify alpha wedges. Our dynamic lumpy investment model predicts that neither the marginal revenue products of capital nor marginal (or average) q are equalized across firms, even when there are no informational inefficiencies (i.e., when alphas are zero).

⁶That is, the joint null hypothesis that (1) the model employed indeed reflects the efficient benchmark, and (2) the actual economy is efficient.

⁷See, for instance, Whited (1992); Kiyotaki and Moore (1997); Gomes, Yaron, and Zhang (2003), and Hennessy and Whited (2007).

⁸Stein (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.

⁹See, e.g., Hou, Xue, and Zhang (2016) who show that factors motivated by the firm's first-order conditions help explain the cross-section of expected returns. The firm's first-order conditions also naturally hold in our framework.

¹⁰See also Baron and Xiong (2016) for a cross-country analysis suggesting a relation between aggregate credit expansion and bank overpricing.

¹¹One 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.

¹²See Section VIII.B for further discussion on this issue.

¹³The full Markov matrices are listed in Appendix B.

¹⁴Our setup can also easily accommodate grids of log-capital that are non-equidistant.

¹⁵Models with a continuous support are in any case approximated by a discrete support when solved numerically.

¹⁶Biases are for example common in the literature on heterogeneous beliefs. See, e.g., Lintner (1969); Williams (1977); Abel (1989); Harris and Raviv (1993), and David (2008).

¹⁷See, e.g., the literature on rational inattention following Sims (2003).

¹⁸Empirical evidence suggests that many anomalies that have been uncovered by the finance literature were indeed related to imperfect information processing. McLean and Pontiff (2016) show that asset pricing anomalies are attenuated substantially after they are documented in academic publications.

¹⁹If this condition was violated, the annualized expected abnormal return on a claim to $d\Pi_t$ at time t under the objective probability measure would be either plus infinity or minus infinity.

²⁰After conditioning on date t information, the actual net payout $d\Pi_t$ over the interval $[\tau, \tau + dt)$ is still stochastic because of the random nature of a firm's investment technology.

²¹See Appendix D for details.

²²As noted in model description (Section III.A), $i^+ = 0$ when $k = \max[\Omega_k]$ and $i^- = 0$ when $k = \min[\Omega_k]$.

²³The literature informing the calibrated parameters includes Bansal and Yaron (2004) and Chen (2010), and the references therein. As discussed in Internet Appendix B, our

main results are also robust to considering only Brownian aggregate shocks (i.e., eliminating Z -shocks). In this environment, a simple one-factor CAPM holds exactly, and the stationary distribution across Markov states is equal to the stationary cross-sectional distribution.

²⁴As stated in Section III.B, we impose the technical condition that the boundaries of the sets Ω_a^1 and Ω_a^2 are reached with probability zero under the objective measure. Thus, the sets Ω_a^1 and Ω_a^2 technically each contain two additional boundary states that are reached with probability zero under the objective measure.

²⁵More alpha states can be accommodated, but a three-state process turns out to be sufficient to capture the cross-sectional moments we target. Exploring more complex alpha processes (e.g., involving more states) and other anomalies is an important avenue for future research.

²⁶Since 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 III.D. 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.

²⁷Related to this effect, Berk (1995) argues that sorts on lagged market values mechanically capture either omitted risk factors or mispricings.

²⁸Given the estimated parameters, firms' market valuations in the model (measured em-

pirically 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. See Opp (2017) for a study of optimal firm default in the presence of debt in a similar model environment.

²⁹Pindyck (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.*”

³⁰As discussed in the literature review, related counterfactual analyses are employed in the existing misallocation literature. For example, in Hsieh and Klenow (2009), the efficient firm policies — which maximize firm value when firms face undistorted input prices in the efficient economy — also do *not* maximize firm value when firms face the distorted (after-tax) prices prevailing in the actual economy. As in our model, the culprit is that the effective prices firms face are distorted, not that firms’ behavior is distorted conditional on these prices.

³¹See, e.g., Philippon (2010), Philippon and Reshef (2012), and Philippon (2015) for papers analyzing financial sector compensation.

³²In this context it is interesting to note that our estimated model under-represents the negative alpha for the 1st decile (extreme growth firms) relative to the data. In that sense, the magnitude of our reported gain measure is likely conservative.

³³These 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 0.00056).

