Keloharju, Matti; Linnainmaa, Juhani T.; Nyberg, Peter

Long-Term Discount Rates Do Not Vary Across Firms

Published in:
Journal of Financial Economics

DOI:
10.1016/j.jfineco.2021.04.031

Published: 01/09/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2021). Long-Term Discount Rates Do Not Vary Across Firms. *Journal of Financial Economics*, *141*(3), 946-967. https://doi.org/10.1016/j.jfineco.2021.04.031

Long-term discount rates do not vary across firms*

Matti Keloharju

Juhani T. Linnainmaa

Peter Nyberg

September 2020

Abstract

Long-term expected returns do not appear to vary in the cross section of stocks. We show that even negligible persistent differences in expected returns, if they existed, would be easy to detect. Markers of such differences, however, are absent from actual stock returns. Our results are consistent with behavioral models and production-based asset pricing models in which firms' risks change over time. Consistent with the lack of long-term differences in expected returns, persistent differences in firm characteristics do not predict the cross section of stock returns. Our results imply stock market anomalies have only a limited effect on firm valuations.

JEL classification: G12; G31

*Keloharju is with Aalto University School of Business, CEPR, and IFN. Linnainmaa is with the Dartmouth College and NBER. Nyberg is with the Aalto University School of Business. We thank Jonathan Berk, John Cochrane, Serhiy Kozak, Mathias Kronlund, Jon Lewellen, Yan Liu (discussant), Anders Löflund, Owen Lamont, Stijn Van Nieuwerburgh (discussant), Shri Santosh (discussant), an anonymous referee, and conference and seminar participants at Florida International University, Lund University, Tinbergen Institute at Amsterdam, University of Illinois at Chicago, University of Illinois at Urbana-Champaign, University of Miami, University of Michigan, LA Finance Day, Society of Financial Studies 2018 Cavalcade, European Finance Association 2019 Meetings, and American Finance Association 2020 Meetings for insightful comments and Jonathan Berk, Timothy Johnson, and Lu Zhang for providing the code for the models simulated in Sections 2 and 5. Linnainmaa and Nyberg are also affiliated with Citadel and Linnainmaa with Research Affiliates. Neither Citadel nor Research Affiliates provided any funding for this research. Financial support from the Academy of Finland and Nasdaq Nordic Foundation is gratefully acknowledged.

1 Introduction

Time-varying risks lie at the heart of production-based asset pricing models. Firms' risks can change either because their asset bases change (Berk, Green, and Naik 1999; Gomes, Kogan, and Zhang 2003) or because firms face systematic productivity shocks and costly reversibility of investment (Zhang 2005; Cooper 2006). If risks change but are stationary, as they are in these models, today's high-risk firms are expected to become less risky and low-risk firms are expected to become riskier. As risks converge, so do expected returns.

The prediction that differences in expected returns die out is not specific to rational models; it is, in fact, inherent to any behavioral explanation for asset pricing anomalies. If a stock's expected return is too high because the stock is underpriced, the return can remain elevated only for as long as the mispricing persists. If, for example, a stock is underpriced by 50% relative to its fair value and its expected return is consequently 10% too high, the mispricing would vanish in 6.9 years.[1] The prediction that expected returns converge is therefore very general. In risk-based models this prediction requires only that the risks that firms face change over time; in behavioral models, it is true by assumption.

In this paper, we first show the data strongly support the idea of cross-sectional differences in expected returns converging to zero over time. Even if a stock's expected return today lies far above or below the average, we cannot reject the null that its expected return after five years equals the average. Second, we show the convergence in expected returns is commensurate with that in risks, regardless of whether we measure them using betas or firm characteristics. Thus, today's high-risk firms indeed become less risky and low-risk firms become riskier.

In Panel A of Figure 1, we use U.S. stock data and assign stocks into deciles based on estimates of

[1] Suppose that a stock's fair value is P and, at that price, it would earn an expected return of r. If the stock is underpriced by 50% so that its expected return exceeds r by δ then, with continuous compounding, the mispricing will vanish in t years, with t satisfying $Pe^{rt} = \frac{1}{2}Pe^{(r+\delta)t}$; with $\delta = 0.1$, $t = \ln(2)/\delta = 6.93$ years.

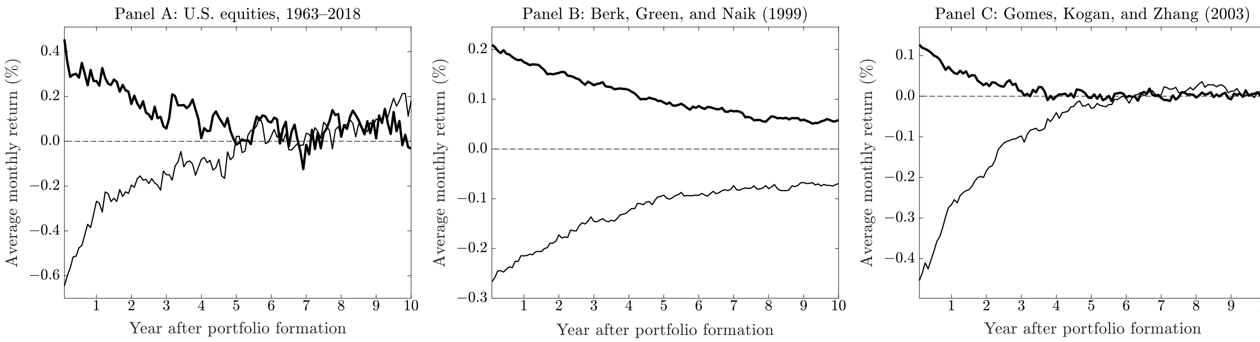

Figure 1: **Average monthly returns on stocks sorted by expected returns.** Panel A assigns U.S. stocks into deciles each month from 1963 through 2018 based on a combination of 55 return predictors and plots the average (non-cumulative) monthly returns for the top and bottom deciles over the next ten years. Panels B and C simulate data from Berk et al.'s (1999) and Gomes et al.'s (2003) models using cross-sectionally demeaned returns. The number of stocks and time periods in the simulations match the average number of stocks and number of months in the U.S. data.

stocks' expected returns. In this computation, which we detail below, we combine 55 return predictors into a proxy for firms' expected returns. We report average *noncumulative* returns for the top and bottom deciles over the next ten years; that is, the month-t estimate is the average return in month t after portfolio formation, not the average return from today to month t. Panels B and C illustrate the extent to which the data conform to two models, Berk, Green, and Naik (1999) and Gomes, Kogan, and Zhang (2003). We simulate data from these models using the original papers' parameters and rank firms into deciles each month based on their expected returns. In both the simulations and actual data, differences in average returns are initially large but collapse to zero as we extend the maturity.[2]

We illustrate our key finding first with a characteristics-free bootstrap procedure. We take the cross section of monthly stock returns, preserve both the covariance structure and distribution of returns, but set all expected returns to zero. We then inject small cross-sectional differences into expected returns and measure our ability to detect these differences. We show returns would be highly predictable even if the persistent differences in expected returns were negligible. Suppose, for example, that the

[2] We compute averages of cross-sectionally demeaned returns to remove market-level variation, which is why the averages in Figure 1 converge towards zero.

cross-sectional standard deviation of the expected monthly stock returns is just 0.4%. This number is small, as it would imply that differences in expected returns would account for just $R^2 = 0.07\%$ of the cross-sectional variance of realized stock returns.[3] Our bootstrap procedure shows that, among all-but-microcaps, the prior five-year return would nevertheless predict the cross section of returns with a t-value of 2.07. In actual data, however, past returns do not positively predict returns outside the prior one-year period (momentum). In the five-year regression, for example, the t-value is -3.50.

Why do past returns predict the cross section of returns with a negative sign? We show that production-based asset pricing models predict a positive association between past and future returns *only* if the differences in discount rates are persistent enough. Long-term reversals emerge when (1) there are large differences in expected returns that (2) rapidly converge to zero. Under these conditions, past returns predominantly measure changes in discount rates. For example, a stock with a high past return likely has experienced a decrease in its expected return—it is *not* one whose expected return was high. This interpretation for long-term reversals differs markedly from the behavioral interpretation proposed by De Bondt and Thaler (1985): returns reverse because expected returns converge towards the mean, not because investors overreact to information. The existence of long-term reversals is thus consistent with our main thesis: long-term discount rates do not vary across stocks.

As in Figure 1, we complement our bootstrapping results by sorting stocks into portfolios by different combinations of return predictors. Although some variables predict the cross section of stock returns a few years out, none of them identify long-term differences in expected returns. When we sort stocks into portfolios by a combination of 55 predictors, the average return difference between the top and bottom deciles is 0.77% in the year following portfolio formation (t-value = 4.90). But by the third

[3] The average cross-sectional variance of monthly U.S. stock returns between July 1963 and December 2018 is 0.0229. If the cross-sectional standard deviation of expected returns is 0.4%, this variation in expected returns would account for $\frac{(0.004)^2}{0.0229} = 0.07\%$ of the cross-sectional variation in stock returns; or, conversely, 99.93% of the cross-sectional variation in monthly stock returns would be unrelated to persistent differences in expected returns.

and fifth years, these return differences fall to just 0.31% (t-value = 2.10) and 0.15% (t-value = 1.12), respectively.

The long-run return estimates are precise enough to bound the amount of cross-sectional variation in expected returns. Consider, for example, the CAPM alphas. In the year following portfolio formation, the 95% confidence interval for the CAPM alpha for the difference between the top and bottom deciles runs from 77 basis points to 130 basis points. In year ten, this confidence interval runs from −21 basis points to 22 basis points. We can thus reliably identify significant differences in short-term expected returns but, as Figure 1 shows, these differences evaporate quickly.

Whereas cross-sectional differences in average returns collapse to zero in five years, some of the differences in betas and firm characteristics are persistent. The difference in the convergence rates of discount rates and characteristics generates the following testable prediction: if the differences in returns die out but those in characteristics do not, the *persistent* differences in characteristics cannot associate with differences in average returns.[4] We test this prediction by decomposing firm characteristics into permanent and transitory components. We find that the transitory differences in characteristics command premiums and discounts that vanish in approximately five years. Persistent differences in characteristics do not associate with any differences in average returns. These results offer further support to our main thesis: there are no persistent cross-sectional differences in average returns.

Literature documents sizable short-term differences in expected returns. Martin and Wagner (2019), for example, estimate that "there is considerably more variation in expected returns... than has previ-

[4] Whether this prediction holds in the data is not a foregone conclusion in the light of existing literature. This literature generally does not differentiate between persistent and transitory sources of risk. For example, in his Presidential Address to the American Finance Association, Cochrane (2011) points out that "[A]n implicit assumption underlies everything we do: expected returns, variances, and covariances are *stable* functions of characteristics such as size and book-to-market ratio" (p. 1062, emphasis ours). When further discussing the valuation effects of various return predictors, he suggests that "long-lasting characteristics are likely to be more important" than short-run predictors such as momentum which lasts less than a year (p. 1064). Our argument is that even if a characteristic is persistent, only the transitory movement in that characteristic can predict differences in returns.

ously been acknowledged." We contribute to the literature by showing a wide gap between short-term and long-term returns—and little evidence of persistent differences in expected returns between firms. Because any differences in expected returns are short-lived, they carry little weight in the average discount rate and in firm valuation.[5]

Assuming all anomalies reflect mispricing gives insight into mispricing's potential economic significance. More persistent anomalies are likely to generate greater distortions in the economy (Binsbergen and Opp 2019). An anomaly that generates an expected return of 10% per annum and lasts for a year causes prices to be wrong by 10%; an otherwise similar anomaly lasting only for one month causes prices to be wrong by less than one percent. If most anomalies are short-lived—as our estimates suggest—the stock market can be inefficient in returns but close to efficient in prices.[6] Its perceived efficiency depends on the investment horizon. An arbitrageur could reap great rewards by trading anomalies but, at the same time, the market would be close to efficient to a buy-and-hold investor.

Some readers suggest our result on the lack of differences in long-term discount rates is not surprising; others suggest it is too surprising to be true.[7] The assumption that long-term discount rates vary in the cross section permeates most areas of research. It features prominently, for example, in portfolio choice theory, measurement of cost of capital, and implied cost of capital and equity duration computations.[8] This assumption also underlies almost all present value computations. Equation $M_t = \sum_{\tau=1}^{\infty} \frac{\mathrm{E}(\tilde{D}_{t+\tau})}{(1+r)^{\tau}}$,

[5]If a firm's continuously compounded discount rate is r over the first year and r' thereafter, its k-year discount rate is $\frac{1}{k}r + \left(1 - \frac{1}{k}\right) r'$ per year. Differences in short-run discount rates carry over to long-run discount rates, but increasing the horizon k dilutes their role.

[6]Black (1986) considers a market in which assets are mispriced by no more than a factor of two efficient: "[w]e might define an efficient market as one in which price is within a factor of 2 of value. . . it seems reasonable to me, in the light of sources of uncertainty about value and the strength of the forces tending to cause price to return to value. By this definition, I think almost all markets are efficient almost all of the time. 'Almost all' means at least 90%."

