Real Anomalies: Are Financial Markets a Sideshow?

Jules H. van Binsbergen* Christian C. Opp†

June 15, 2016

Preliminary, comments welcome.

Abstract

We examine the importance of financial market imperfections such as asset pricing anomalies on the real economy. In order to assess distortions quantitatively, we estimate the joint dynamic distribution of firm characteristics that have been linked to financial imperfections and other firm variables, such as investment, capital, and value added. Based on a model that matches these joint dynamics we then evaluate the counterfactual dynamic distribution of these quantities absent financial market imperfections and find that they can cause large and persistent deviations. This implies that financial intermediaries that can reduce and/or eliminate such market imperfections can provide large value added to the economy. We show that both the persistence and the amount of mispriced capital are a major determinant of the real economic consequences. Our framework can be extended to address the welfare costs of incomplete risk sharing and other financial market imperfections.

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

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

1. Introduction

In the past few decades a vast literature has developed that attempts to document and explain the behavior of asset prices both in the cross section and the time series. The seminal paper on excess volatility (Shiller (1981)) has spurred a literature that attempts to explain why stock markets are so volatile and whether or not such volatility is excessive (irrational), relative to the existing models. Similarly, many different "anomalies" have been uncovered in the cross-section of asset prices, such as the value premium puzzle (firms with high Book-to-Market ratios earn on average higher returns than firms with low Book-to-Market ratios) and momentum (past winners earn higher returns than past losers).[1] One important question that naturally follows from these documented pricing effects is how important such effects are for the real economy. In this paper we quantitatively assess this question and show under what conditions these effects are important.

In the finance literature, mispricings are typically estimated based on realizations of so-called alphas, that is, deviations of average returns from a benchmark asset-pricing model. While realizations of alpha indicate that imperfections exist, they can be poor indicators of the economic importance of anomalies for at least three reasons. First, they only represent (partial) corrections of asset mispricings. Mispricing is thus an inherently dynamic phenomenon — as alphas are realized over time, firms are only temporarily affected by distortions in their cost of capital or biased expectations regarding their cash flows. That is, realized alphas are partial corrections of existing mispricings. As a consequence, we need to also consider the persistence of the mispricing. Second, as alphas are *return* measures, they do not give a accurate representation of the *value* of the mispricing. Just as the internal rate of return cannot be used to measure the value of an investment opportunity (it is the net present value that does), the alpha cannot be used to measure the economic importance of an anomaly.[2] Only by multiplying the alpha of a portfolio by the amount of capital in that portfolio can we get a sense of the value affected by the anomaly. Thirdly, and most importantly, it is not clear from studying alphas to what extent these distortions in the cost-of-capital translate into real investment and value added distortions. For example, one may wonder whether for the aggregate economy it is worse for firms to have a very short-lived alpha of 10% versus a

[1] See papers as early as Rosenberg, Reid, and Lanstein (1985) for value and Jegadeesh and Titman (1993) for momentum.

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

very persistent alpha of 1%. From an investors' point of view, the short-lived alpha may seem interesting, particularly if a fraction of firms gets resorted into this portfolio with large alpha, but for the firms investment decisions, it seems hard to imagine that such short-lived mispricings matter.

In this paper we thus deviate from a large fraction of the finance literature that is interested in how investors perform when chasing alphas, and instead focus on the economy-wide implications of such distortions. In order to assess the documented distortions quantitatively, we estimate the joint dynamic distribution of firm characteristics that have been linked to mispricing and other firm variables, such as investment, capital, and output. Based on a structural model that matches these joint dynamics we then evaluate the counterfactual dynamic distribution of capital, investment, and value added absent anomalies. Deviations between the counterfactual and the actual distribution allow us to assess the magnitude of real inefficiencies caused by asset-pricing anomalies. As such, while the existing literature evaluates asset pricing anomalies primarily based on the statistical significance of alphas, we aim to provide a framework to gauge economic significance, as measured by distorted investment and the present value of losses in value added. Economic significance will depend on the persistence in the underlying characteristic that is associated with alpha, the fraction, size and investment opportunities of firms most affected by an anomaly, and affected firms flexibility to adjust the scale of their operations inter-temporally.

We find that cross-sectional anomalies can have important effects on value added and investment. Even if we assume that the aggregate firm's cash flows are correctly priced by the market (the CAPM alpha of the index is by definition 0), cross-sectional anomalies have important effects. Due to distortions in the cost of capital for individual firms, some firms overinvest and others underinvest, both leading to suboptimal investment decisions and value destruction. That is, even though in our setting the cash flows of the aggregate firm are correctly valued by the market, the value added generated by these firms would be higher absent cross-sectional mispricings. We show that in our calibration, cross-sectional distortions can represent up to 1.3% in value added when just considering the value premium puzzle. As a consequence, a financial sector that is able to eliminate these mispricings, and extracts the rents from this process, can justifiable represent a fraction of 1.3% of the total market capitalization.

Our calculations shed light on another important debate in the literature on financial intermediation. One often heard critique of active mutual funds is Sharpe's arithmetic.

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

2. Related Literature

The influence of prices on the allocation efficiency of resources goes at least as far back as (Hayek (1945)). The allocational role of prices in secondary financial markets and their influence on real investment has been studied in papers by Leland (1992), Dow and Gorton (1997), Subrahmanyam and Titman (2001). Our paper also relates to Morck, Shleifer, and Vishny (1990) who argue there are three ways in which stock market valuations can influence real economic outcomes. First, managers could rely on the stock market as a source of information. Given how much uncertainty there is regarding the correct discount rate for projects, managers may want to use the stock market's valuation of the firm/sector to learn about the appropriate discount rate. Second, the stock market can directly affect investment through its influence on the cost of external financing. When firms issue stock to finance new investment, the market's valuation of the company is an important determinant of the cost of financing the new project. Third, outside investors can exert direct pressure on directors inside the firm. A low stock market valuation may increase the probability of a takeover or a forced removal of top management. In this paper, we are agnostic about the exact mechanism through

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

which the stock market's valuation affects investment. We assume that managers use the market's valuation in their investment decision in the sense that they adjust the discount rate (or equivalently the cash flow expectations) for projects to reflect the stock market's assessment of their value. This assumption is motivated by the recent empirical success of the so-called investment-based asset pricing literature, which argues that the first-order conditions of the firm (as opposed to the first-order condition of the investors), seems consistent with documented anomalies (see for instance Hou, Xue, and Zhang (2015) and the references therein).