³⁴See, e.g., Peters and Taylor (2016).

³⁵Hou et al. (2016) for example find that the investment factor is not subsumed by a Fama-French (1993) three-factor model, the Carhart (1997) four-factor model, and the Pastor-Stambaugh (2003) model, which adds a liquidity factor to the three-factor model.

³⁶This perspective differs from analyses where decision makers within the economy recognize that their models can be misspecified, as for example, in Hansen and Sargent (2001, 2008).

³⁷See, e.g., Edmans et al. (2012) and the references therein. See also the recent papers by Foucault and Fresard (2014) and Dessaint, Foucault, Fresard, and Matray (2016).

³⁸Relatedly, Gomes (2001) shows that Tobin's q should capture most of investment dynamics even when there are credit constraints.

³⁹Berk and van Binsbergen (2014) provide other arguments for why Sharpe's arithmetic is flawed.

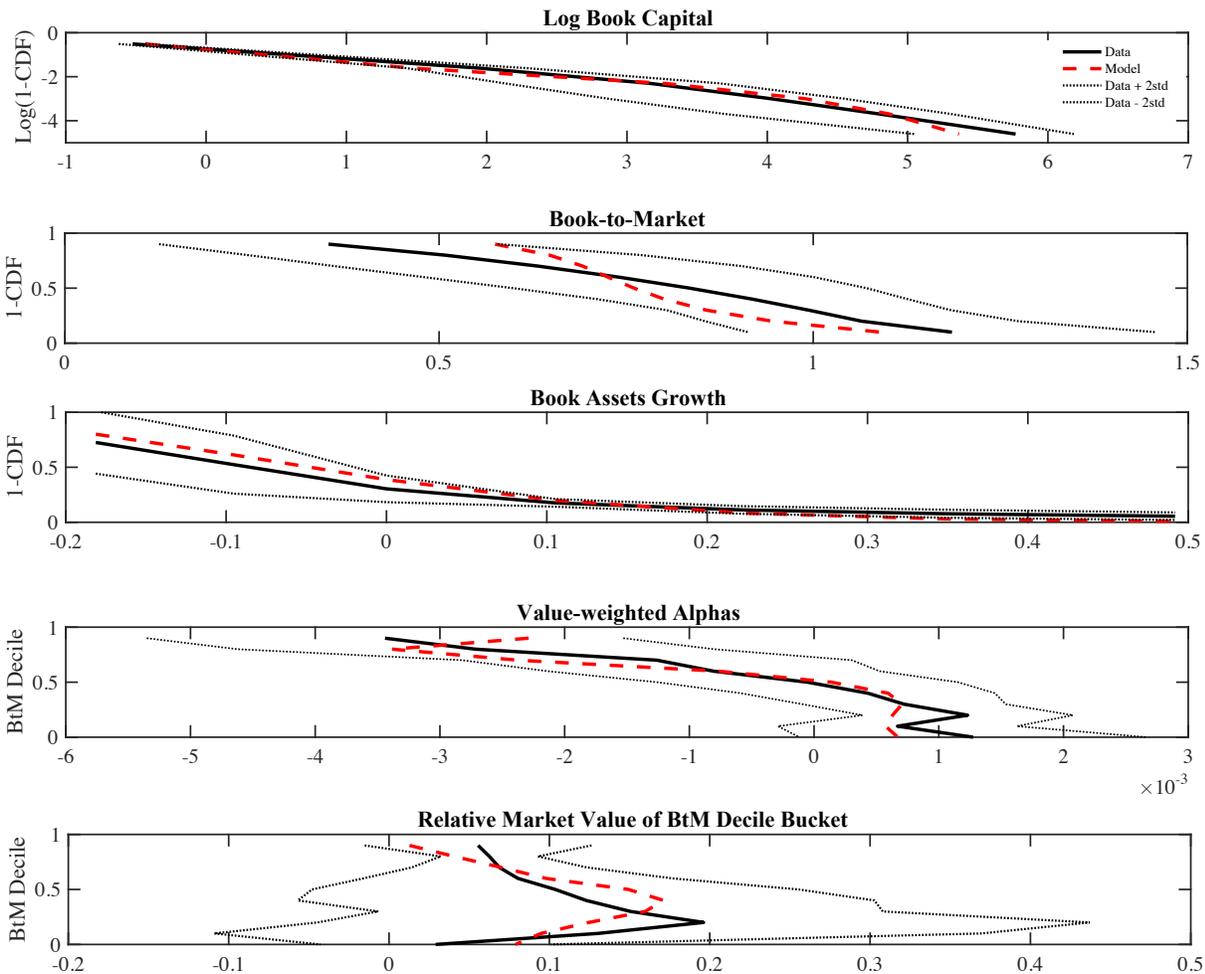