[7]We discuss the latter argument, which is about unlevered and levered returns, at the end of the paper.

[8]See, for example, MacKinlay and Pastor (2000), Cvitanić, Lazrak, Martellini, and Zapatero (2006), and DeMiguel, Garlappi, and Uppal (2009) (portfolio choice), Gebhardt, Lee, and Swaminathan (2001), Hou, Van Dijk, and Zhang (2012), and Hann, Ogneva, and Ozbas (2013) (implied cost of capital), and Dechow, Sloan, and Soliman (2004) (equity duration). Studies that measure variation in the cost of capital often assume that differences in one-period ahead average returns—such as one-month or one-year returns—represent differences in cost of capital. See, for example, Fama and French (1997, 1999), Francis, Nanda, and Olsson (2008), and Barth, Konchitchki, and Landsman (2013).

which is stated in Chapter 1 in most valuation textbooks, only works if expected returns are constant. When this equation is invoked to justify, e.g., the value, profitability, and investment effects (Novy-Marx 2013; Fama and French 2015), this seemingly innocuous assumption becomes anything but. If expected returns vary, high expected profitability, for example, need not imply that a firm's discount rate today is high.

Although we fail to uncover any evidence of differences in long-term discount rates, this does not imply that there is none: a failure to reject the null hypothesis does not grant us the license to accept it. What we *can* say is that any differences in the long-term rates are significantly smaller than those in short-term rates. Moreover, our three approaches—Fama-MacBeth regressions, portfolio sorts, and permanent-transitory decomposition—all lead us to the conclusion that the differences in long-term rates, *if any*, must be negligible. If these differences are so small that we cannot detect them, it seems reasonable to operate under the null, that is, treat all long-term discount rates as equal.

The assumption that discount rates are constant is one of convenience. Models and methods grow more complicated if expected returns are stochastic. Present value computations no longer separate into the two steps of computing expected cash flows and discount rate; one instead needs to specify how expected returns and cash flows vary and covary.[9] Although researchers assume constant discount rates for tractability, our results suggest we likely sacrifice more than we gain by assuming something that runs significantly counter to the data. Any differences in expected returns last for but a few years; insights derived under the assumption that they are permanent are likely misguided. We do not mind if our conclusions, after the fact, do not seem surprising if this means that researchers will grow reluctant to assume something they should not.

[9] Ang and Liu (2004), for example, compute present values under the assumption of predictable risk premiums and conditional betas. They also discuss other studies that compute present values under other special cases, such as under the assumption that risk premiums vary predictably.

2 Expected returns in production-based asset pricing models

Many production-based asset pricing theories seek to explain cross-sectional return patterns associated with, for example, size and book-to-market. These models feature time-varying risks. In this section, we describe and simulate data from four models. We show that—and explain why—cross-sectional firm-level differences in expected returns in these models tend to vanish.

2.1 Models and mechanisms

Model 1: **Berk, Green, and Naik (1999): assets in place versus growth options.** Firms encounter new projects each period and they accept those with positive NPVs. Projects differ in their amount of systematic risk, and old projects turn obsolete at random. Because new projects are drawn from the same distribution, firms are asymptotically identical. In other words, high-risk firms will, on average, encounter projects that lower their risk, and low-risk firms will tend to encounter projects that increase their risk. Firm valuation in the model buttresses this point. A firm's value is the sum of the value of the assets in place and the value of future growth options. Of these two terms, the value of the future growth options is the same across firms, because all firms expect to encounter the same projects.

Model 2: **Gomes, Kogan, and Zhang (2003): general equilibrium with growth options.** Gomes, Kogan, and Zhang (2003) take the investment mechanism of Berk, Green, and Naik (1999) to general equilibrium. Whereas Berk et al. assume the process describing the pricing kernel, Gomes et al. model the household sector and let the markets clear. Returns in this model are completely described by a conditional CAPM; size and book-to-market predict returns because they correlate with the true conditional market betas. Also similar to Berk et al., new projects are distributed randomly across all firms with equal probabilities. Therefore, all firms derive the same value from future growth options, and expected returns perfectly converge over time.

Model 3: **Zhang (2005) and Lin and Zhang (2013): production shocks and costly reversibility of investment.** Production in this model requires capital, and firm-level productivity is subject to aggregate and idiosyncratic shocks. Firms have to pay to install new capital and to adjust it. The pricing kernel is parametrized directly to have a countercyclical price of risk. Productivity shocks alter firms' riskiness; in bad times, when the price of risk is high, low-productivity firms find it costly to shed unproductive capital. Because productivity shocks mean revert, firms' expected returns converge over time.

Model 4: **Hackbarth and Johnson (2015) and Gu, Hackbarth, and Johnson (2017): operating leverage and real options.** Similar to Zhang (2005), production in this model requires capital, and productivity is subject to aggregate and idiosyncratic shocks. Firms face both quasi-fixed and variable costs for both upward and downward adjustments to capital. The random shocks cause risks and, by extension, expected returns, to mean revert.

2.2 Changes in expected returns

We simulate 1,000 months of return data from the models described above using the parameters used in the original studies; when a study considers multiple sets of parameters, we use those of the baseline specification. Studies use different methods to choose the parameters. They are typically "fixed" based on prior literature, directly estimated, or calibrated to match some features of the data, such as the levels and volatilities of equity premium and interest rate. We discard the first 400 months of data to ensure the simulations stabilize.

We assign stocks into deciles based on expected returns at the end of month t, and compute average returns for stocks in these deciles over the next 15 years. We then average the estimates over all starting

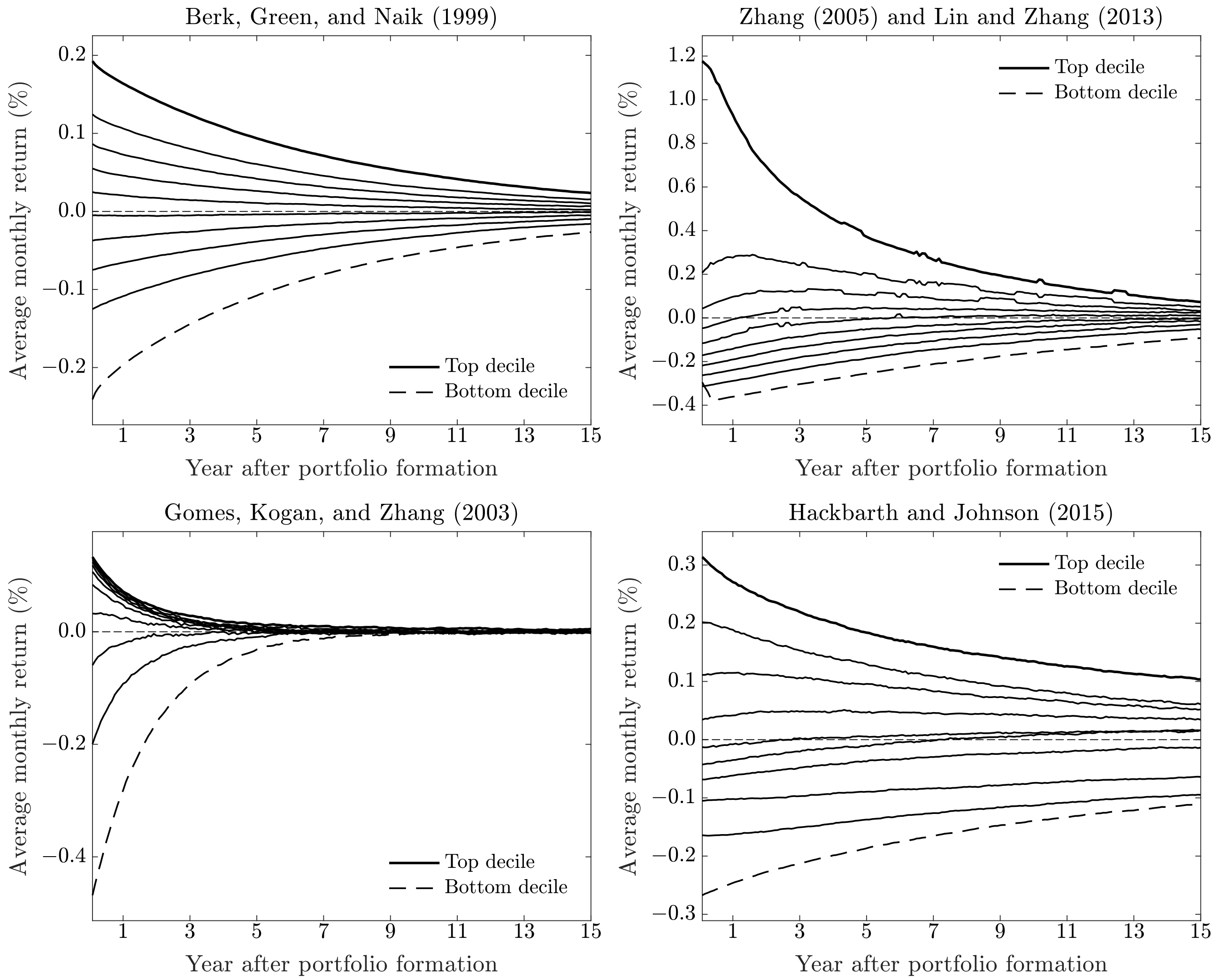

Figure 2: **Average monthly returns on stocks sorted by expected returns.** We simulate 1,000 months of return data from the four models described in Section 2. We run these simulations using the same parameters as those used in the original studies. We discard the first 400 months and then begin ranking stocks into deciles based on expected returns. We report the average cross-sectionally demeaned monthly returns for these deciles over the next 15 years after portfolio formation.

months t. We repeat each simulation 1,000 times to reduce simulation-specific noise; this is different from the graphs in the introduction's Figure 1 in which we plot the data from single runs.

Figure 2 shows average, cross-sectionally demeaned monthly returns from the four models discussed above. The models differ in the amount of dispersion in expected returns. In Berk, Green, and Naik (1999), for example, the difference in expected monthly returns between the top and bottom deciles is just over 40 basis points; in Zhang (2005), this difference is over 140 basis points. In some models, such as Berk, Green, and Naik (1999), the cross-sectional distribution of expected returns is nearly symmetric; in others, such as Gomes, Kogan, and Zhang (2003), it is considerably left-skewed. The common element of these models, however, is the convergence in expected returns. Although the models differ in the speed of convergence—they model different economic mechanisms, and they are parametrized differently—expected returns converge in all of them.[10] Figure A1 in the appendix shows that as expected returns converge toward the mean, so do market betas.

The pattern in Figure 2 is not specific to the models we consider. If risks change but are stationary, they must mean-revert. Moreover, unless a model builds in *permanent* cross-sectional differences in, for example, production technology, each firm's risk (and its expected return) must be expected to converge toward the common mean. We are not aware of any study on production-based asset pricing models that has found the need to build in persistent differences across firms.[11]

3 Data

We use the daily and monthly CRSP return data from January 1963 through December 2018 on stocks listed on the NYSE, Amex, and NASDAQ. We exclude securities other than ordinary common shares. We also exclude financials, which are identified as firms with SIC codes between 6000 and 6999. We use CRSP delisting returns; if a delisting return is missing and the delisting is performance-related, we impute a return of −30% for NYSE and Amex stocks (Shumway 1997) and −55% for Nasdaq stocks

[10]For example, Hackbarth and Johnson (2015, Table 2) impose significantly more firm-level persistence than the other models or what is seen in the data. In their baseline model, the autocorrelation in profitability is 0.97, whereas it is 0.51 in the data.

[11]In the risk-based models, market rationally expects expected returns to converge, even though the changes in discount rates come as surprises. In Berk, Green, and Naik (1999), for example, a firm's expected return changes when the firm's old project (randomly) dies off or it (randomly) encounters a new project it chooses to undertake. Expected returns could also converge deterministically—an asset's expected return could, for example, be 10%, 9%, 8%, . . . over its life—and the market would discount the cash flows given these expected changes.

(Shumway and Warther 1999).

We use balance sheet and income statement information from the annual and quarterly Compustat files to construct various return predictors that have been proposed in the literature. We describe these predictors in Section 6.1.