Our paper further relates to Warusawitharana and Whited (2016), who consider the influence of equity misvaluation on a representative firm. Our framework differs from theirs in that we focus on the effects of *cross-sectional* mispricings on real quantities. Even if the representative firm is correctly priced, these cross-sectional anomalies can have large effects. To our knowledge, we are the first to quantitatively assess the importance of cross-sectional mispricings. To be able to study this, the paper also presents a methodological contribution by presenting a simple continuous time framework that features many of the important characteristics that have been posed in the literature, yet still maintains a simple tractable setting that allows us to evaluate the stationary cross-sectional distribution of the quantities of interest in closed-form.

Several papers in the literature already provide evidence that firms respond to stock market valuations and/or equity mispricings. Barro (1990) shows that changes in stock prices have substantial explanatory power for U.S. investment, especially for long-term samples, and even in the presence of cash flow variables. The specification he employs outperforms standard Tobin's Q regressions. Polk and Sapienza (2009) use discretionary accruals as a proxy for mispricing and find a positive relation between abnormal investment and discretionary accruals. Baker, Stein, and Wurgler (2003) test the prediction that stock prices have a stronger impact on the investment of equity-dependent firms – firms that need external equity to finance marginal investments — and find strong support for it. Gilchrist, Himmelberg, and Huberman (2005) argue that dispersion in investor beliefs and short-selling constraints can lead to stock market bubbles and that firms, unlike investors, can exploit such bubbles by issuing new shares at inflated prices. This lowers the cost of capital and increases real investment. They use the variance of analysts earnings forecasts to proxy for the dispersion of investor beliefs, and find that increases in dispersion cause increases in new equity issuance, Tobins Q, and real investment, as predicted by their model. Goldstein, Ozdenoren, and Yuan (2013) study a model in which a capital provider learns from the price of a firm's security in deciding

how much capital to provide for new investment. This feedback effect from the financial market to the investment decision gives rise to trading frenzies, in which speculators all wish to trade like others, generating large pressure on prices. Chen, Goldstein, and Jiang (2007) show that two measures of the amount of private information in stock priceprice nonsynchronicity and probability of informed trading (PIN)have a strong positive effect on the sensitivity of corporate investment to stock price. They argue that firm managers learn from the private information in stock price about their own firms fundamentals and incorporate this information in the corporate investment decisions. Even though these papers argue that firm investment is affected by stock market fluctuations, they do not provide estimates of the associated loss in aggregate value added.

3. Reduced-Form Estimates of Distortions and Their Shortcomings

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

3.1. Alphas

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

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

and plot the intercept of this regression $\alpha_{i,btm}$ in Figure I. The figure shows values ranging from -2% to 7% for BtM sorted portfolios. We then repeat the same analysis for momentum-sorted portfolios (the return over the past 12 months). For momentum, alphas range from -12% to 5%. To get a first sense of the aggregate mispricing (MP)we

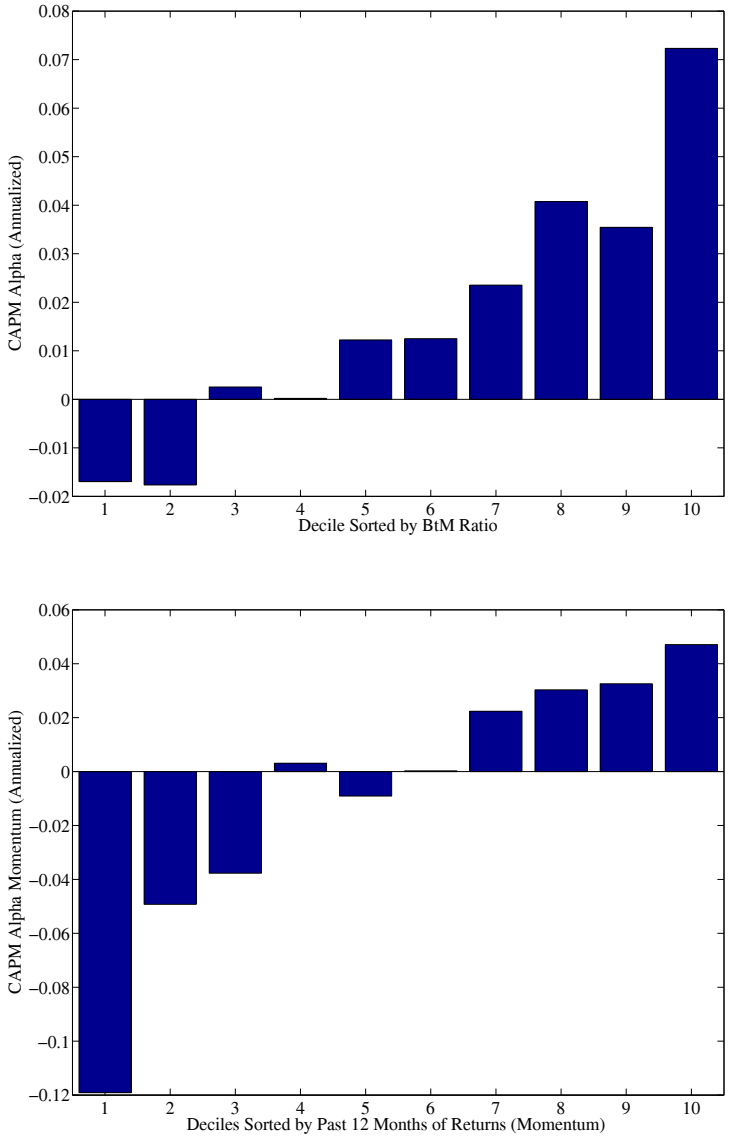


FIGURE I
Annualized CAPM Alphas of Book-to-Market (Value) and Momentum Sorted Portfolios

compute the average absolute value of the alpha for BtM sorted portfolios:

$$MP_{btm} = \frac{\sum_{i=1}^{10} \mid \alpha_{i,btm} \mid}{10} = 2.3\%, \tag{2}$$

as well as momentum sorted portfolios:

$$MP_{mom} = \frac{\sum_{i=1}^{10} \mid \alpha_{i,mom} \mid}{10} = 3.5\%. \tag{3}$$