Figure 1. Model fit. The panels plot for each variable the model-implied values (red dashed line) and compare them with the data (black solid line), including 95% confidence bounds (black dotted lines). The confidence bounds are based on the standard deviations of the annual time series observations of percentile breakpoints. 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).

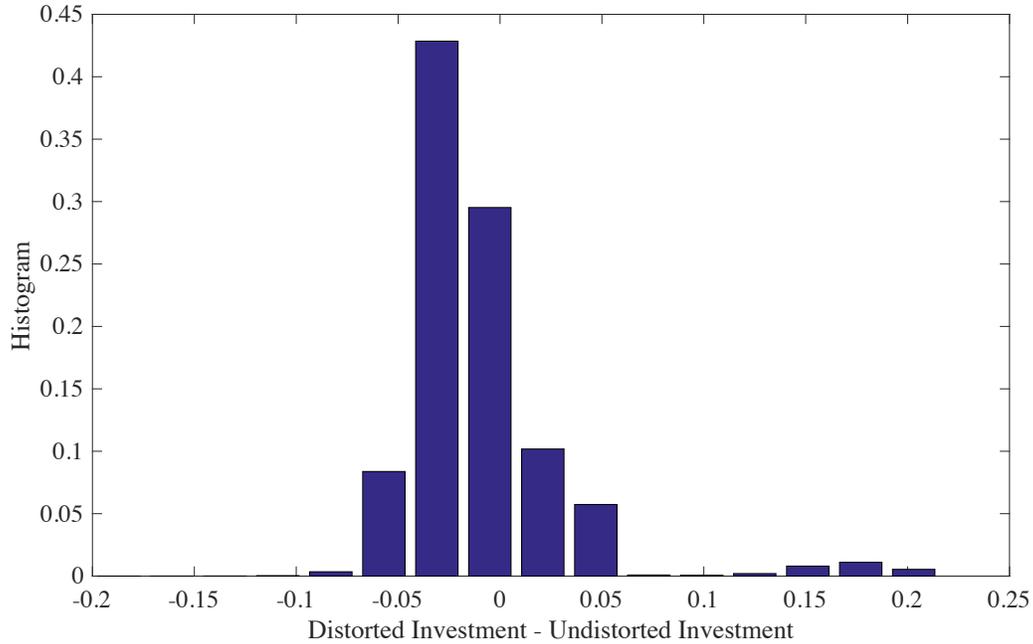


Figure 2. Investment distortions. The graph plots a histogram of the difference between the distorted and undistorted expected investment rate in each state, where the expected investment rate refers to $(i_+ - i_-)$ in our model. The histogram is computed based on the stationary distribution of the initial distorted economy, the parameters of which are detailed in Table II. The undistorted expected investment rates in each state are computed by solving the model with the same parameters, except that we set $\alpha_i = 0 \forall i$.