4 Detecting differences in long-term discount rates: A bootstrap approach

4.1 Cross-sectional regressions

In this section, we estimate regressions that predict the cross section of stock returns with each stock's average past return. For each month t, we estimate a cross-sectional regression

$$r_{it} = a_t + b_t \times \bar{r}_{i,t-k_1,t-k_2} + e_{it}, \tag{1}$$

where $\bar{r}_{i,t-k_1,t-k_2}$ is stock i's average return from month $t-k_2$ to $t-k_1$. If stock returns contain persistent differences in expected returns, the slope estimate from these regressions is proportional to cross-sectional variance of these differences, $\hat{b} \sim \hat{\sigma}_\mu^2$.[12] For example, if expected returns are constant and return innovations IID, then

$$\hat{b}_t = \frac{\mathrm{cov}^{\mathrm{cs}}(r_{it}, \bar{r}_{i,t-k_1,t-k_2})}{\mathrm{var}^{\mathrm{cs}}(\bar{r}_{i,t-k_1,t-k_2})} = \frac{\mathrm{cov}^{\mathrm{cs}}(\mu_i + \varepsilon_{i,t}, \mu_i + \frac{1}{k_2-k_1+1}\sum_{t'=t-k_2}^{t-k_1} \varepsilon_{i,t'})}{\mathrm{var}^{\mathrm{cs}}(\mu_i + \frac{1}{k_2-k_1+1}\sum_{t'=t-k_2}^{t-k_1} \varepsilon_{i,t'})} = \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + \frac{1}{k_2-k_1+1}\hat{\sigma}_\varepsilon^2} \geq 0, \tag{2}$$

[12] Conrad and Kaul (1998, p. 491) suggest that this dispersion-in-means mechanism is responsible for some of the momentum profits: "The repeated purchase of winners from the proceeds of the sale of losers will, on average, be tantamount to the purchase of high-mean securities from the sale of low-mean securities. Consequently, as long as there is some cross-sectional dispersion in the mean returns of the universe of securities, a momentum strategy will be profitable."

where μ_i is stock i's expected return, σ_μ^2 is the cross-sectional variance of expected returns, $\varepsilon_{i,t}$ is the firm-specific return, and σ_ε^2 is the cross-sectional variance of these returns. We estimate the regressions in equation (1) using actual and bootstrapped stock return data. We construct the bootstrapped data to preserve both the distributions of returns and their covariance structure.

4.2 Methodology

We generate each draw of simulated data in five steps:

1. We draw each stock's expected return from a normal distribution with a mean of zero and a standard deviation σ_μ.

2. We cross-sectionally demean month-t returns and divide them by their standard deviation to generate a vector of studentized residuals with unit variance. We denote this $N_t \times 1$ vector of residuals by ϵ_t, where N_t is the number of stocks in month t.

3. We estimate the $N_t \times N_t$ covariance matrix of stock returns, Σ_t, using monthly data from month $t - 30$ to month $t + 30$, that is, five years plus one month. We estimate pairwise covariances; that is, we do not require all stocks to have non-missing returns for the entire estimation period. Because the number of stocks exceeds the length of the time series, Σ_t is singular.

4. We replace the covariance matrix Σ_t with its nearest positive definite matrix using the algorithm of Higham (2002). We denote this positive definite matrix by S_t.

5. We generate month-t stock-specific shocks ε_t by randomizing the elements of ϵ_t and by post-multiplying them by the Cholesky factor of S_t. Stock i's return in month t is then $\mu_i + \varepsilon_{i,t}$.

These simulated returns have appealing properties. First, the resulting data matrix has the same dimensions as the actual return data; that is, it has the same number of months and the same number of stocks each month. Second, the factor structure of returns is the same as that in the actual data and, given the rolling estimates of covariances, this factor structure changes over time as it does in the data. Third, because we extract the shocks from actual stock returns, the distribution of stock returns each month resembles the actual distribution of returns.

We draw 100 random samples of returns using this procedure. Using each simulated dataset, we estimate the cross-sectional regressions and record the average regression slopes and their t-values. We then compare these estimates with those from the actual data, and examine how the estimates change as we increase the amount of variation in expected returns, σ_μ.

4.3 Estimates

Table 1 reports results from cross-sectional regressions that we estimate using both actual and simulated data. Panel A uses data on all stocks; Panel B restricts the sample to all-but-microcaps. All-but-microcaps are stocks with market capitalizations above the 20th percentile of the NYSE distribution.

The estimates in the first column of Table 1 use actual data; the remaining columns simulate data with values of σ_μ ranging from 0% to 1.4%. The average cross-sectional variance of stock returns is 0.0229. Therefore, when $\sigma_\mu = 1\%$, cross-sectional variation in expected returns explains $(0.01)^2/0.0229 = 0.44\%$ of the cross-sectional variation in realized returns.

We estimate the cross-sectional regressions in equation (1) using various past return windows. The first row predicts the cross section of month t returns using month $t-1$ returns; the last row uses prior 20-year returns skipping a month. Each regression contains stocks that have non-missing returns in month t and for at least half of the past return period. The regressions on the last row, for example, include stocks that have at least ten years of non-missing returns.

The estimates that use actual return data show the effects of short-term reversals (Jegadeesh 1990),

Table 1: Fama-MacBeth regressions with actual and bootstrapped data

This table reports average coefficients and t-values from regressions to predict the cross section of monthly stock returns. The explanatory variable is the stock's average return over the window specified in the first column. Column "Actual data" uses returns on common stocks listed on NYSE, Amex, and NASDAQ from 1963 through 2018. The sample excludes financials, which are identified as firms with SIC codes between 6000 and 6999. Columns "Bootstrapped data" use data that are first randomized to set the variation in expected returns to zero. These data are generated in five steps. First, month-t returns are demeaned and divided by their standard deviation to generate a vector of residuals, ϵ_t, with unit variance. Second, the covariance matrix of individual stock returns, Σ_t, is estimated using data from 30 months prior to 30 months after month t. Third, this covariance matrix is transformed to the nearest positive definite matrix S_t using the algorithm of Higham (2002). Fourth, month-t returns are generated by randomizing the elements of ϵ_t and post-multiplying this vector by the Cholesky factor of S_t. Fifth, μ_i is added to stock i's return each month, where μ_is are drawn from a distribution with a mean of zero and a standard deviation of σ_μ. The simulations vary the σ_μ from 0 to 1.4%. We repeat the bootstrapping procedure 100 times for each value of σ_μ; this table reports the average coefficients and t-values across these simulations. Panel A uses data on all stocks; Panel B uses all-but-microcaps. All-but-microcaps are stocks with market capitalizations above the 20th percentile in the NYSE distribution.

Panel A: All stocks

Historical return horizon	Actual data	Bootstrapped data: Cross-sectional variation in expected returns, σ_μ							
		0.0%	0.2%	0.4%	0.6%	0.8%	1.0%	1.2%	1.4%
Fama-MacBeth coefficient estimates									
$[-1, -1]$	−0.05	−0.00	−0.00	0.00	0.00	0.00	0.00	0.01	0.01
$[-12, -2]$	0.04	−0.00	−0.00	0.00	0.01	0.02	0.04	0.06	0.07
$[-60, -13]$	−0.16	−0.00	0.01	0.03	0.07	0.11	0.17	0.22	0.27
$[-120, -61]$	−0.20	−0.01	0.01	0.07	0.15	0.24	0.32	0.40	0.47
$[-120, -2]$	−0.26	−0.00	0.02	0.07	0.15	0.23	0.31	0.39	0.46
$[-240, -121]$	−0.22	−0.01	0.04	0.16	0.29	0.41	0.51	0.59	0.66
$[-240, -2]$	−0.26	−0.01	0.04	0.16	0.29	0.41	0.51	0.59	0.66
t-values									
$[-1, -1]$	−11.80	−0.05	−0.02	0.06	0.20	0.40	0.65	0.95	1.32
$[-12, -2]$	2.66	−0.19	−0.10	0.18	0.64	1.29	2.10	3.08	4.24
$[-60, -13]$	−5.98	−0.06	0.25	1.14	2.58	4.52	6.90	9.66	12.83
$[-120, -61]$	−4.54	−0.24	0.37	2.08	4.73	8.17	12.22	16.95	22.28
$[-120, -2]$	−6.83	−0.12	0.55	2.45	5.37	9.17	13.61	18.77	24.54
$[-240, -121]$	−3.34	−0.30	0.93	4.17	8.85	14.59	21.30	29.19	38.00
$[-240, -2]$	−4.58	−0.24	1.11	4.67	9.79	16.01	23.22	31.65	40.98

Panel B: All-but-microcaps

Historical return horizon	Actual data	Bootstrapped data: Cross-sectional variation in expected returns, σ_μ 0.0%	0.2%	0.4%	0.6%	0.8%	1.0%	1.2%	1.4%
		Fama-MacBeth coefficient estimates							
$[-1, -1]$	-0.02	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02
$[-12, -2]$	0.10	-0.00	0.00	0.01	0.03	0.06	0.08	0.12	0.15
$[-60, -13]$	-0.10	-0.00	0.01	0.06	0.12	0.20	0.28	0.35	0.42
$[-120, -61]$	-0.09	-0.01	0.03	0.13	0.24	0.36	0.46	0.55	0.62
$[-120, -2]$	-0.15	-0.01	0.03	0.12	0.23	0.35	0.45	0.54	0.61
$[-240, -121]$	-0.10	-0.01	0.07	0.24	0.40	0.54	0.64	0.72	0.77
$[-240, -2]$	-0.16	-0.01	0.07	0.23	0.40	0.53	0.63	0.71	0.77
		***t*-values**							
$[-1, -1]$	-3.48	0.00	0.06	0.24	0.53	0.96	1.48	2.13	2.90
$[-12, -2]$	4.98	-0.14	0.05	0.64	1.60	2.96	4.63	6.67	9.02
$[-60, -13]$	-3.50	-0.15	0.42	2.07	4.74	8.33	12.59	17.60	23.23
$[-120, -61]$	-1.60	-0.26	0.77	3.69	8.22	14.17	20.96	28.92	37.80
$[-120, -2]$	-3.41	-0.19	0.90	3.98	8.77	15.01	22.14	30.52	39.71
$[-240, -121]$	-1.38	-0.27	1.59	6.64	14.08	23.37	33.97	45.99	59.47
$[-240, -2]$	-2.47	-0.22	1.77	7.12	14.98	24.69	35.63	48.14	61.93

momentum (Jegadeesh and Titman 1993), and long-term reversals (De Bondt and Thaler 1985). Short-term reversals show up as negative slope coefficients in cross-sectional regressions of month t returns against prior month returns. The estimates in the full and all-but-microcaps samples are -0.05 (t-value $= -11.80$) and -0.02 (t-value $= -3.48$). Momentum registers as positive coefficients in regressions against prior one-year returns skipping a month. The estimates in the full and all-but-microcaps samples are 0.04 and 0.10, and these estimates are associated with t-values of 2.66 and 4.98. Long-term reversals appear as negative slope coefficients when we lag average returns by more than a year. For example, in regressions that predict returns using prior five-year returns skipping a year, the slope coefficients are -0.16 and -0.10, and the t-values are -5.98 and -3.50.

Data in which expected returns are constant over time, vary across firms, and stock-specific innovations are serially and cross-serially uncorrelated cannot produce *negative* regression slopes. The inevitability of positive regression coefficients under these assumptions is best illustrated by considering the Lo and MacKinlay (1990) decomposition of trading profits.[13] They consider a strategy that weighs stock i by

$$w_{i,t} = \frac{1}{N}\left(\bar{r}_{i,t-k_1,t-k_2} - \bar{r}_{m,t-k_1,t-k_2}\right), \tag{3}$$

where $\bar{r}_{i,t-k_1,t-k_2}$ is stock i's average return from month $t-k_2$ to $t-k_1$, and $\bar{r}_{m,t-k_1,t-k_2}$ is the return on the equal-weighted market index. The expected trading profit in month t, $\mathrm{E}(\pi_t) = \mathrm{E}\left(\sum_{i=1}^{N} w_{i,t} r_{i,t}\right)$, then decomposes into

$$\mathrm{E}(\pi_t) = \underbrace{\frac{N-1}{N^2}\mathrm{tr}(\Sigma)}_{\text{Autocovariances}} - \underbrace{\frac{1}{N^2}\left[1'\Sigma 1 - \mathrm{tr}(\Sigma)\right]}_{\substack{\text{Cross-serial}\\ \text{covariances}}} + \underbrace{\sigma_\mu^2}_{\substack{\text{Persistent}\\ \text{variation}\\ \text{in means}}}, \tag{4}$$

where $\mathrm{tr}(\Sigma)$ is the trace of the covariance matrix, $1'\Sigma 1 - \mathrm{tr}(\Sigma)$ is the sum of the cross-serial covariances, and σ_μ^2 is the cross-sectional variance of unconditional expected returns. The last term is always positive; persistent cross-sectional variation in mean returns increases the profitability of trading strategies in which weights increase in realized returns. Profits can therefore be negative only if stock returns are negatively autocorrelated (the first term of equation (4) is negative) or positively cross-serially correlated (the second term of equation (4) after the minus sign is positive). That is, a negative slope coefficient emerges only if a high return on an asset predicts a low return on that asset or high returns on other assets.