One potential downside of this measure of aggregate mispricing, however, is that it does not properly account for size differences across the deciles. That is, if most (least) capital is concentrated in the deciles with the least mispricing, the measure in (2) overstates (understates) the amount of mispricing. To address this issue, we recompute the measures above using the aggregate market value of equity (E) of the decile as weights in the computation. Define the weights:

$$we_{i,btm} = \frac{1}{T}\sum_{t=1}^{T} \frac{E_{i,btm}}{\sum_{i=1}^{10} E_{i,btm}}, \tag{4}$$

then the equity-weighted mispricing measure is given by:

$$EMP_{btm} = \sum_{i=1}^{10} we_{i,btm} \mid \alpha_{i,btm} \mid = 1.6\%. \tag{5}$$

The equivalent measure for momentum is:

$$EMP_{mom} = 2.4\%. \tag{6}$$

Both measures are lower than their equal-weighted counterparts, indicating that most mispricing is concentrated in deciles with small equity capitalziation.

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

$$wm_{i,btm} = \frac{1}{T}\sum_{t=1}^{T} \frac{E_{i,btm} + L_{i,btm}}{\sum_{i=1}^{10} E_{i,btm} + L_{i,btm}}, \tag{7}$$

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

$$VMP_{btm} = \sum_{i=1}^{10} wm_{i,btm} \mid \alpha_{i,btm} \mid = 2.3\%. \tag{8}$$

For momentum this measure equals:

$$VMP_{mom} = 2.4\%. \tag{9}$$

Interestingly, we find that by weighting by total firm size, the weighted mispricing is about the same size as the unweighted one.

Even though these measures could be useful as first estimates of the importance of mispricing, they miss one important feature which is the persistence of the mispricing. Momentum is a short-lived phenomenon, whereas value is a long-lived one. One way to assess the persistence is to compute Markov transition matrices that summarize how firms migrate across deciles. In Table 1 we report these Markov matrices for both value and momentum strategies. As expected, the table shows that momentum is a much shorter-lived anomaly than the value anomaly. The average diagonal element for value is 0.34, with values above 0.5 for the two extreme portfolios. For momentum the average diagonal element is 0.12 and the first diagonal element (the (1,1) element) equals 0.17. Given that momentum is so much less persistent than value we would expect it to have relatively little influence on firm policy.

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

3.2. A Different Counterfactual

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

$$V_{i0}(\alpha_{i0}) = \frac{E_0\left[CF_{i1}\right]}{1 + r_f + \beta_i\psi_0 + \alpha_{i0}}. \tag{10}$$

where r_f is the risk free rate, ψ_0 is a measure of the risk premium and β_i measures the usual scaled covariation with the stochastic discount factor. Interpreting α_{i0} as a distortion in the discount rate is isomorphic to a distortion in the beliefs (probabilities) regarding the cash flows. Suppose we observe $\alpha_{i0} = \phi$ in the data. It is clear that the

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	Exit
Dec 1	0.577	0.215	0.071	0.033	0.015	0.008	0.006	0.004	0.006	0.009	0.056
Dec 2	0.160	0.352	0.216	0.100	0.044	0.025	0.012	0.007	0.010	0.013	0.059
Dec 3	0.044	0.182	0.289	0.198	0.100	0.045	0.025	0.015	0.016	0.017	0.069
Dec 4	0.018	0.067	0.177	0.258	0.195	0.099	0.041	0.026	0.024	0.023	0.072
Dec 5	0.009	0.028	0.074	0.172	0.246	0.197	0.089	0.045	0.036	0.030	0.075
Dec 6	0.005	0.014	0.035	0.079	0.171	0.244	0.188	0.094	0.061	0.038	0.071
Dec 7	0.003	0.008	0.016	0.036	0.077	0.168	0.264	0.207	0.099	0.048	0.075
Dec 8	0.003	0.005	0.013	0.021	0.039	0.079	0.198	0.302	0.194	0.065	0.083
Dec 9	0.003	0.006	0.011	0.019	0.031	0.056	0.088	0.182	0.338	0.178	0.087
Dec 10	0.006	0.008	0.011	0.018	0.024	0.031	0.039	0.059	0.152	0.527	0.124
Momentum Sorted Portfolios											
From/To	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10	Exit
Dec 1	0.176	0.110	0.082	0.056	0.049	0.047	0.045	0.048	0.066	0.129	0.191
Dec 2	0.122	0.115	0.099	0.089	0.076	0.074	0.069	0.075	0.088	0.107	0.085
Dec 3	0.084	0.100	0.104	0.097	0.093	0.088	0.089	0.093	0.090	0.092	0.070
Dec 4	0.063	0.085	0.092	0.105	0.108	0.109	0.107	0.099	0.095	0.078	0.060
Dec 5	0.056	0.075	0.090	0.105	0.108	0.117	0.113	0.108	0.096	0.074	0.057
Dec 6	0.049	0.071	0.088	0.099	0.111	0.122	0.119	0.114	0.099	0.071	0.057
Dec 7	0.046	0.072	0.087	0.103	0.113	0.112	0.114	0.115	0.104	0.073	0.061
Dec 8	0.054	0.074	0.088	0.101	0.108	0.106	0.116	0.108	0.099	0.079	0.068
Dec 9	0.069	0.087	0.093	0.095	0.099	0.097	0.098	0.098	0.097	0.093	0.074
Dec 10	0.123	0.114	0.099	0.083	0.077	0.068	0.069	0.075	0.093	0.116	0.084