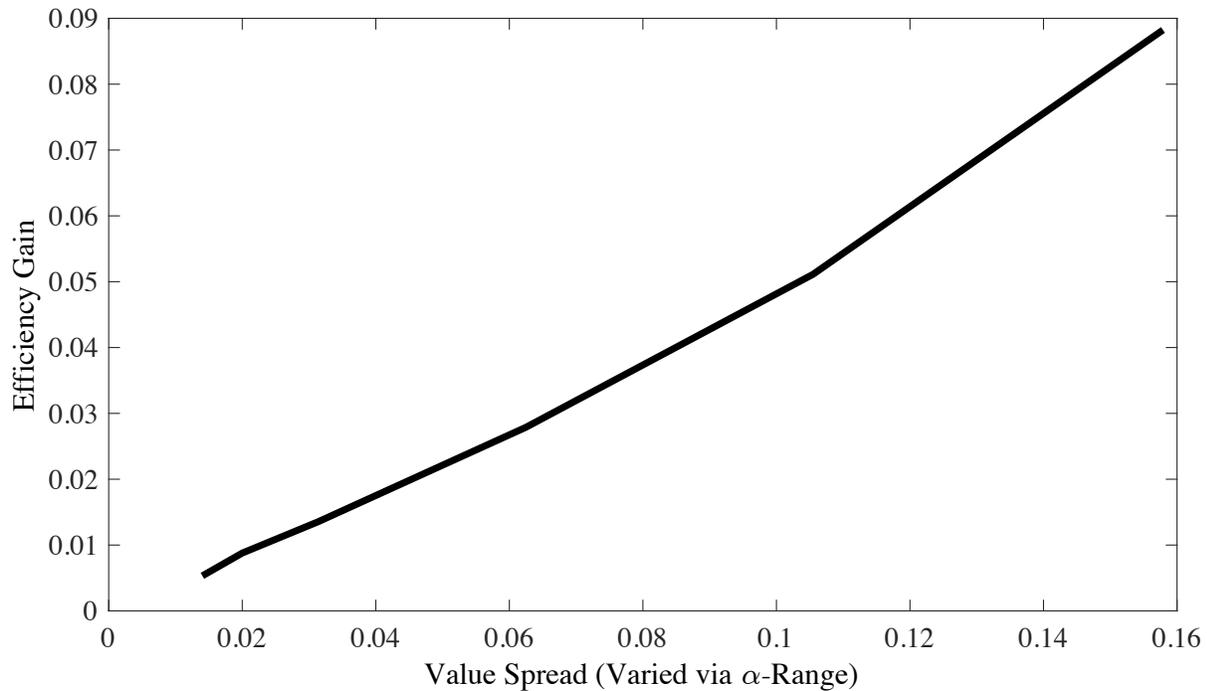


Figure 3. Aggregate efficiency gain as a function of the value spread. This graph plots the aggregate *gain* measure in equation (29) for different economies that are indexed by their value spread. The value spread is measured by the difference between the alpha of the top decile portfolio and the alpha of the bottom decile portfolio. We vary the value spread by increasing the range of the alpha process while maintaining a value-weighted alpha of zero in each economy. Starting with the alpha process parameters specified in Table II we increase or decrease the value for α_3 . Afterwards, the values of all alpha states are shifted by an equal amount to ensure that the value-weighted average alpha is still zero.