The estimated regression slope in equation (2), which assumes IID innovations, is nonnegative; the IID assumption sets the first two terms of the decomposition in equation (4) to zero. Therefore, in Table 1's bootstrapped data, the slope coefficients start at zero when $\sigma_\mu = 0$ and become increasingly

[13]Lo and MacKinlay (1990) decompose profits to strategies that trade long-term reversals. Lewellen (2002) considers profits to strategies that trade the momentum in individual stocks or portfolios of stocks.

positive as we increase the amount of cross-sectional variation in expected returns. When $\sigma_\mu = 0.4\%$, average past returns over various horizons begin to be identified as statistically significant predictors of returns in the full sample. For the $[-240, -2]$ window in Panel A, for example, the average slope coefficient is 0.16 and the t-value associated with this estimate is 4.67. In the all-but-microcaps sample, the estimates are even more statistically significant when $\sigma_\mu = 0.4\%$ or greater.

The simulations in Table 1 show that if even a small amount of persistent variation was present in expected returns, our regressions would have the power to detect it. A volatility parameter of $\sigma_\mu = 0.4\%$, for example, corresponds to a world in which a cross-sectional regression of *realized* month-t returns r_{it} against expected returns μ_i has an R^2 of 0.07%. Moreover, if expected returns remained constant in the cross section, past returns should become more informative about returns as we increase the length of the past-return window; a wider window yields more precise estimates of expected returns. In the simulations, the t-values indeed increase in the length of the estimation window. In the data, however, the slope coefficients are and remain negative after the one-year momentum period.[14]

These comparisons suggest the amount of persistent cross-sectional variation in expected returns must be negligible; whatever the variation might be, it is completely overshadowed by long-term reversals in individual stock returns. The term *persistent* can also be interpreted loosely here. If stocks' expected returns change over time, but slowly, a regression against five-year average returns would still typically return a positive coefficient. In the data, the cutoff is one year. It is instructive to consider momentum to understand the amount of predictive power we would expect to find at longer lags if cross-sectional differences in expected returns persisted. In Panel B's all-but-microcaps sample, the slope coefficient on momentum is 0.10 with a t-value of 4.98. In the bootstrapped data, we approximately match this

[14]Jegadeesh and Titman (2001) also note that returns on momentum-sorted portfolios turn negative after 12 months, and that these negative estimates are inconsistent with Conrad and Kaul's (1998) explanation that persistent differences in expected returns drive momentum profits.

coefficient when $\sigma_\mu = 1.0\%$. Here, the slope coefficient is 0.08 with a t-value of 4.63. If we were to attribute momentum to persistent cross-sectional variation in mean returns, long-horizon past returns would be tremendously powerful predictors of the cross section of stock returns. For example, in regressions against the prior ten-year returns skipping a month, the t-value would be 22.14.

5 Production-based models, changes in discount rates, and long-term reversals

5.1 Intuition

The Fama-MacBeth regressions in Section 4 show that past returns predict the cross section of average returns with negative signs. These negative associations correspond to the long-term reversals identified by De Bondt and Thaler (1985). The negative sign must be due either to negative autocorrelations or positive cross-serial correlations in stock returns (Lo and MacKinlay 1990). The lack of persistent differences in expected returns is a potential source of negative autocorrelations. If differences in expected returns are short-lived, some of the stock return variation is, by definition, due to changes in discount rates. Long-term reversals could emerge if there are large but short-lived differences in expected returns; in that case, past returns predict future returns with a negative sign because past returns negatively correlate with changes in discount rates.

Figure 3 gives a textbook illustration of the association between past and expected returns in the presence of discount rate shocks.[15] We assume all firms pay just one cash flow to shareholders at time $t = 5$ and the expected value of this terminal cash flow is $\mathrm{E}(\widetilde{\mathrm{CF}}_1) = 100$ at time $t = 0$. A firm's market value at any date prior to the terminal date is the present value of the expected cash flow. In Panel A,

[15] Cochrane (2011) provides a detailed discussion and analysis of the variation in discount rates at the market level.

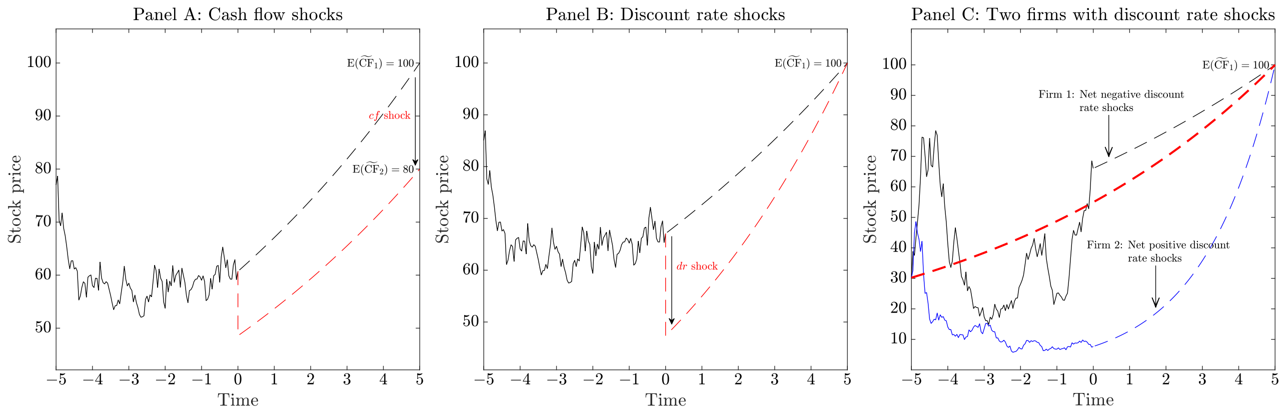

Figure 3: **Cash flow shocks, discount rate shocks, and the information content of past returns.** This figure illustrates the association between past and expected returns when return variation is due to cash flow shocks (Panel A) or discount rate shocks (Panel B). When return variation is due to cash flows shocks, expected returns are unrelated to past returns. When return variation is due to discount rate shocks, the relation between past and expected returns is negative. In Panel C two firms have identical discount rates at time $t = -5$. The differences in these firms' discount rates at time $t = 0$ correlate perfectly with the differences in past returns because their return variation is only due to discount rate shocks.

all return variation is due to cash flow shocks. Because the discount rate is constant, differences in past returns are uninformative about differences in expected returns. If a firm's market value decreases significantly—as it does at time $t = 0$—it must do so because the expected cash flow decreased. In Panel B all return variation is due to discount rate shocks. If a firm's market value decreases significantly, it must be because the firm's expected return increased. In this case, past returns correlate negatively with expected returns. A negative realized return signifies an increase in the expected return.

In Panel C, we endow two firms the same expected return at time $t = -5$ and randomly shock these firms' discount rates until time $t = 0$. Because one firm is hit with net-positive discount rate shocks and the other with net-negative discount rate shocks, one firm's expected return is higher than that of the other at time $t = 0$. Moreover, because the firms started with identical expected rates at time $t = -5$, the differences in expected returns at time 0 correlate perfectly with differences in past returns.[16] Past returns' informativeness about expected returns depends on the intensity of discount rate shocks relative to cash flow shocks. If firms' expected returns rapidly converge toward a common mean, past returns may mostly measure discount rate shocks. Here, past returns no longer measure differences in the *levels* of discount rates.[17]

The relative importance of discount-rate and cash-flow news in driving stock returns does not matter for this argument. Vuolteenaho (2002), for example, attributes one-third of the variation in unexpected firm-level stock returns to discount-rate news and two-thirds to cash-flow news. This result does not contradict ours. The question "how much of the variation in stock returns emanates from discount-rate news?" is distinct from the one we ask, "how persistent are differences in discount rates?" Our results suggest that the movements in discount rates are not persistent; it does not matter how common or uncommon such movements are relative to the return shocks emanating from cash-flow news.

5.2 Evidence from production-based models

In Table 2 we report estimates from cross-sectional regressions similar to those in Table 1. However, instead of using the actual or bootstrapped data, we estimate these regressions using data simulated from the four models described in Section 2. We simulate data using each model's baseline parameters and generate hypothetical datasets that match the length and size of the cross section of the actual return data. With the Berk, Green, and Naik (1999) model, we also alter the model's depreciation rate

[16]The log return on firm i from $t-5$ to t is $\ln(P_{i,t}/P_{t-5}) = p_{i,t} - p_{t-5}$. The expected log return from t to $t+5$ is $\ln(\mathrm{E}(\widetilde{\mathrm{CF}}_1)/P_{i,t}) = \ln(\mathrm{E}(\widetilde{\mathrm{CF}}_1)) - p_{i,t}$. Because p_{t-5} and $\ln(\mathrm{E}(\widetilde{\mathrm{CF}}_1))$ are common across firms, the past and expected returns encode the same information and correlate perfectly in the time-series (within a firm) and the cross section (across multiple firms).

[17]Pástor and Stambaugh (2009) describe the association between past and future *market* returns as follows: "Suppose that recent returns have been unusually low. On the one hand, one might think that the expected return has declined, since a low mean is more likely to generate low realized returns, and the conditional mean is likely to be persistent. On the other hand, one might think that the expected return has increased, since increases in expected future returns tend to be accompanied by low realized returns. When ρ_{uw} [the correlation between realized returns and the changes in expected returns] is sufficiently negative, the latter effect outweighs the former and recent returns enter negatively when estimating the current expected return." They find that, in the data, high (low) recent market returns imply that expected returns are low (high).

Table 2: Long-term reversals in production-based asset pricing models

This table reports average coefficients and t-values from regressions to predict the cross section of monthly stock returns. The explanatory variable is the stock's average return over the window specified in the first column. We estimate these regressions using data simulated from the models of Berk, Green, and Naik (1999), Gomes, Kogan, and Zhang (2003), Lin and Zhang (2013), and Hackbarth and Johnson (2015). Each simulated sample matches the dimensions of the actual stock data in Table 1. We use each model's baseline parameters. With the Berk et al. (1999) model, we also vary the depreciation rate, $1-\pi$, from the 1% rate of the baseline model. We repeat the simulations 10,000 times for each specification and report the average estimates and t-values.

	Model						
Historical return horizon	Berk, Green, and Naik — Depreciation rate, $1-\pi$: 0.5%	1.0%	2.0%	3.0%	Gomes, Kogan, and Zhang	Lin and Zhang	Hackbarth and Johnson
	Coefficient estimates						
$[-1,-1]$	−0.00	−0.00	−0.00	−0.01	−0.01	−0.01	0.00
$[-12,-2]$	−0.00	0.00	−0.03	−0.06	−0.07	−0.05	0.02
$[-60,-13]$	0.01	0.01	−0.08	−0.10	−0.10	−0.13	0.05
$[-120,-61]$	0.01	0.01	−0.03	−0.02	−0.01	−0.09	0.04
$[-120,-2]$	0.02	0.02	−0.16	−0.19	−0.21	−0.31	0.10
$[-240,-121]$	0.01	0.01	−0.01	−0.00	−0.00	−0.09	0.05
$[-240,-2]$	0.03	0.03	−0.18	−0.19	−0.22	−0.49	0.14
	t-values						
$[-1,-1]$	−0.64	−0.39	−3.25	−4.64	−9.11	−2.96	2.71
$[-12,-2]$	0.10	0.53	−4.89	−6.79	−17.27	−3.20	5.77
$[-60,-13]$	1.29	0.97	−4.77	−6.50	−14.25	−3.29	6.82
$[-120,-61]$	1.21	1.05	−3.55	−3.47	−2.45	−3.22	6.14
$[-120,-2]$	1.94	1.07	−4.85	−6.48	−16.28	−3.39	7.32
$[-240,-121]$	1.13	0.85	−1.29	−0.51	−0.22	−2.95	5.70
$[-240,-2]$	2.08	1.13	−4.73	−6.21	−14.16	−3.36	7.36

and run additional simulations. We estimate cross-sectional regressions using each simulated sample and record the average slope estimates and their t-values. Table 2 reports averages of the estimates and t-values over 10,000 simulations for each specification.

The first four columns in Table 2 use data simulated from Berk et al. (1999). The original parametrization of the model uses a depreciation rate of $1-\pi = 1\%$. Berk et al. (1999, p. 1573)

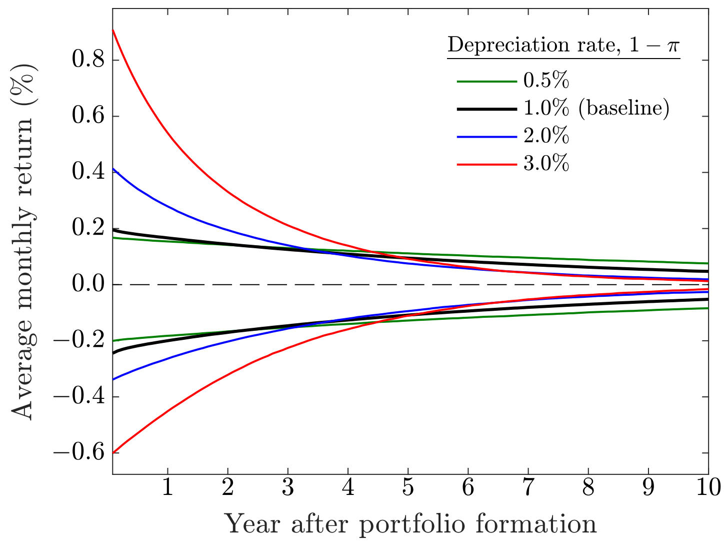

Figure 4: **Average monthly returns in the Berk, Green, and Naik (1999) model with different depreciation rates.** We simulate 1,000 months of return data from the model of Berk et al. (1999) using four alternative values for the depreciation rate, $1 - \pi$. All other parameters of the model are set to the values used in the original study. We discard the first 400 months from each simulation and then begin ranking stocks into deciles based on expected returns. We compute the average cross-sectionally demeaned monthly returns for these deciles over the next 10 years after portfolio formation. This figure averages over 10,000 simulations and reports the average monthly returns for the top and bottom deciles.

determine this rate based on the calibrations of real business cycle models (Kydland and Prescott 1982; Christiano and Eichenbaum 1992). In the model, the depreciation rate accounts for the probability that a firm's project becomes obsolete and ceases to generate cash flows. We vary this rate because it affects the rate of convergence of expected returns and, through the discount-rate mechanism, the relation between past and future returns.