Table 1
Annual Markov Matrices of Decile Portfolios Sorted by Book-to-Market and Momentum.

value of the firm V_{i0} is affected by this distortion at time 0. However, as long as the actual cash flows of the firm are not affected by α_{i0}, and thus real economic quantities are unaffected, the misvaluation will resolve itself at time 1 through the higher return, and no further losses to the economy occur. Computing the ratio of $V_{i0}(\alpha_{i0} = 0)$ to $V_{i0}(\alpha_{i0} = \phi)$ in this one-period example is equally informative as ϕ itself.

In this paper, we are interested in the value distortions that happen when the cash flows of the firm *are* affected by α_{i0}. That is, we model the real investment decisions of the firm as a function of α_{i0}, i.e. $CF_{i1}(\alpha_{i0})$.[4] We then compare the valuation under the actual firm policies:

$$V_{i0}^{act} = \frac{E_0\left[CF_{i1}(\alpha_{i0} = \phi)\right]}{1 + r_f + \beta_i \psi_0} \tag{11}$$

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

to the valuation under the optimal firm policies:

$$V_{i0}^{opt} = \frac{E_0\left[CF_{i1}(\alpha_{i0}=0)\right]}{1+r_f+\beta_i\psi_0}. \tag{12}$$

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

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

4. The Model

The economy we study is in continuous time. We consider an AK model for firms with adjustment costs that features a particular stochastic structure that ensures computational tractability and exact solutions for the stationary distribution of all quantities in the presence of complex dynamics for various structural parameters. After all, characterizing the whole cross-section of investment and value added is required to properly assess the influence of cross-sectional mispricings on real economic quantities. The structural parameters of the model are governed by a set of exogenous state variables, which are described in detail in Section 4.2 below. For notational convenience, we will omit parameters' functional dependence on these states elsewhere in the model description.

4.1. Firm technology

The firm generates an output flow $AK^{\eta}dt$, where K denotes the firm's capital stock. The capital stock is affected by firm investment I and depreciation shocks. Capital K takes values in a discrete set indexed by $\kappa \in \Omega_\kappa = \{1, 2, ..., N_\kappa\}$, where K is given by:

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

By choosing Δ small enough the model can approximate a model with a continuous support for K arbitrarily well. The discrete state space structure, however, increases the computational tractability of the model and allows obtaining exact solutions.[5]

When the firm chooses an expected investment rate $i_+ = \frac{E[I]}{K} \geq 0$ it incurs a cost that is quadratic in this rate: $(c_1 i_+ + c_2 i_+^2)K$. Investing yields an increase in capital K:

$$Ke^{\Delta} = K + I,$$

with Poisson intensity $i_+/(e^{\Delta} - 1)$.[6] When the firm has reached the upper bound of the capital stock, $K(N_\kappa)$, investment is assumed to be ineffective in generating further increases in capital. By choosing N_κ high enough, this boundary will have no effects on the results, as optimal investment will be zero above some endogenous threshold for capital.

Capital depreciates stochastically from K to $Ke^{-\Delta}$ with hazard rate $\frac{\delta}{(1-e^{-\Delta})}$, except at the lower boundary, $K(1) = K_l$. Here, we assume that a firm fails if it is hit by another depreciation shock. In all other states, the expected depreciation rate is thus δ.

The firm can also search for buyers of its capital in order to disinvest. Choosing an expected disinvestment rate $i_- \geq 0$ leads to a decline in capital from K to $Ke^{-\Delta}$ with hazard rate $\frac{i_-}{(1-e^{-\Delta})}$. Trying to find a buyer is costly, leading to a quadratic search cost of $c_{2-} \cdot i_-^2 \cdot K dt$. The expected revenue from capital sales over an instant is $c_{1-} \cdot i_- \cdot K dt$.[7] We assume that disinvestment is impossible when capital has reached the lower bound $K = K_l$. Again, by choosing K_l low enough we can ensure that the firm would never optimally attempt to disinvest at the lower bound in any case, so that this restriction is also non-binding.

Let N^+ and N^- denote counting processes that keeps track of successful capital acquisitions and sales, and let N^δ denote a counting process that keeps track of depreciation shocks. Capital evolves according to a jump process:[8]

$$d\log(K_t) = \Delta(dN_t^+ - dN_t^- - dN_t^\delta),$$

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

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

[7]In principle, we can capture mergers and acquisition as well as large asset sales in this fashion as well (jumps by $N \cdot \Delta$).

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

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

4.2. Exogenous State Variables

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

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

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

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

$$dY_t/Y_t = \mu(Z_t)dt + \sigma(Z_t)dB_t.$$

The variable Y can capture a macro trend growth. Y is assumed to enter both a firm's cost function and its output linearly, that is, the cost function parameters c_1 and c_2 and the TFP variable A are all linear in Y. The trend reflects a gradual increase in output and the price of capital.

4.3. Pricing

Undistorted stochastic discount factor. We consider a partial equilibrium analysis in the sense that we quantify efficiency losses, taking as given a particular stochastic

discount factor (SDF). We assume that the SDF follows a Markov-modulated jump diffusion process,

$$\frac{d\xi_t}{\xi_{t-}} = -r_f\left(Z_{t-}\right)dt - \nu(Z_{t-})dB_t + \sum_{Z'\neq Z_{t-}} (e^{\phi(Z_{t-},Z')} - 1)dM_t\left(Z_{t-}, Z'\right), \tag{13}$$

where r_f is the risk free rate, ν is the price of risk for aggregate Brownian shocks, $\phi(Z, Z')$ is a jump risk premium, and $dM(Z, Z')$ is a compensated Poisson process capturing switches between the macroeconomic Markov states Z and Z'.[9] Let $\tilde{\Lambda}_Z$ denote the generator under the risk neutral measure, that is, $\tilde{\lambda}(Z, Z') = e^{\phi(Z,Z')}\lambda(Z, Z')$.[10]

Mispricing. Firms can face a distorted cost of capital, as characterized by a firm's instantaneous α. The parameter α can depend on both the firm-specific state z and the macro state Z.