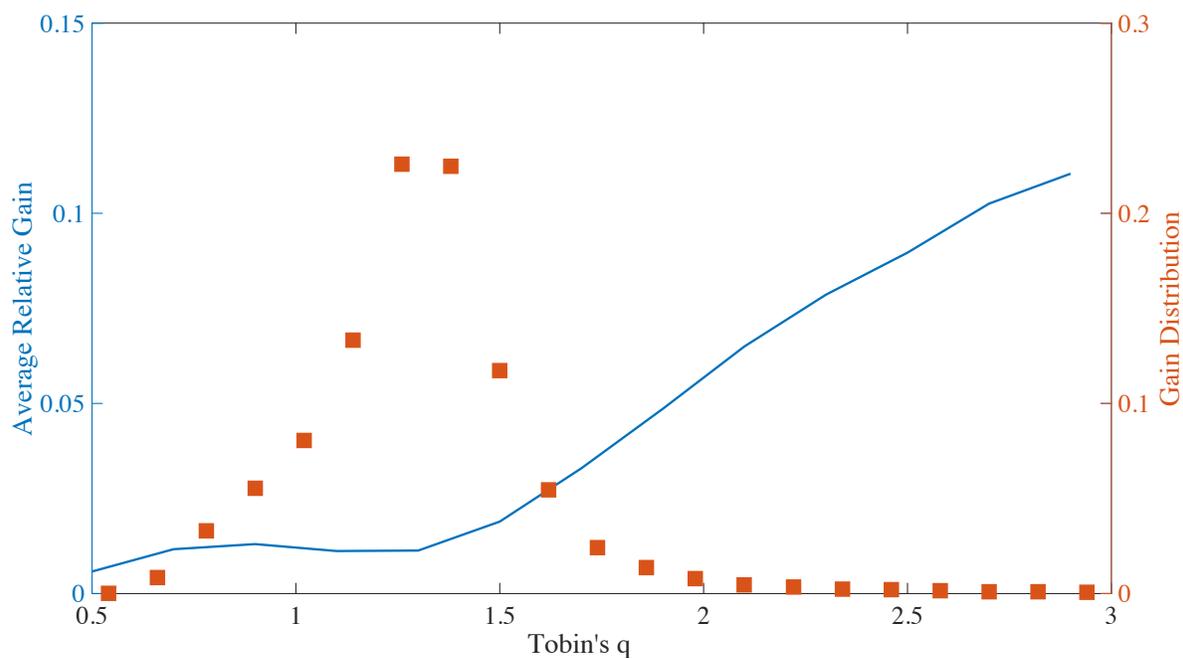


Figure 4. Gains and Tobin's q . The blue solid line plots the average relative gain (left vertical axis) for a typical firm with a given Tobin's q (horizontal axis) from moving to undistorted investment policies. The orange squares plot a histogram of how the aggregate value *gain* in (29) is distributed across firms (right vertical axis) with different Tobin's q . The parameters of the economy are detailed in Table II.

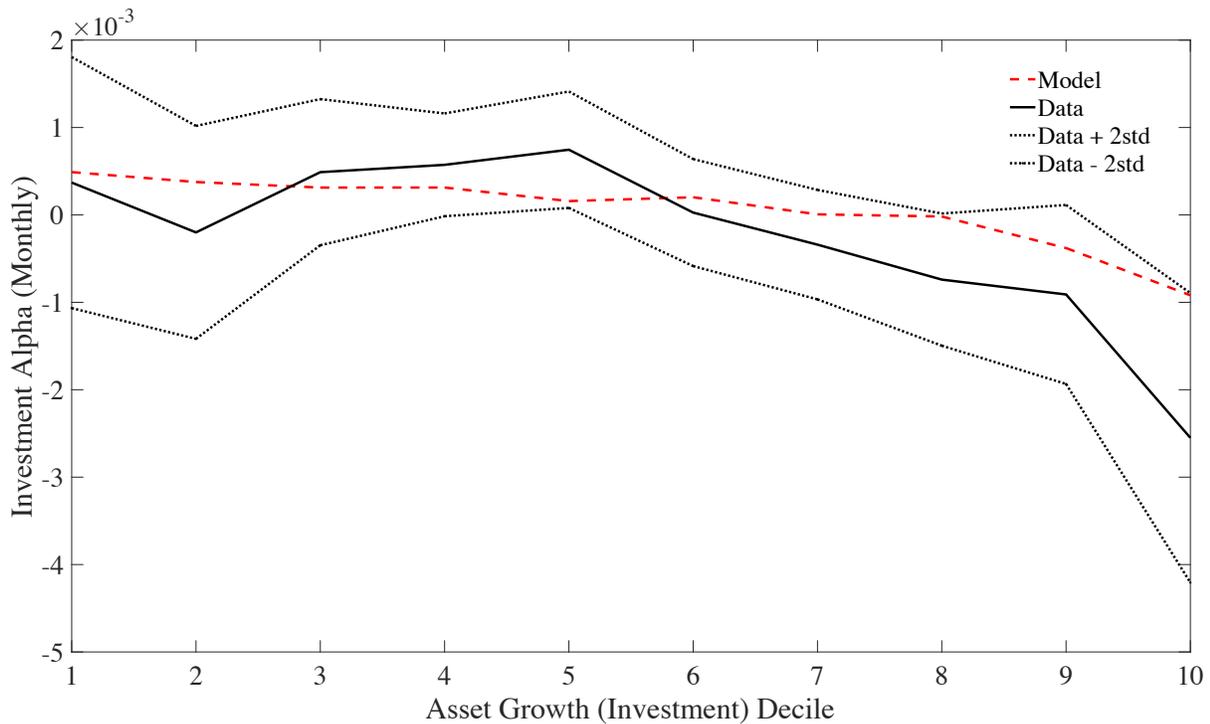


Figure 5. Investment-alpha relation. The graph plots the CAPM alphas of investment-sorted decile portfolios in the model (red dashed line) and compares them with the data (black solid line), including 95% confidence bounds (black dotted lines). The parameters of the economy are detailed in Table II.