Figure 4 shows how the convergence of expected returns in the Berk et al. (1999) model depends on the depreciation rate. When the depreciation rate increases, firms' risks converge faster toward the common mean; when we decrease the rate below that in the baseline model, expected returns become more persistent. An increase in the depreciation rate therefore increases the amount of return variation that is due to changes in discount rates.

Table 2 shows that, in the baseline specification of Berk et al. (1999), past returns do not significantly predict the cross section of average returns. However, outside of this Goldilocks point, the association is statistically significant. If we lower the depreciation rate, past returns positively predict future returns. In this case, discount rates rarely change, and past returns mostly measure differences in (relatively) persistent expected returns. If we increase the depreciation rate, past returns negatively predict future returns. This parametrization is close to the world of Panel C of Figure 3 in which all return variation is due to changes in discount rates. Even when the depreciation rate is very high, the association between past returns and future returns is weak only when the historical returns are over ten years old. What happened between 10 and 20 years ago is uninformative about a firm's expected return today because these returns have already been followed by another 10 years of changes in discount rates.

The estimates from the other three models in Table 2 suggest the strength (or lack thereof) of long-term reversals depends on the rate at which the discount rates converge. In Figure 2 the rate of convergence varies across the three models. The rate of convergence is the fastest in Gomes et al. (2003); it is the second fastest in Lin and Zhang (2013); and, in Hackbarth and Johnson (2015), discount rates are very persistent. Consistent with this ordering, the estimates in Table 2 show the strongest long-term reversals for Gomes et al. (2003), followed by Lin and Zhang (2013). In Hackbarth and Johnson (2015), expected returns are so persistent that past returns positively predict future returns.

Figure 3 and Table 2 provide a compelling explanation for the negative slope estimates found in the cross-sectional returns-on-past-returns regressions. These estimates would be positive if cross-sectional differences in expected returns were very persistent. The apparent reason for the negative estimates is that expected returns are not very persistent. Expected returns vary dramatically across firms (Martin and Wagner 2019), but these differences, on average, converge rapidly to zero. It is this convergence that imprints the negative serial correlation into the data. Long-term reversals are therefore the consequence of the lack of differences in long-term discount rates.

6 Portfolio sorts

6.1 Return predictors

We complement the bootstrap analysis by sorting stocks into portfolios using 55 return predictors. We form portfolios using these predictors and then measure the persistence in their average returns and alphas. The predictors we include are among those analyzed in McLean and Pontiff (2016). We also add a few additional predictors from Linnainmaa and Roberts (2018) that were published after McLean and Pontiff (2016) created their list. We group predictors into (1) fundamental, (2) event, (3) market, and (4) valuation subgroups using the McLean and Pontiff (2016, p. 19) classification scheme. We recompute and rebalance all factors monthly.[18] We assume that firms' accounting numbers are available six months after the end of the fiscal year. We compute the two factors that relate to earnings announcements, standardized unexpected earnings (SUE) and three-day cumulative abnormal return around earnings announcement (CAR3), using the quarterly earnings dates from Compustat.

Table 3 reports monthly average month $t+1$ returns and CAPM and three-factor model alphas for HML-style factors based on each of the 55 return predictors. We sort stocks into six portfolios by market capitalization and the predictor, and then compute value-weighted returns on these portfolios. We sign each predictor so that high values identify the stocks that the initial study identified as earning higher returns. The breakpoint for size is the 50th NYSE percentile, and the breakpoints for the predictor are the 30th and 70th NYSE percentiles. A factor's return is the average return on the two high portfolios

[18] We maximize the timeliness of the signals by using monthly rebalancing. That is, the lag between any future period and the signal is always the smallest possible. Despite monthly rebalancing, any signal that uses accounting information typically updates only once a year. Gross profitability, for example, is defined as gross profits divided by total assets from the fiscal year that ended at least six months earlier; this measure therefore updates only once a year except when a firm changes its fiscal year.

minus the average return on the two low portfolios.[19] Table 3 shows that most of the 55 return predictors that we examine explain differences in average returns over the 1963 through 2018 sample period: 39 factors' average returns are positive and statistically significant at the 5% level; for the CAPM and three-factor model alphas, the counts are 45 and 40, respectively.

The differences between the CAPM and three-factor model alphas are consistent with the literature. Long-term reversals, for example, has a CAPM alpha of 29 basis points per month (t-value = 3.01) but a three-factor model alpha of 7 basis points (t-value = 0.86). Fama and French (1996) note that the addition of the size and value factors removes this alpha because "long-term past losers load more on SMB and HML," that is, they are predominantly small value stocks. The three-factor model alphas for "firm size" and "book-to-market" are not zero in Table 3 because, first, we rebalance the factors monthly (instead of annually) and because Fama and French use a proprietary mapping file to link CRSP and Compustat (instead of using the link supplied by CRSP).

6.2 Average long-term returns

In Figure 5 and Table 4, we form decile portfolios using the 55 return predictors and examine differences in value-weighted average returns up to ten years after portfolio formation. In these computations, the portfolios are formed at date t and then held unchanged for up to ten years. In addition to using all 55 returns predictors, in Table 4 we also form portfolios using four subsets of predictors: (1) fundamental, (2) event, (3) market, and (4) valuation.

We convert all predictors into percentile ranks to make them comparable. We first take NYSE-listed stocks and compute cross-sectional percentile ranks of each predictor. We then place all non-NYSE stocks onto these distributions with linear interpolation. By using NYSE stocks as the reference

[19] The size factor is the only exception to this rule. This factor is long the three small portfolios and short the three big portfolios when sorting stocks into portfolios by size and book-to-market.

Table 3: Average returns and CAPM and three-factor model alphas of 55 return predictors

This table reports average returns and CAPM and three-factor model alphas for 55 return predictors. We construct HML-style factors by first sorting stocks into six portfolios by size and predictor. These sorts are independent and use NYSE breakpoints; the size breakpoint is the median and the predictor breakpoints are the 30th and 70th percentiles. A predictor's return is the average return on the two value-weighted high portfolios minus the average return on the two value-weighted low portfolios. The high and low portfolios are determined so that "high" corresponds to those stocks that the initial study identified as earning higher returns. We rebalance all factors monthly.

		Average return		CAPM alpha		FF3 alpha	
Predictor	Start year	$\bar{r}$	$t(\bar{r})$	$\hat{\alpha}$	$t(\hat{\alpha})$	$\hat{\alpha}$	$t(\hat{\alpha})$
Fundamental							
Gross profitability	1963	0.36	4.13	0.35	3.90	0.47	5.67
Piotroski's F-score	1963	0.32	4.02	0.42	5.71	0.40	5.90
Abnormal investment	1963	0.23	4.31	0.22	4.06	0.23	4.53
Leverage	1963	0.15	1.25	0.21	1.71	−0.15	−2.03
Accruals	1963	0.27	4.34	0.29	4.75	0.25	4.15
Net operating assets	1963	0.39	6.26	0.37	5.98	0.43	7.17
O-score	1963	0.08	1.09	0.15	2.09	0.22	3.34
Profit margin	1963	0.03	0.36	0.15	1.75	0.21	3.01
Sales growth	1963	0.05	0.58	0.16	1.97	−0.04	−0.65
Asset growth	1963	0.31	3.48	0.41	4.87	0.18	3.13
Z-score	1964	0.05	0.50	−0.02	−0.19	0.21	3.00
Investment-to-capital	1963	0.24	2.22	0.41	4.32	0.17	2.59
Investment-to-assets	1963	0.35	5.12	0.41	6.08	0.28	4.89
QMJ profitability	1963	0.36	4.40	0.43	5.30	0.54	7.51
Distress	1963	0.33	2.34	0.57	4.76	0.56	5.71
Operating profitability	1963	0.29	2.89	0.35	3.50	0.36	4.06
Operating leverage	1963	−0.01	−0.15	−0.09	−1.10	−0.06	−0.74
Tax-to-price	1970	0.41	2.36	0.27	1.54	0.22	1.21
Cash-based op. profitability	1963	0.56	8.65	0.61	9.59	0.69	12.10
Return on assets	1963	0.23	2.41	0.31	3.30	0.39	4.47
Return on equity	1963	0.11	1.41	0.15	1.84	0.24	3.21
Asset turnover	1963	0.24	4.47	0.26	4.95	0.23	4.37

Predictor	Start year	Average return $\bar{r}$	Average return $t(\bar{r})$	CAPM alpha $\hat{\alpha}$	CAPM alpha $t(\hat{\alpha})$	FF3 alpha $\hat{\alpha}$	FF3 alpha $t(\hat{\alpha})$
Event							
Inventory growth	1963	0.29	4.20	0.35	5.37	0.29	4.57
Net working capital changes	1963	0.26	4.53	0.29	5.03	0.28	5.08
Share issuance, 1 year	1963	0.24	3.58	0.32	5.03	0.22	4.20
Share issuance, 5 years	1963	0.24	3.70	0.31	5.11	0.23	4.35
Sustainable growth	1963	0.18	2.18	0.27	3.47	0.08	1.25
Total external financing	1972	0.38	3.67	0.53	5.79	0.42	5.94
Ind.-adjusted CAPX growth	1963	0.17	3.58	0.21	4.55	0.16	3.50
Sales-minus-inv. growth	1963	0.26	5.38	0.26	5.44	0.28	5.77
Inventory growth rate	1963	0.24	3.93	0.29	5.07	0.19	3.62
Earnings surprise (SUE)	1975	0.52	6.84	0.56	7.98	0.60	8.72
Earnings surprise (CAR3)	1975	0.44	7.00	0.46	7.69	0.48	7.83
Market							
52-week high	1963	0.52	3.21	0.76	5.30	0.82	5.90
Amihud's illiquidity	1963	0.37	1.85	0.48	2.49	0.31	1.72
Market beta	1963	0.04	0.25	0.41	3.28	0.33	3.25
Idiosyncratic volatility	1963	0.22	1.20	0.58	3.99	0.49	4.82
Industry momentum	1963	0.56	4.23	0.58	4.36	0.61	4.57
Long-term reversals	1963	0.28	2.91	0.29	3.01	0.07	0.86
Maximum daily return	1963	0.30	1.92	0.60	4.78	0.50	5.30
Momentum	1963	0.65	4.04	0.71	4.50	0.84	5.36
Intermediate momentum	1963	0.59	4.91	0.58	4.85	0.69	5.85
Nominal stock price	1963	−0.07	−0.44	−0.23	−1.77	−0.47	−3.81
Short-term reversals	1963	0.46	3.85	0.36	3.11	0.32	2.68
High volume premium	1963	0.50	8.02	0.53	8.74	0.49	8.16
Share volume	1963	−0.03	−0.18	−0.29	−2.44	−0.19	−1.85
Coskewness	1963	0.13	1.80	0.14	1.95	0.09	1.28
Firm size	1963	0.20	1.60	0.06	0.49	−0.07	−2.06
Valuation							
Earnings-to-price	1963	0.29	2.15	0.45	3.73	0.17	2.03
Enterprise multiple	1963	0.41	3.75	0.48	4.63	0.21	2.68
Sales-to-price	1963	0.34	2.79	0.36	3.01	−0.01	−0.11
Book-to-market	1963	0.25	2.14	0.35	3.11	−0.03	−0.60
Advertising-to-price	1973	0.19	1.41	0.27	2.02	−0.04	−0.45
R&D-to-price	1973	0.39	3.08	0.29	2.38	0.29	2.39
Cash flow-to-price	1963	0.32	2.46	0.46	3.77	0.11	1.57