5. Analysis

5.1. Positive Analysis

In this subsection we analyze firm behavior in the presence of pricing distortions (α) that all agents, including firm managers, mistakenly view as a part of a firm's fair discount rate. As argued before, this interpretation is isomorphic to biased beliefs (distortions) regarding the probability distribution of cash flows. Managers take the prices of the financial market as given and maximize the going market value of the firm under this distorted valuation.

Let the firm's value function be denoted by $V(\kappa, z, Z, Y)$, where

$$V(\kappa, z, Z, Y) = \max_{i_+, i_- \geq 0} \mathbb{E}_t \int_t^\infty \frac{\xi\left(Z_u\right)}{\xi\left(Z_t\right)} e^{-\int_t^u \alpha(z_s)ds} \pi(u)du, \tag{14}$$

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

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

and the conditional expected after-tax profit:

$$\pi(t) = (1-\tau)AK_t^{\eta} - (c_{1+}^{\tau} i_t + c_{2+}^{\tau} i_t^2 - c_{1-}^{\tau} i_- + c_{2-}^{\tau} i_-^2)K_t + \tau\delta K_t. \tag{15}$$

where we define the tax-adjusted cost parameters:

$$c_{1+}^{\tau} = c_{1+} - \tau(c_{1+} - 1), \tag{16}$$
$$c_{2+}^{\tau} = (1-\tau)c_{2+}, \tag{17}$$
$$c_{1-}^{\tau} = c_{1-} + \tau(1 - c_{1-}), \tag{18}$$
$$c_{2-}^{\tau} = (1-\tau)c_{2-}. \tag{19}$$

Since the parameters A, c_{1+}, c_{2+}, c_{1-}, and c_{2-} are assumed to be linear in the trend component Y, we can conjecture that the value function is linear in Y, that is, $V(\kappa, z, Z, Y) = Y \cdot \tilde{V}(\kappa, z, Z)$. Let a tilde indicate that a variable or parameter is scaled by Y. Then the Hamilton-Jacobi-Bellman equation implies that $\tilde{V}(\kappa, z, Z)$ solves the following set of equations:[11]

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

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

The first-order conditions yield the optimal controls for investment and disinvestment:

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

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

Note that conditional on investment and disinvestment the system is linear in v, which will make determining an exact solution easy and fast.

[11] See Appendix A for details.

Boundaries of the state space. If the firm is in state $\kappa = 1$ and receives another depreciation shock it fails (has zero value). In the state $\kappa = N_\kappa$ we have $i_+ = 0$.

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

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

Quantitatively, most of the risk premium will be driven by the jump risk premia (the growth rate risk), since most of the firm dynamics are driven by changes in Markov states rather than innovations to the common trend Y. Both betas and aggregate risk premia are state-dependent.

5.2. Stationary Distribution with Entry and Exit

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

6. Estimating the Model

6.1. Specification of the Markov Processes

As a simple starting point, we consider a parsimonious specification of the model where it is useful to represent the firm-specific state z as a tuple $(\tilde{A}, g)$ where the productivity parameter $\tilde{A}$ can take values in the set $\{\tilde{A}_1, ..., \tilde{A}_6\}$ and where the state g governs not only a firm's transition rates between productivity parameters $\tilde{A}$ but also determines the values of other firm-specific parameters such as depreciation, adjustment cost associated with investment, and alphas. As a first step, we will consider only two g regimes.

6.2. Calibration and Estimation

We calibrate the parameters of the macroeconomy (relating to the state Z) based on the existing literature (e.g. Chen, Cui, He, and Milbradt (2015)) and estimate the firm-specific parameters using a method of moments approach. When estimating parameters it is useful to specify boundaries of the capital grid ($K(1)$ and $K(N_\kappa)$) that are sufficiently extreme to be almost never reached by a firm and that can be computed quickly in closed-form. For this purpose we compute hypothetical steady-state values of capital that would obtain if a firm stayed in the the worst or best technology state forever (see Appendix B for details). We assume that capital takes value on a grid with $N_\kappa = 70$ points, which also pins down the value of Δ (see caption Table 2).

Targeted moments. We estimate 23 parameters by targeting 34 moments related to the cross-sectional distribution of firms: 9 book to market decile breakpoints, 9 sales-to-book breakpoints, 6 book capital breakpoints and 10 CAPM alphas corresponding to the book-to-market deciles. The calibrated and estimated parameters are summarized in Table 2 and all fall in a range that is broadly consistent with the literature. Table 3 summarizes the model moments and compares them to the data moments. The table also reports the error bounds around the data point estimates. As the table shows, even this very parsimonious model with a sparse number of states does a reasonable job fitting the cross-sectional moments in the data. Where the model fails is in the extreme tail of both book-to-market as well as book values. This fit could be enhanced by expanding the state space.