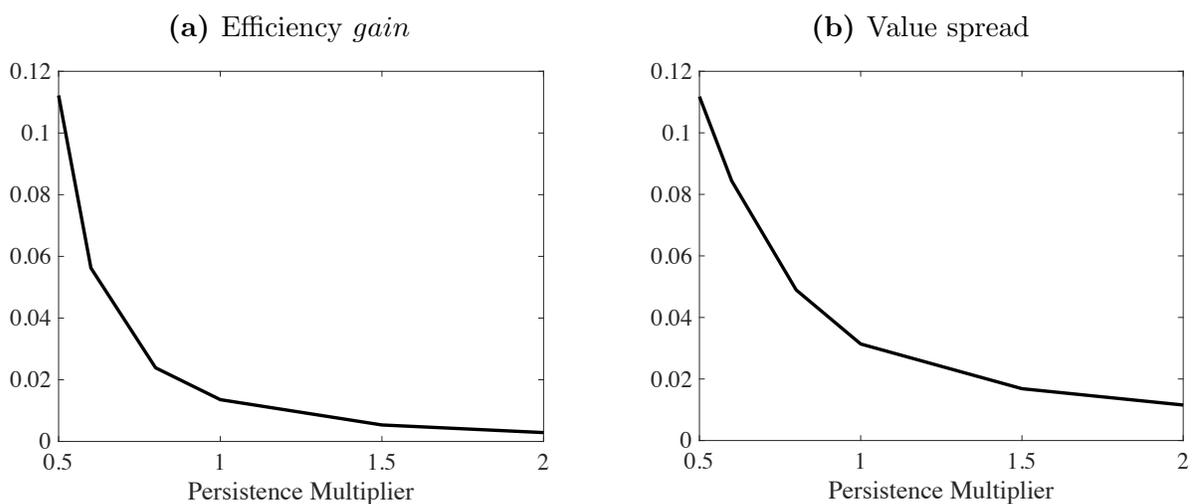


Figure 6. Alpha process persistence and aggregate gains. The graphs illustrate the effects of changing the persistence of the α -process on the aggregate efficiency *gain* and on the value spread. We multiply the transition rates ($h_{\alpha}^{+}, h_{\alpha}^{-}$) of the baseline parameterization by a factor $[0.5, 2]$. Afterwards, the values of all alpha states are shifted by an equal amount to ensure that the value-weighted average alpha is still zero in each economy. The other parameters of the economy are detailed in Table II. The value spread is measured by the difference between the alpha of the top decile portfolio and the alpha of the bottom decile portfolio.

Table I. 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.

Decile	1	2	3	4	5	6	7	8	9	10
Panel A: Raw Returns										
BtM	0.0100	0.0092	0.0103	0.0105	0.0111	0.0114	0.0124	0.0139	0.0138	0.0168
Invest	0.0132	0.0122	0.0122	0.0127	0.0118	0.0111	0.0106	0.0104	0.0113	0.0090
Panel B: CAPM Alphas										
BtM	-0.0014	-0.0014	0.0001	0.0001	0.0010	0.0011	0.0020	0.0034	0.0029	0.0060
Invest	0.0018	0.0015	0.0021	0.0029	0.0020	0.0012	0.0001	-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	0.1153	0.0993	0.0931	0.0776	0.0545	0.0204
Invest	0.0219	0.0505	0.0888	0.1207	0.1309	0.1434	0.1492	0.1221	0.1038	0.0686
Panel D: Time Series Average of Decile's Firm Value (Equity plus Debt) as Fraction of Total										
BtM	0.0556	0.0625	0.0687	0.0806	0.1037	0.1231	0.1509	0.1961	0.1298	0.0290
Invest	0.0252	0.0566	0.0961	0.1220	0.1302	0.1489	0.1511	0.1215	0.0914	0.0568
Panel E: Persistence as Measured by Diagonal Element of Decile in Markov Matrix										
BtM	0.5769	0.3522	0.2890	0.2579	0.2460	0.2439	0.2635	0.3023	0.3376	0.5274
Invest	0.2741	0.1789	0.1630	0.1515	0.1529	0.1498	0.1502	0.1605	0.1852	0.2273