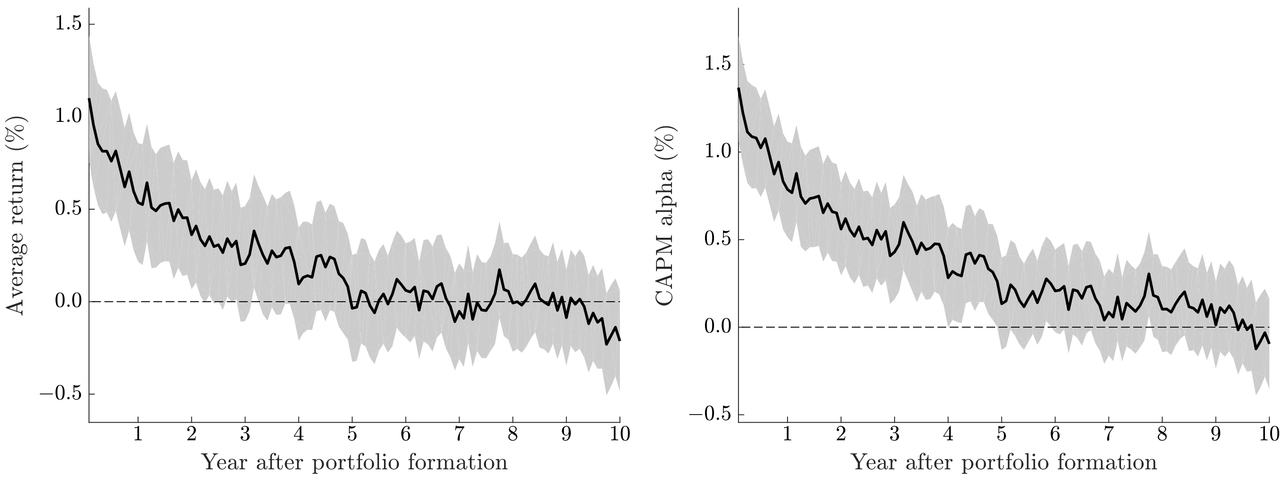

Figure 5: **Average monthly returns and CAPM alphas on long-short strategies in months 1–120 after portfolio formation.** We sort stocks into portfolios by a combination of the 55 return predictors listed in Table 3. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We construct decile portfolios each month and compute value-weighted returns for these portfolios for up to ten years. We plot the average return difference between the top and bottom deciles (left panel) and the monthly CAPM alphas for the return difference between the top and bottom deciles (right panel). The average return in month t is only the month t return, not the average return from today to month t. The shaded areas indicate 95% confidence intervals.

distribution, this percentile-rank method is a generalization of the common methodology of sorting stocks into portfolios using NYSE breakpoints. We sign predictors so that high values correspond to high average returns based on the original study. A stock's combined signal is the average of its non-missing percentile ranks. There are, for example, 11 "event" predictors (see Table 3); if a firm has non-missing values for all of them, its "event" signal is the average percentile rank of these 11 predictors.

In Figure 5 we take the difference between the top and bottom deciles and report the average value-weighted monthly returns (left panel) and CAPM alphas (right panel) up to 120 months after portfolio formation. These returns are non-cumulative. That is, the return at horizon k is the average return that an investor would have earned in month k after portfolio formation. Table 4 reports average returns, CAPM alphas, and the associated t-values for the month following portfolio formation and for years 1, 2,. . . , 10 following portfolio formation. Because we rebalance the portfolios monthly, we adjust standard

Table 4: Annual post-formation spread portfolio returns

We sort stocks into decile portfolios using 55 return predictors ("all") or four predictor subsets: (1) fundamental, (2) event, (3) market, and (4) valuation. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We form value-weighted portfolios each month and hold these portfolios for up to ten years. We report average monthly returns and CAPM alphas for high-minus-low portfolios for different horizons following portfolio formation. The holding periods are nonoverlapping: year 3, for example, is the average return over just the third year after portfolio formation. We adjust standard errors for overlapping returns using the Jegadeesh and Titman (1993) method. Estimates that are statistically significantly different from zero at the FDR $\leq$ 5% level with the Benjamini-Hochberg-Yekutieli correction for multiple-hypothesis testing are denoted with the + sign.

		Subset of predictors					Subset of predictors			
Horizon	All	(1)	(2)	(3)	(4)	All	(1)	(2)	(3)	(4)
	Average monthly returns (%)					Monthly CAPM alphas (%)				
1 month	1.10^{+}	0.67^{+}	0.79^{+}	1.18^{+}	0.41	1.37^{+}	0.91^{+}	0.98^{+}	1.35^{+}	0.44^{+}
Year 1	0.77^{+}	0.48^{+}	0.52^{+}	0.46^{+}	0.44^{+}	1.03^{+}	0.71^{+}	0.69^{+}	0.59^{+}	0.49^{+}
Year 2	0.50^{+}	0.27	0.35^{+}	0.03	0.42^{+}	0.72^{+}	0.48^{+}	0.51^{+}	0.11	0.45^{+}
Year 3	0.31	0.13	0.11	0.03	0.33	0.52^{+}	0.32^{+}	0.29	0.11	0.35
Year 4	0.25	0.15	0.09	0.14	0.20	0.45^{+}	0.33^{+}	0.24	0.21	0.22
Year 5	0.15	0.15	0.03	0.03	0.13	0.33^{+}	0.32^{+}	0.20	0.06	0.15
Year 6	0.03	0.10	0.04	0.02	0.04	0.19	0.26	0.20	0.05	0.05
Year 7	0.01	0.07	0.07	0.06	0.03	0.15	0.21	0.20	0.07	0.03
Year 8	0.01	−0.03	−0.02	−0.10	0.18	0.15	0.11	0.11	−0.10	0.16
Year 9	0.01	−0.08	−0.06	−0.17	0.08	0.12	0.06	0.06	−0.17	0.05
Year 10	−0.10	−0.11	−0.07	0.02	−0.11	0.00	0.02	0.05	0.04	−0.14
	t-values					t-values				
1 month	6.33^{+}	4.08^{+}	5.63^{+}	7.46^{+}	2.32	8.95^{+}	6.20^{+}	7.52^{+}	8.73^{+}	2.51^{+}
Year 1	4.90^{+}	3.10^{+}	4.29^{+}	3.70^{+}	2.63^{+}	7.70^{+}	5.06^{+}	6.39^{+}	5.02^{+}	2.92^{+}
Year 2	3.36^{+}	1.79	2.74^{+}	0.24	2.69^{+}	5.53^{+}	3.58^{+}	4.40^{+}	0.95	2.88^{+}
Year 3	2.10	0.83	0.86	0.29	2.11	4.13^{+}	2.34^{+}	2.35	1.05	2.21
Year 4	1.73	1.03	0.66	1.28	1.26	3.51^{+}	2.51^{+}	1.92	1.90	1.39
Year 5	1.12	1.04	0.21	0.26	0.85	2.65^{+}	2.49^{+}	1.52	0.50	0.94
Year 6	0.23	0.73	0.28	0.16	0.28	1.73	2.03	1.51	0.41	0.29
Year 7	0.05	0.56	0.57	0.61	0.22	1.30	1.72	1.67	0.68	0.20
Year 8	0.12	−0.26	−0.18	−0.92	1.10	1.44	0.89	0.97	−0.97	1.02
Year 9	0.13	−0.52	−0.52	−1.55	0.54	1.14	0.44	0.51	−1.55	0.31
Year 10	−0.87	−0.80	−0.54	0.18	−0.67	0.01	0.15	0.41	0.40	−0.91

errors for overlapping observations using the Jegadeesh and Titman (1993) procedure.

Figure 5 and Table 4 suggest that little, if any, persistent differences exist in average returns. Using the full set of 55 predictors, the average return is 110 basis points (t-value $= 6.33$) one month after portfolio formation. These returns decline monotonically in the horizon. In years 1 through 5, the returns are 77, 50, 31, 25, and 15 basis points per month, respectively. Because the precision of the estimates is approximately constant—these are non-cumulative returns—t-values decline hand in hand with the point estimates, falling from 4.90 in year one to 1.12 by year five. Figure 5 shows that this full set of predictors is not informative about long-term differences in average returns.

Differences in CAPM alphas persist longer than those in average returns. Figure 5 and Table 4 show statistically significant differences in CAPM alphas for up five or six years after portfolio formation. Market adjustment helps both by increasing point estimates and, by virtue of removing market-wide variation in returns, by lowering standard errors. When we use all 55 return predictors, the CAPM alpha in the first year after portfolio formation is 103 basis points with a t-value of 7.70. The fact that CAPM alphas are higher than average returns is consistent with Table 3's result that most anomalies are stronger on a market risk-adjusted basis.

Because we test 11 hypotheses in each column, we adjust critical p-value thresholds for the multiple comparisons problem using the Benjamini-Hochberg-Yekutieli method.[20] This method orders p-values from the smallest to largest and finds the largest p-value below the critical threshold $p_{(k)} < \frac{k}{m}\alpha$, where k is the kth smallest p-value, α is the desired false discovery rate (here, $\alpha = 0.05$), and m is the total number of tests (here, $m = 11$). If, for example, the fifth smallest p-value lies below the threshold, we reject all hypotheses up to and including the fifth. The Benjamini-Hochberg-Yekutieli method ensures that the expected proportion of false discoveries (Type I errors) is no greater than α. The

[20]See, for example, Harvey, Liu, and Zhu (2016) and Chordia, Goyal, and Saretto (2020) for discussions of the multiple comparisons problem in finance applications and an overview of different methods for addressing this problem.

$p_{(k)} < \frac{k}{m}\alpha$ critical threshold assumes that the tests are either independent or positively correlated. In Table 4 we mark estimates that are statistically significantly different from zero when controlling for the false discovery rate at the 5% level using the Benjamini-Hochberg-Yekutieli method with the + sign. Adjusted for the multiple comparisons problem, Table 4 suggests that statistically significant differences in average returns persist for up to two years after portfolio formation; statistically significant differences in CAPM alphas persist for up to five years.[21]

Some predictors are more informative about long-term returns than others. Table 4 shows that "market" predictors—which only use price, return, or volume information—are highly informative about one-month and year 1 returns, but become economically and statistically insignificant predictors already by year 2.[22] Predictors in the valuation subset, by contrast, predict alphas up to year 2 under the Benjamini-Hochberg-Yekutieli method. Nevertheless, the combined set of all predictors is always at least as predictive of long-term returns as any of these subsets. That is, although some of the predictors are short-lived, they do not reduce the full set's power to predict long-term returns.

Figure 5 shows that our inability to detect differences in future average returns is unlikely due to lack of power. If a lack of power were the issue, we would expect the standard errors to increase as a function of time from portfolio formation. Yet they do not: in Figure 5, the width of the confidence interval remains approximately unchanged while the point estimates converge toward zero in all specifications. In fact, we can estimate differences in average returns and CAPM alphas with enough precision to bound the amount of variation in long-term average returns to significant extent. Consider, for example, the CAPM alphas for portfolios formed using all predictors. The 95% confidence interval for year one returns runs from 77 basis points to 130 basis points. In year ten, this interval runs from −21 basis

[21]If we define "persistence" as the last holding period after which the average return has a p-value below 0.05 for three consecutive months, many predictors, such as short-term reversals and idiosyncratic volatility, persist for just one month. The two longest-lasting predictors—quality-minus-junk profitability at 59 months and cash-based operating profitability at 61 months—lose their predictive powers around the five-year mark.

[22]Figure A4 in the appendix shows average returns and CAPM alphas for the same four predictor subgroups.

points to 22 basis points. Therefore, strong statistical evidence suggests differences in expected returns change as a function of time from portfolio formation.

Our results on the short-livedness of cross-sectional differences in expected returns are not specific to any one part of our sample. Figure A2 in the appendix divides the sample period into two halves and shows our results are similar before and after 1990. Concerns related to firm survival also cannot plausibly explain our results. Much of the convergence in expected returns happens already within the first year, when any expected return effects due to a change in the set of firms must be small. Our results are also highly similar in a sample restricted to firms that will survive for ten years (Figure A3) and in the unrestricted sample (Figure 5). Table A1 in the appendix shows that while differences in *equal-weighted* portfolio returns are more persistent than those in value-weighted returns, even these differences last for approximately eight years.[23]

In principle, this section's evidence of the lack of persistence in average returns could be specific to the chosen predictors, but such a scenario is unlikely. None of the predictors in Table 3, when taken in isolation, contain information about differences in average returns beyond year five. We would expect to find some degree of spurious persistence just by luck when testing 55 return predictors; the fact that even the most persistent predictor is not *that* persistent when sidestepping the multiple comparisons issue suggests that long-term differences in average returns must be small or nonexistent. Moreover, we are not aware of any other predictors outside this list that predict differences in average returns farther out than five years. In fact, Section 4's bootstrap procedure suggests the search for such predictors should be futile. If persistent differences existed in expected returns, the Fama-MacBeth regression slopes would be positive. In the data, they are not.

[23]Some of the *alphas* in Table 3 are more persistent—such as the three-factor model alpha of the five-year net stock issuance anomaly, which persists for 75 months—but the persistence in alphas is not the same as the persistence in average returns; an anomaly's alpha can remain significant because the anomaly's market beta changes over time. This distinction is important: an investor at date 0 cannot construct a buy-and-hold portfolio based on five-year net stock issuance that earns positive alphas for 81 consecutive months.