Parameters of the Macroeconomy (Calibrated)			
Parameter	Variable	Boom $Z = G$	Recession $Z = B$
Transition rates for aggregate states	λ	0.1	0.5
Trend growth	μ	0.03	-0.01
Trend risk exposure	σ	0.16	0.16
Risk-free rate	r_f	0.02	0.02
Local risk price	ν	0.165	0.255
Jump-risk premium	$\exp(\phi)$	2.0	0.5
Tax rate (personal + corporate)	τ	0.53	

Firm-specific Parameters (Estimated)			
Parameter	Variable	g_1	g_2
Rate of leaving g-state	$\lambda_{g,g'}$	0.31	0.50
Rate of moving to next-higher $\tilde{A}$	h^{up}	0.51	0.45
Rate of moving to next-lower $\tilde{A}$	h^{dn}	0.51	0.51
Abnormal return	α	-0.11	0.10
Depreciation rate	δ	0.11	0.01
Purchase price of capital	c_{1+}	1.25	1.00
Upward adjustment cost	c_{2+}	4.83	4.83
Sales price of capital	c_{1-}	0.50	
Downward adjustment cost	c_{2-}	60.25	
Decreasing returns to scale	η	0.88975	

Table 2

Parameters. The two tables list parameters of the macroeconomy and firm-specific parameters. The parameters of the macroeconomy are calibrated. We estimate firm-specific parameters via a method of moments approach. Factor productivity is estimated to be $\tilde{A}_i = A_1 e^{\sum_{j<i} a_j}$ where $a_j \in \{1.2017, 1.4045, 1.503, 0.81715, 1.028\}$ and $A_1 = 0.0046183$. The capital grid is specified by $K_l = 1.166710^{-8}$, $N_\kappa = 70$, and $\Delta = 0.6093$. We consider two g regimes, denoted by g_1 and g_2 which govern the parameters in the bottom panel.

Book-to-Market Breakpoints										
Breakpoint	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
Model	0.265	0.35857	0.456	0.550	0.695	0.86586	1.085	1.376	1.7197	
Data	0.348	0.505	0.631	0.739	0.834	0.919	0.996	1.06	1.1951	
Data std.	0.117	0.136	0.139	0.133	0.119	0.10497	0.0954	0.105	0.137	
				Book Breakpoints						
Breakpoint	0.4	0.8	0.9	0.95	0.975	0.99				
Model	0.5509	6.303	21.32	39.21	39.21	72.13				
Data	0.586	7.496	25.39	60.346	129.51	343.58				
Data std.	0.0260	1.496	8.193	20.84	44.57	98.52				
				Alphas by Book-to-Market Decile						
Decile	1	2	3	4	5	6	7	8	9	10
Model	-0.00214	-0.000727	0.000797	0.00101	0.00172	0.000960	0.00127	0.00117	0.00248	0.00543
Data	-0.00183	-0.00130	-0.000237	-0.000208	0.00119	0.00103	0.00173	0.00285	0.00251	0.00475
Data std.	0.00120	0.00072	0.00072	0.00074	0.00092	0.00099	0.0011	0.0013	0.00152	0.00197

Table 3
Moments used in the estimation of firm-specific parameters.

Entry and exit rates. In addition to the parameters listed in Table 2 we directly specify entry and exit rates. We impose that each firm that exits is replaced by another firm in the exact same state. To match exit rates in the data we estimate historical (average) exit rates in each sales-to-book decile. A firm's exogenous exit rate is then determined via interpolation as a function of its sales-to-book ratio.

7. Results

7.1. Under- and Overinvestment

First we assess the influence of cross-sectional distortions on investment. The bottom panel of Figure II plots the cumulative distribution function of the log investment rate in the model. As can be seen from the figure, only about 44% of firms are investing and the remaining 56% are not. Secondly, only 10% of firms have an investment rate that is higher than 30% (i.e. a log investment rate of -1.2 on the x-axis). The top panel of Figure II We plot optimal log investment rates from the model on the x-axis against distorted log investment rates on the y-axis. Each dot represents a state, and the color of each dot represents the probability of that state occurring, as indicated by the colorbar on the right hand side. The figure shows that there is substantial heterogeneity across firms due to distortions. For example, for an optimal log investment rate of -2 (an investment rate of 13.5%), an actual log investment rate of -3 (an investment rate of 5%) implies under-investment by 8.5%.