Table II. Parameters. The three panels list parameters of the macroeconomy and firm-specific parameters. The parameters of the macroeconomy are calibrated to values commonly used in the existing literature. We normalize the scaled cost of installing capital (the conversion ratio) \tilde{p}^+ to one. We estimate firm-specific parameters via a GMM approach. The lower bound for the set of objectively possible log-productivity states is given by $\min\{a \in \Omega_a^1 : \Pr[a] > 0\} = -5.334$ for the first set of firms, where $\Pr[a]$ denotes the unconditional probability of the log-productivity value a under the objective probability measure. The constant shifter that differentiates the sets Ω_a^1 and Ω_a^2 is given by $\underline{a} = 0.099$. The grid increments Δ_a^- and Δ_a^+ for the sets Ω_a^1 and Ω_a^2 are given by $\{0.652, 0.597, 0.542, 0.452, 0.362, 0.550, 0.738, 0.534, 0.329\}$, where we determine every second value via interpolation and estimate the remaining ones. The fraction of firms that have the upward-shifted productivity process is estimated to be 16.95%. The log-capital grid is characterized by the lower bound $\min\{\Omega_k\} = -6.034$ (which can be interpreted in the context of a median value close to zero), the number of capital grid points $n_k = 160$, and the log-change in capital between grid points $\Delta_k = 0.1$.

Parameters of the Macroeconomy (Calibrated)			
Parameter	Variable	Boom	Recession
Transition rates for aggregate states	λ	0.1000	0.5000
Trend growth	μ	0.0245	-0.0145
Trend risk exposure	σ	0.0290	0.0290
Risk free rate	r_f	0.0200	0.0200
Local risk price	ν	0.7000	1.3000
Jump in m upon leaving state Z	$e^\phi - 1$	1.0000	-0.5000
Tax rate (personal + corporate)	τ	0.450	

Constant Firm-specific Parameters		
Parameter	Variable	Estimated Values
Rate of moving to next-higher a -state	h_a^+	2.828
Rate of moving to next-lower a -state	h_a^-	2.238
Search costs for capital acquisitions	$\tilde{\theta}^+$	0.833
Search costs for capital divestments	$\tilde{\theta}^-$	4.734
Unit yield from deinstalling capital	\tilde{p}^-	0.615
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_α^+	2.006	1.418	-
Rate of moving to next-lower α -state	h_α^-	-	0.354	0.415
Local alpha	α	-0.184	-0.034	0.023

Table III. Annual 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).

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	0.58	0.22	0.07	0.03	0.02	0.01	0.01	0.00	0.01	0.01
Dec 2	0.16	0.35	0.22	0.10	0.04	0.03	0.01	0.01	0.01	0.01
Dec 3	0.04	0.18	0.29	0.20	0.10	0.05	0.02	0.02	0.02	0.02
Dec 4	0.02	0.07	0.18	0.26	0.20	0.10	0.04	0.03	0.02	0.02
Dec 5	0.01	0.03	0.07	0.17	0.25	0.20	0.09	0.04	0.04	0.03
Dec 6	0.01	0.01	0.04	0.08	0.17	0.24	0.19	0.09	0.06	0.04
Dec 7	0.00	0.01	0.02	0.04	0.08	0.17	0.26	0.21	0.10	0.05
Dec 8	0.00	0.01	0.01	0.02	0.04	0.08	0.20	0.30	0.19	0.06
Dec 9	0.00	0.01	0.01	0.02	0.03	0.06	0.09	0.18	0.34	0.18
Dec 10	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.15	0.53

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	0.27	0.14	0.08	0.06	0.04	0.04	0.04	0.04	0.05	0.08
Dec 2	0.16	0.18	0.14	0.09	0.07	0.06	0.05	0.05	0.04	0.05
Dec 3	0.08	0.13	0.16	0.14	0.10	0.09	0.07	0.06	0.05	0.04
Dec 4	0.06	0.10	0.13	0.15	0.13	0.11	0.08	0.07	0.06	0.04
Dec 5	0.05	0.07	0.10	0.13	0.15	0.13	0.11	0.08	0.07	0.04
Dec 6	0.04	0.06	0.09	0.10	0.13	0.15	0.13	0.10	0.08	0.05
Dec 7	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.13	0.10	0.06
Dec 8	0.04	0.06	0.06	0.07	0.09	0.11	0.14	0.16	0.13	0.09
Dec 9	0.06	0.06	0.06	0.06	0.07	0.08	0.10	0.14	0.19	0.13
Dec 10	0.10	0.07	0.06	0.06	0.05	0.06	0.07	0.09	0.14	0.23