6.3 Implied long-term discount rates: A back-of-the-envelope computation

What do the estimates in Table 4 imply about long-term discount rates? Although the first-year difference in discount rates between the top- and bottom-decile firms is $12 \times 0.77\% = 9.3\%$, this difference applies only when we discount next year's cash flows. We now describe a back-of-the-envelope computation of implied long-term discount rates to assess how much of the variation in short-term rates carries over to firm valuations.

We assume that all firms pay a \$1 dividend next year and that these dividends grow at a rate g per year forever. We set g to either 1% or 5%. Table 5 takes the estimates from column "All" of Table 4 to compute the term structure of discount rates for firms located in the top and bottom deciles. We assume that the risk-free rate is zero and the equity risk premium is 6% per year at all maturities. The per-month difference in average returns between the top- and bottom-decile firms over the first year is 0.77%, which implies that the appropriate discount rates are $6\% + \frac{1}{2} \times 12 \times 0.77\% = 10.6\%$ and 1.4% for the high- and low-discount rate firms, respectively.

Table 5 shows that the *second-year* discount rates are 9.0% and 3.0%. We would use these rates to discount cash flows from year 2 to 1, and then the first-year rates of 10.6% and 1.4% to discount them further to date 0. The two-year discount rates, for moving cash flows from year 2 to 0, for the top- and bottom-decile firms are therefore $\sqrt{(1 + 10.6\%)(1 + 9.0\%)} - 1 = 9.8\%$ and 2.2%, respectively. In Table 5 we show both sets of discount rates—those used to discount cash flows from year t to year $t - 1$ and those used to discount them from year t to year 0. We assume that, starting in year 11, both firms' discount rates are equal to the 6% equity risk premium.

Using these discount rates together with the expected dividends, we can compute the prices of the low- and high-discount rate firms. If $g = 1\%$, the prices of the low- and high-discount rate firms are

Table 5: Implied long-term discount rates: A back-of-the-envelope computation

We assume that the spread in low- and high-discount rate firms' discount rates are those shown in column "All" of Table 4. The risk-free rate is zero and the equity risk premium is 6% per year at all maturities. The discount rates in columns "Year $t-1$ to year t" are the per-year rates used to discount cash flows from year t to year $t-1$; the rates in columns "Year $t-1$ to year 0" are the per-year rates used to discount cash flows from year t to year 0. Both firms pay a dividend of \$1 next year and these dividends grow at a rate of $g = 1\%$ or 5% per year forever. We compute the prices of the low- and high-discount rate firms by discounting the expected dividends back using the term structures of discount rates. A firm's implied discount rate is $r_{\text{implied}} = \frac{D}{P} + g$, where P is computed using the actual term structure of discount rates.

	Discount rates				Cash flows	
	Year $t-1$ to year t		Year 0 to year t		Growth rate, g	
Year t	Bottom decile	Top decile	Bottom decile	Top decile	1%	5%
1	1.4%	10.6%	1.4%	10.6%	1.00	1.00
2	3.0%	9.0%	2.2%	9.8%	1.01	1.05
3	4.2%	7.8%	2.8%	9.1%	1.02	1.10
4	4.5%	7.5%	3.2%	8.7%	1.03	1.16
5	5.1%	6.9%	3.6%	8.4%	1.04	1.22
6	5.8%	6.2%	4.0%	8.0%	1.05	1.28
7	6.0%	6.0%	4.3%	7.7%	1.06	1.34
8	5.9%	6.1%	4.5%	7.5%	1.07	1.41
9	5.9%	6.1%	4.6%	7.4%	1.08	1.48
10	6.6%	5.4%	4.8%	7.2%	1.09	1.55
11	6.0%	6.0%	4.9%	7.0%	1.10	1.63
12	6.0%	6.0%	5.0%	7.0%	1.12	1.71
...						
100	6.0%	6.0%	5.9%	6.1%	2.68	125.24
...						

Firm valuations and implied discount rates, $r = \frac{D}{P} + g$

Growth rate, g	Firm	Price	Implied discount rate, r_{implied}
1%	Bottom decile	22.25	5.50%
	Top decile	18.04	6.54%
5%	Bottom decile	101.24	5.99%
	Top decile	81.44	6.23%

\$22.25 and \$18.04. Now, instead of applying the term structure of discount rates, suppose that we discount all cash flows back at the same rate r. A firm's value is then $P = \frac{D}{r-g}$. Given valuation P, dividend d, and the growth rate g, a firm's *implied* long-term discount rate is therefore $r_{\text{implied}} = \frac{D}{P} + g$. If $g = 1\%$, the implied discount rates are 5.5% and 6.5% for the low- and high-discount rate firms, respectively. This 1% spread in discount rates is far smaller than the 9.3% spread in year-1 rates from Table 4 because the discount rates converge over time.

The spread in discount rates decreases as g increases. Long-dated cash flows, which we discount back at more similar rates, carry now more weight. Table 5 shows that when $g = 5\%$, the prices of the low- and high-discount rate firms are \$101.24 and \$81.44, and the implied discount rates are 6.0% and 6.2%, respectively. The spread in implied long-term discount rates is now merely 0.2%.

7 Are long-term expected return differences too small to be justified by characteristics?

7.1 Measuring persistence in market betas and firm characteristics

In production-based asset pricing models, expected returns change as firms' risks change. In a model such as Gomes, Kogan, and Zhang (2003), a single-factor conditional CAPM describes returns. However, because risks (and, therefore, betas) change, firm characteristics may predict future returns better than betas estimated from historical data (Lin and Zhang 2013). If firm characteristics predict returns because they identify over- or underpriced stocks, differences in firm characteristics should converge to zero at the same rate as return differences.

In Figure 6, we measure changes in both market betas (Panel A) and firm characteristics (Panel B). We estimate betas using one quarter of daily data. We assign stocks into deciles using quarter-q

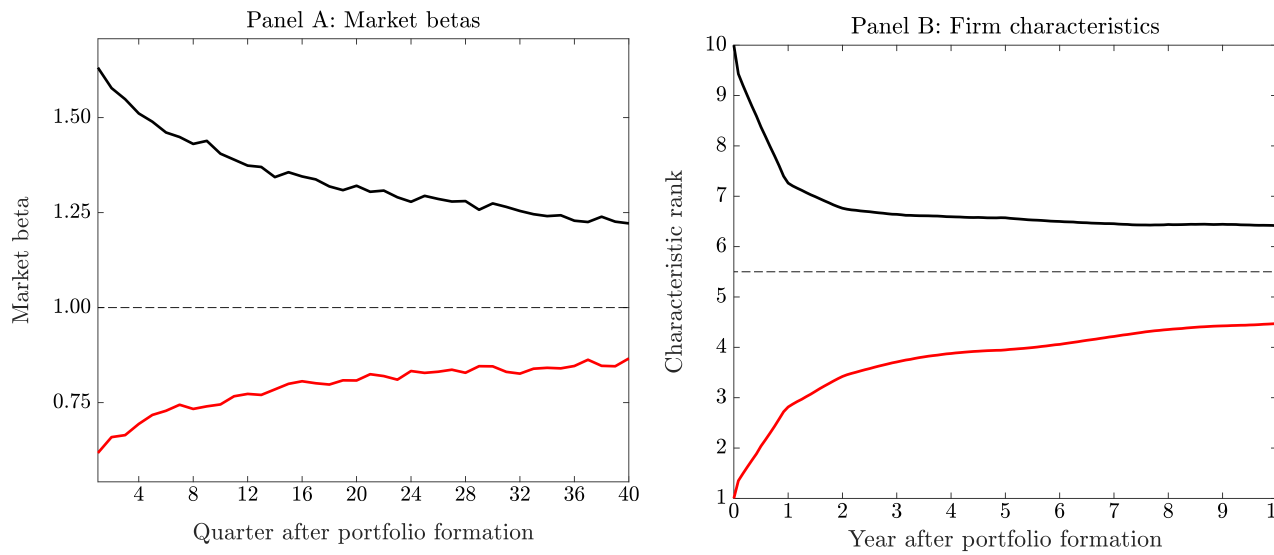

Figure 6: **Changes in market betas and firm characteristics.** This figure reports average market betas and firm-characteristic ranks for 10 years after portfolio formation. In Panel A we assign firms into deciles by quarter-q market betas and report average betas starting in quarter $q + 1$. We estimate each market beta using one quarter of daily data so that the quarterly periods do not overlap. In Panel B, we assign firms into deciles at the end of each month by a combination of 55 return predictors. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's characteristic rank (decile) is determined by the average of its non-missing percentile ranks.

estimates and then report average estimates for the next 40 quarters. By using non-overlapping data, the estimation errors at the time of the portfolio sort are uncorrelated with those in the measurement period.

Panel A shows that market betas converge toward the cross-sectional mean. In the first quarter after portfolio formation, the average beta of the firm in the top decile is 1.63; that of the firm in the bottom decile is 0.62. This initial beta gap of 1.01 narrows rapidly at first and then at a slower pace. The gap is half of its initial value after five years, and after ten years, it is 0.36. That is, 65% of the initial differences in market betas disappear in ten years.

In Panel B, we assign firms into deciles at the end of each month by a combination of the 55 return predictors listed in Table 3. As before, we convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. A firm's decile is determined by the average of its non-missing percentile ranks. By construction, the average rank difference between the top and bottom deciles is 9 at the time of portfolio formation. One year later, this difference is 4.4; after five years, it is 2.6; and after ten years, it is 1.9. Thus, similar to market betas, differences in firm characteristics evaporate rapidly at first. However, although betas and characteristics of stocks in the top and bottom deciles converge toward each other, large differences remain even after ten years. Figure A5 in the appendix shows that the predictors in the fundamental and valuation subgroups are more persistent than those in the market and event subgroups.

7.2 Do persistent and transitory firm characteristics command the same premiums?

Our bootstrapping analysis suggests that there are no differences in long-term discount rates; and when we form portfolios using multiple return predictors, any differences in expected returns collapse to zero by year five. Differences in market betas and firm characteristics, by contrast, are more persistent. To reconcile these two seemingly contradictory sets of results, we hypothesize that *persistent* differences in characteristics cannot associate with differences in average returns.

To see the intuition for this hypothesis, suppose that we can decompose firm characteristics $c_{i,t}$ into persistent and transitory components, $c^p_{i,t}$ and $c^e_{i,t}$. We then assume that expected returns are a stable function of firm characteristics,

$$\mathrm{E}_t(r_{i,t+1}) = \delta' c^p_{i,t} + \gamma' c^e_{i,t}, \tag{5}$$

where δ and γ give the associations between expected returns and the persistent and transitory firm characteristics. If all firm characteristics, no matter whether they are permanent or transitory features of the firm, have the same association with returns, $\delta = \gamma$. If so, a firm's gross profitability, for example, would predict returns the same way for firms that have always been profitable ($c^p_{i,t}$ is high) and those that are newly profitable ($c^e_{i,t}$ is high). We test the hypothesis that $\delta = \gamma$; that is, whether firm characteristics, no matter their "age", have the same associations with expected returns.

Figure 7 shows that firm characteristics are so persistent that we would expect them to associate with significant differences in average long-term returns. The black line in this figure is the same as that in Panel A of Figure 5; it represents the difference in the average monthly returns between the top and bottom deciles generated from all 55 predictors. The shaded red line plots the same top-minus-bottom return difference after replacing each stock with its characteristics-matched portfolio of stocks. We construct this return series by tracking, as in Panel B of Figure 6, a stock's decile assignment over time. If a stock's decile in month t after portfolio formation is d, we replace the stock's actual return with the value-weighted return of stocks that today belong to decile d.

The difference between the black and red lines in Figure 7 measures the difference in the premiums and discounts commanded by permanent and transitory firm characteristics. In terms of equation (5), our hypothesis $\delta = \gamma$ implies that the red and black lines should overlap. The fact that the two lines deviate implies that persistent characteristics command significantly smaller premiums than transitory characteristics.