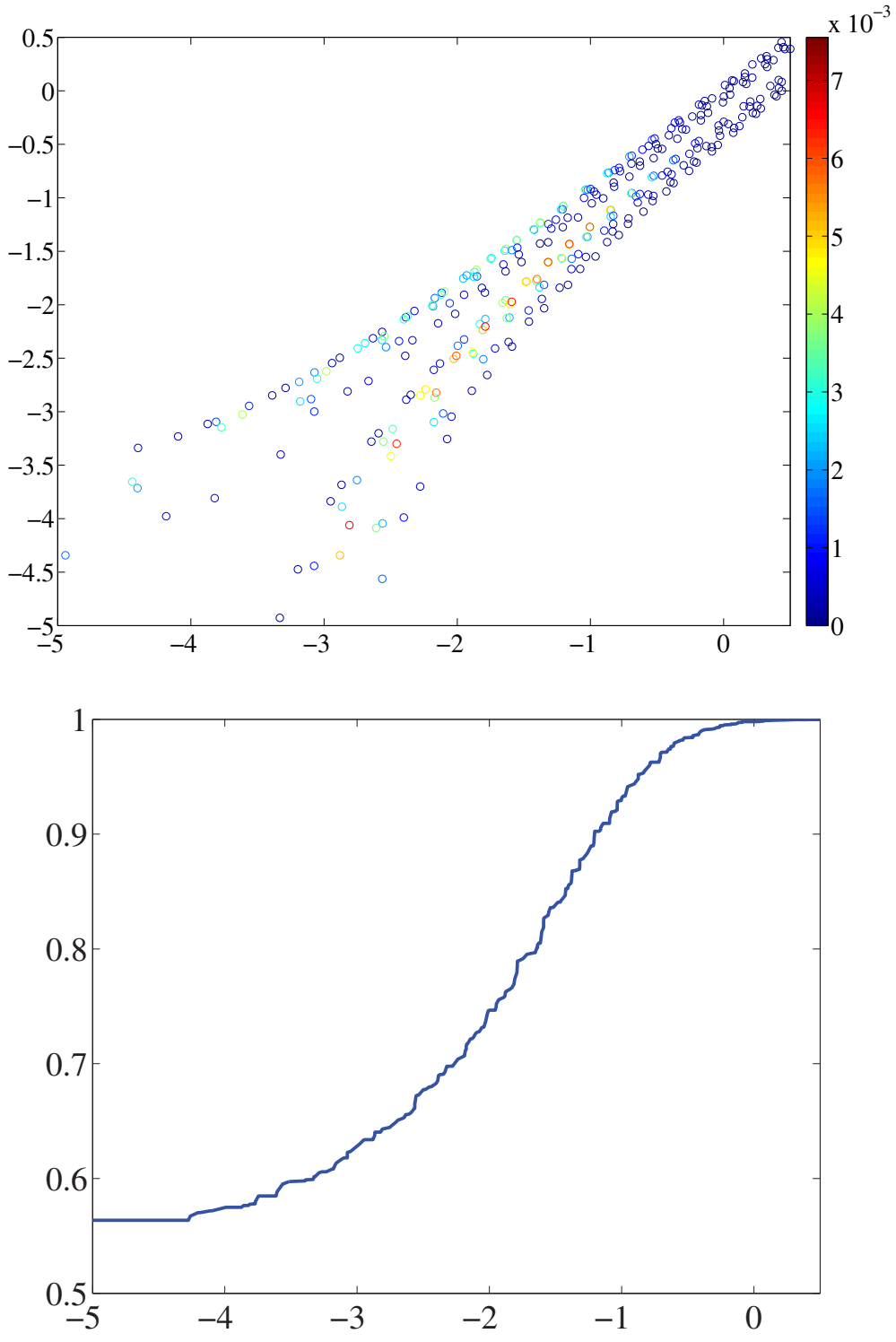


FIGURE II

Model: Investment under Actual vs. Optimal Policy. The first graph plots $\log(i_+)$ in the model under the actual (distorted) policy (vertical axis) against $\log(i_+)$ under the optimal policy (horizontal axis). The second graph plots the CDF of $\log(i_+)$ under the optimal policy. Note that more than 50% of firms do not invest (that is, $i_+ = 0$), which is why the plotted CDF starts at a value above 0.5.

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

7.2. Unconditional efficiency measures

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

If, starting today, firms switch from following suboptimal policies to following optimal policies the present value of surplus rises, on average, by:

$$\frac{p^{act} \cdot V^{opt}(\alpha = 0)}{p^{act} \cdot V^{act}(\alpha = 0)} - 1 = 1.31\%. \tag{23}$$

Alternatively, we can scale the expected increase in the true market capitalization by the expected value added as measured by the difference between firm market values and firm book values. This measure yields:

$$\frac{p^{act} \cdot V^{opt}(\alpha = 0) - p^{act} \cdot V^{act}(\alpha = 0)}{p^{act} \cdot (V^{opt}(\alpha = 0) - K)} = 5.41\%.$$

In other words, when compared to the surplus that installed capital creates $V^{opt}(\alpha = 0) - K$ the value gain represents an even higher fraction.

7.3. Sensitivity Analysis

In this section we consider the sensitivity to varying the persistence of the g-regime. Recall that there are two regimes that influence firms' growth opportunities and mispricing, as well as several other parameters summarized in Table 2. Changing the persistence of this Markov process allows us to gauge how important the persistence of an anomaly is for the aggregate value losses. Figure III plots the change in value as computed in

Equation 23 for different levels of persistence. The x-axis is the multiplier on the transition rates $\lambda_{g,g'}$ of the baseline parameterization by a factor that ranges between 0.1 and 4. When the multiplier is 1, we obtain the baseline value loss of 1.31%. The graph shows that the value losses are highly sensitive to the persistence of the anomaly, thereby confirming the intuition that non-persistent anomalies, such as the momentum effect, are unlikely to have a large effect on value added. On the other hand, if anomalies are more persistent than the value premium effect, they can create very large real inefficiencies.

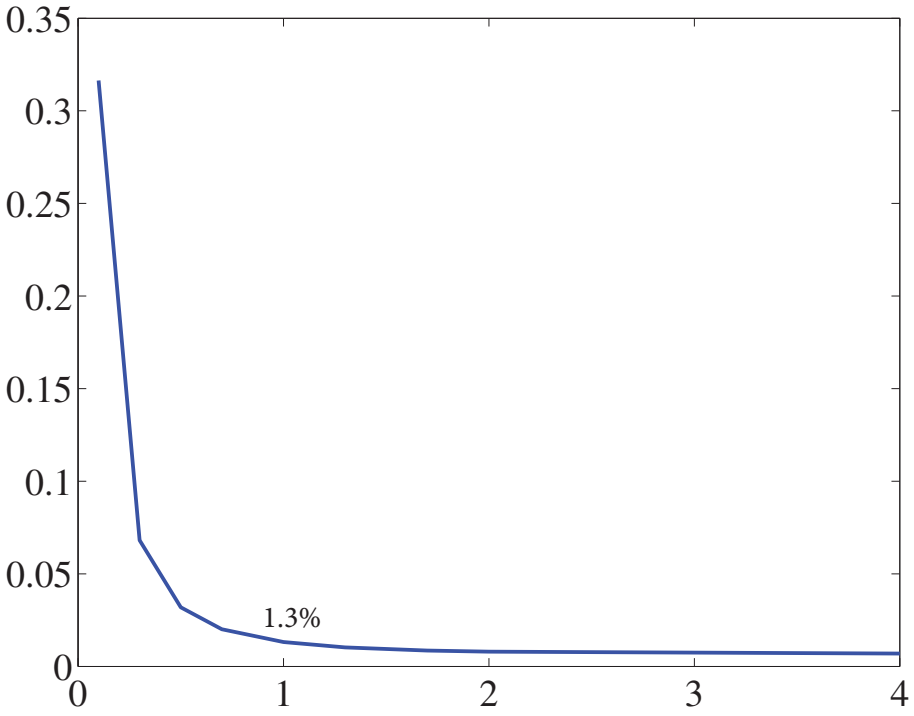


FIGURE III

Changing g-state persistence. The graphs illustrate the effects of changing the persistence of the g-state by multiplying the transition rates $\lambda_{g,g'}$ of the baseline parameterization by a factor $[0.1, 4]$. The first graph plots the present value of gains as measured by equation (23).

8. Conclusion

Cross-sectional stock pricing anomalies are a widely studied topic. A large fraction of this literature studies the implications of these anomalies for investors, and whether or not these strategies are implementable and profitable after taking into account transaction costs. Another important fraction studies whether or not the first order conditions of the firm are consistent with the observed financial returns and argues that even if the first order conditions of consumers are not able to price assets appropriately, at least

the investment behavior of corporate managers seems more consistent with the observed return patterns.

In this paper we ask a more fundamental question. Taking as given that managers respond to financial markets, we investigate to what extent cross-sectional anomalies in financial markets can lead to distortions in the real economy. In order to assess distortions quantitatively, we estimate the joint dynamic distribution of firm characteristics that have been linked to financial imperfections and other firm variables, such as investment, capital, and output. Based on a model that matches these joint dynamics we then evaluate the counterfactual dynamic distribution of capital, investment, and value added absent financial market imperfections and find that they can cause large and persistent deviations. Just the value premium effect represents up to 1.3% in lost value added. This implies that financial intermediaries that can reduce and/or eliminate such market imperfections can provide large value added to the economy. As such, our paper contributes to the debate on the role and optimal size of the financial sector.

Even though we find that financial intermediaries can potentially add a lot of value to the economy by resolving anomalies, we do not show that they are currently engaged in that activity. That is, it is unclear how large cross-sectional anomalies would be absent these financial intermediaries. What is clear, is that using Sharpe's arithmetic to argue that financial intermediaries such as active mutual funds do not add value is flawed. What our framework does allow us to evaluate is how large and persistent cross-sectional anomalies absent financial intermediation need to be to justify their current size. Furthermore, there may be large aggregate (as opposed to cross-sectional) mispricings, that will further enhance the value that financial intermediaries could add to the economy.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A. Proof: HJB Equation

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

$$\begin{aligned}0 = \max_{i_+, i_- \geq 0} [&\pi(\kappa, z, Z, Y) - (r_f(Z) + \alpha(z, Z))V(\kappa, z, Z, Y) \\ &+ \frac{i_+}{(e^{\Delta} - 1)}(V(\kappa + 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\ &+ \frac{\delta + i_-}{(1 - e^{-\Delta})}(V(\kappa - 1, z, Z, Y) - V(\kappa, z, Z, Y)) \\ &+ \Lambda_Z(Z)\mathbf{V}^Z(z, \kappa, Y) + \Lambda_z(Z)\mathbf{V}^z(Z, \kappa) \\ &+ V_A A\mu(Z) + \frac{1}{2}V_{AA}A^2\sigma(Z)^2 - V_A A\sigma(Z)\nu(Z)], \end{aligned} \tag{24}$$

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

B. Approximate Steady States

Approximate steady states inform our choices for the boundaries of the capital grid. If we set $i_+ = \delta$ and $i_- = 0$ and abstract from Poisson shocks to z and Z we obtain for $\Delta \to 0$:

$$\tilde{V}(K) = \frac{(1 - \tau)\tilde{A}K^{\eta} - (\tilde{c}_{1+}^{\tau}\delta + \tilde{c}_{2+}^{\tau}\delta^2 + \tau\delta)K}{(r_f + \sigma\nu + \alpha - \mu)}$$

and

$$\frac{\partial \tilde{V}(K)}{\partial K} = \frac{\eta(1 - \tau)\tilde{A}K^{\eta - 1} - (\tilde{c}_{1+}^{\tau}\delta + \tilde{c}_{2+}^{\tau}\delta^2 + \tau\delta)}{(r_f + \sigma\nu + \alpha - \mu)}$$

with

$$\bar{i}_+ = \delta = \frac{\frac{\partial \tilde{V}(K)}{\partial K} - \tilde{c}_{1+}}{2\tilde{c}_{2+}}$$

This yields the steady state capital $\bar{K}$:

$$\bar{K} = \left(\frac{\eta(1 - \tau)\tilde{A}}{(2\delta\tilde{c}_{2+} + \tilde{c}_{1+})(r_f + \sigma\nu + \alpha - \mu) + \tilde{c}_{1+}^{\tau}\delta + \tilde{c}_{2+}^{\tau}\delta^2 + \tau\delta} \right)^{\frac{1}{1 - \eta}}.$$

C. Stationary Distribution

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

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

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

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

$$\mathbb{E}\left[d\left(\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\right)\right] = 0 \tag{25}$$

$$\begin{aligned}
\mathbb{E}\left[d\left(\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\right)\right] =& \mathbb{E}\left[\frac{dm}{\mathbf{1}'\tilde{m}} - \frac{\tilde{m}}{\mathbf{1}'\tilde{m}}\frac{\mathbf{1}'dm}{\mathbf{1}'\tilde{m}}\right] && (26)\\
=& (\mathbf{I}_{N_s} - p\mathbf{1}')\frac{\mathbb{E}\left[dm\right]}{\mathbf{1}'\tilde{m}} && (27)\\
=& (\mathbf{I}_{N_s} - p\mathbf{1}')\Lambda'\frac{\tilde{m}}{\mathbf{1}'\tilde{m}}dt && (28)\\
=& (\mathbf{I}_{N_s} - p\mathbf{1}')\Lambda' p dt && (29)
\end{aligned}$$

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

$$(\Lambda' - \mathbf{I}_{N_s}\mathbf{1}'\Lambda' p)p = \mathbf{0} \tag{30}$$

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

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

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

$$\begin{pmatrix} \Lambda' \\ \mathbf{1}' \end{pmatrix} p = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}. \tag{31}$$

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