7.3 Decomposing firm characteristics into persistent and transitory components

We can test the hypothesis that persistent firm characteristics do not predict returns by decomposing firm characteristics into persistent and transitory components. We first estimate cross-sectional regressions in which the dependent variable is today's percentile rank for predictor j and the explanatory variable is the past six-to-ten-year average of these percentile ranks:

$$pr^j_{it} = a_t + b_t \times \overline{pr}^{j,\text{6-to-10}}_{it} + e_{it}. \tag{6}$$

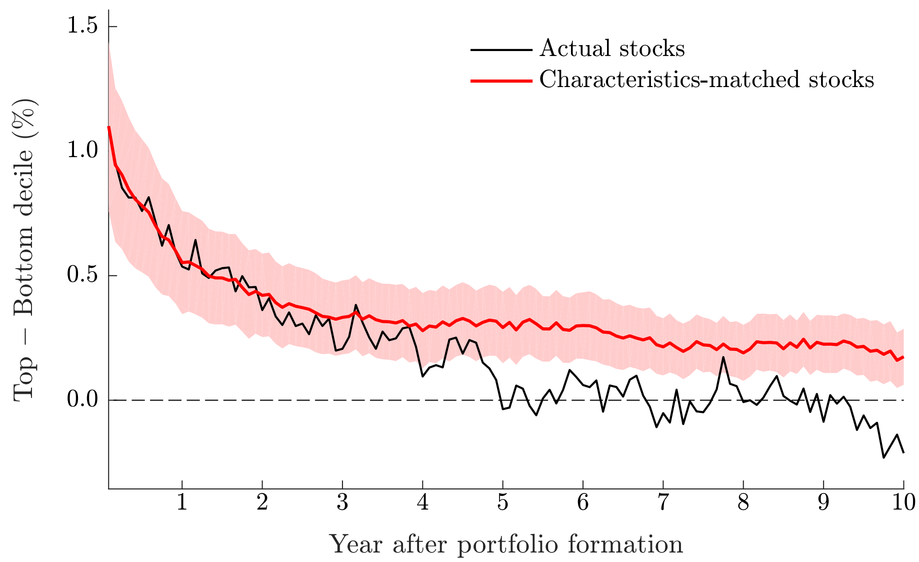

Figure 7: **Returns on characteristics-matched portfolios.** We sort stocks into decile portfolios by 55 predictors listed in Table 3. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's characteristic rank (decile) is determined by the average of its non-missing percentile ranks. We construct value-weighted portfolios each month and hold these portfolios for up to ten years. The thin black line is the average monthly return difference between the top and bottom deciles in month t following portfolio formation. The thick red line uses returns on characteristics-matched portfolios. Each month we calculate the current portfolio rank of a stock that was originally assigned into the top or bottom decile. If a stock's decile in month t is d, we replace the stock's actual return in month t with the value-weighted return on all stocks that would, as of today, belong to decile d. The red shaded area indicates 95% confidence intervals.

A firm's persistent component is the fitted value, $\hat{b}_t \times \overline{pr}_{it}^{j,\text{6-to-10}}$, and the transitory component is the residual, $\hat{e}_{it}$. We compute the average percentile rank after skipping over those five years over which characteristics seem to predict returns; *if* there are long-term differences in discount rates, we have better hope of uncovering them from long-term differences in characteristics not contaminated by any short-term variation.

The persistent component for the average anomaly is economically large. The average R^2 across all predictors and time periods from equation (6) is 20%, but the amount of persistence varies significantly across predictors. Predictors such as momentum and earnings surprise are, by definition, almost fully transitory and, as a consequence, their average R^2s from equation (6) are less than 1%; at the other

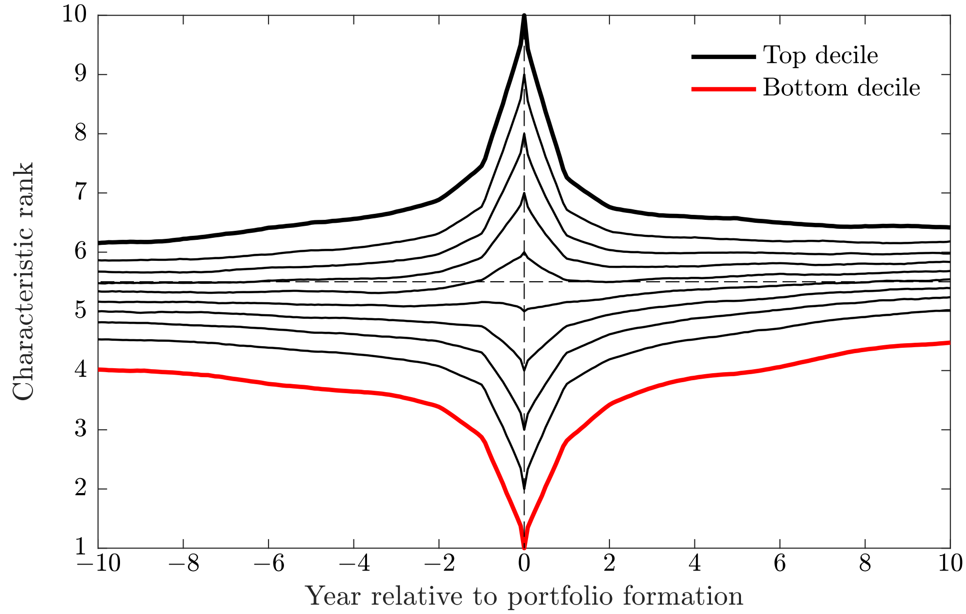

Figure 8: **Changes in firm characteristics before and after portfolio formation.** This figure reports average firm characteristic ranks for 10 years before and after portfolio formation. We assign firms into deciles each month by a combination of the 55 predictors listed in Table 3. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's characteristic rank (decile) is determined by the average of its non-missing percentile ranks.

extreme, fundamental predictors such as gross profitability (65%) and operating leverage (70%) are highly persistent. The same also applies to some market-based predictors: Amihud's illiquidity (72%) and idiosyncratic volatility (40%), for example, are highly persistent even though the return and volume data from which they are computed do not overlap from year to year.

Figure 8 illustrates the rationale behind this decomposition. This figure is the same as Panel B of Figure 6 except that we also track the evolution of firm characteristics *before* portfolio formation. Top-decile stocks tend to fall in rankings over time, and these stocks are also the ones that gained in rankings before portfolio formation. Those in the bottom decile tend to gain in rankings after portfolio formation, and they are the ones that have fallen in rankings before portfolio formation.[24] The regression in equation (6) decomposes firm characteristics by comparing a stock's (percentile) rank at the time of portfolio formation to its average pre-formation rank. By skipping over five years, we seek to isolate persistent variation in firm characteristics. An important difference between Figure 8 and our regressions is that we estimate equation (6) separately for each predictor, taking into account the fact that some predictors are more or less persistent than the average predictor in this figure.

Table 6 uses portfolio sorts similar to those in Table 4 to measure differences in the predictive powers of total, persistent, and transitory firm characteristics. In this table, we sort stocks into portfolios based on the original predictors (as in Table 4) as well as the persistent and transitory components of these predictors. We then examine differences in average returns and CAPM alphas between top and bottom deciles (Panel A) and measure the extent to which these components are incrementally informative over the other components. We note that any estimation errors reduce our ability to find meaningful differences between the persistent and transitory components; even if persistent components do not predict the cross section of returns—in terms of equation (5), $\delta > 0$ and $\gamma = 0$—we can only sort stocks into portfolios by our noisy estimates of these components.

Panel A of Table 6 shows that differences in both average returns and CAPM alphas associated with the transitory firm characteristics are typically greater than those associated with total firm characteristics. At the one-month horizon, the average return for the transitory-component strategy is 1.16% per month (t-value = 6.99); that for the total-characteristic strategy is 0.97% (t-value = 5.71). Our point estimates suggest transitory components better predict differences in average returns up to year 7 after portfolio formation. In terms of CAPM alphas, transitory components outperform total firm

[24]The near-symmetry of Figure 8 is not a coincidence. The post-formation pattern, for example, shows high-decile stocks turn into lower-decile stocks over time. Reading this post-formation pattern backwards therefore implies low-decile stocks were, on average, higher-decile stocks in the past. That is, the post-formation pattern implies the pre-formation pattern. The figure is not perfectly symmetrical, because the observations at the beginning of the sample lack pre-formation data, and those toward the end lack post-formation data.

Table 6: Decomposing predictors into persistent and transitory components

This table examines the performance for strategies that are formed based on a combination of 55 return predictors. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We rebalance the portfolios monthly and compute their value-weighted returns. Columns labeled "Total" use original predictors. Columns labeled "Persistent" and "Transitory" use the fitted values and residuals from cross-sectional regressions of characteristics against the average characteristics over the prior ten years, skipping five years. Panel A reports average returns and CAPM alphas for the difference between the top and bottom decile portfolios for different holding periods. Panel B reports alphas from regressions with two explanatory variables: the excess return on the market and the return on the strategy based on the total, permanent, or transitory firm characteristics. We adjust standard errors for overlapping returns using the Jegadeesh and Titman (1993) method. Holding-period returns in this table begin in July 1968. Estimates that are statistically significantly different from zero at the FDR $\leq$ 5% level with the Benjamini-Hochberg-Yekutieli correction for multiple-hypothesis testing are denoted with the + sign.

Panel A: Average returns and CAPM alphas

	Average returns			CAPM alphas		
Horizon	Total	Pers.	Trans.	Total	Pers.	Trans.
			Average returns and alphas			
1 month	0.97^{+}	0.22	1.16^{+}	1.19^{+}	0.27	1.37^{+}
Year 1	0.69^{+}	0.15	0.80^{+}	0.90^{+}	0.19	1.00^{+}
Year 2	0.45^{+}	0.12	0.56^{+}	0.65^{+}	0.15	0.73^{+}
Year 3	0.31	0.14	0.40^{+}	0.53^{+}	0.16	0.60^{+}
Year 4	0.32	0.18	0.36^{+}	0.51^{+}	0.20	0.52^{+}
Year 5	0.23	0.05	0.23	0.39^{+}	0.08	0.37^{+}
Year 6	0.06	0.03	0.08	0.22	0.07	0.21
Year 7	0.08	0.10	0.12	0.22	0.15	0.23
Year 8	0.06	0.11	−0.01	0.18	0.17	0.05
Year 9	0.10	0.09	0.04	0.18	0.15	0.08
Year 10	−0.19	0.10	−0.13	−0.11	0.16	−0.07
			***t*-values**			
1 month	5.71^{+}	1.98	6.99^{+}	7.94^{+}	2.51	9.14^{+}
Year 1	4.64^{+}	1.39	5.54^{+}	7.12^{+}	1.83	7.85^{+}
Year 2	3.16^{+}	1.14	3.98^{+}	5.15^{+}	1.42	5.76^{+}
Year 3	2.21	1.33	2.85^{+}	4.30^{+}	1.53	4.77^{+}
Year 4	2.21	1.77	2.62^{+}	3.89^{+}	1.92	4.06^{+}
Year 5	1.75	0.49	1.81	3.23^{+}	0.77	2.95^{+}
Year 6	0.48	0.28	0.65	1.98	0.58	1.90
Year 7	0.61	0.86	0.96	1.92	1.22	1.87
Year 8	0.52	0.90	−0.12	1.53	1.36	0.40
Year 9	0.86	0.69	0.30	1.57	1.09	0.63
Year 10	−1.57	0.73	−1.14	−0.94	1.12	−0.62

Panel B: Spanning regressions

	Total		Persistent		Transitory	
Horizon	(1)	(2)	(3)	(4)	(5)	(6)
			Alphas			
1 month	1.14^{+}	0.00	0.13	0.25	0.34^{+}	1.36^{+}
Year 1	0.87^{+}	−0.01	0.08	0.22	0.16^{+}	1.01^{+}
Year 2	0.62^{+}	−0.01	0.06	0.19	0.14	0.75^{+}
Year 3	0.49^{+}	0.01	0.08	0.20	0.12	0.61^{+}
Year 4	0.47^{+}	0.03	0.14	0.24	0.08	0.54^{+}
Year 5	0.37^{+}	0.07	0.01	0.10	0.02	0.37^{+}
Year 6	0.20	0.05	0.01	0.09	0.04	0.21
Year 7	0.19	0.05	0.10	0.18	0.04	0.25
Year 8	0.16	0.14	0.14	0.17	−0.11	0.08
Year 9	0.16	0.12	0.12	0.16	−0.08	0.09
Year 10	−0.12	−0.05	0.17	0.15	0.02	−0.05
			***t*-values**			
1 month	7.67^{+}	0.05	1.17	2.16	4.42^{+}	9.08^{+}
Year 1	7.00^{+}	−0.16	0.72	1.96	3.09^{+}	7.90^{+}
Year 2	4.97^{+}	−0.19	0.55	1.79	2.40	5.89^{+}
Year 3	4.06^{+}	0.11	0.75	1.79	2.01	4.88^{+}
Year 4	3.60^{+}	0.48	1.30	2.25	1.43	4.21^{+}
Year 5	3.13^{+}	1.25	0.08	0.91	0.30	2.99^{+}
Year 6	1.88	0.76	0.06	0.76	0.53	1.97
Year 7	1.67	0.74	0.79	1.45	0.54	2.01
Year 8	1.36	2.16	1.16	1.45	−1.61	0.63
Year 9	1.43	1.88	0.90	1.19	−1.28	0.79
Year 10	−1.01	−0.78	1.19	1.05	0.35	−0.45
RHS Factors:						
Market	×	×	×	×	×	×
Total			×		×	
Persistent	×					×
Transitory		×		×		

characteristics up to year 4. Permanent firm components, by contrast, are largely uninformative about future returns. The month following portfolio formation, the strategy that sorts stocks into portfolios by permanent firm characteristics earns a return of just 0.22% (t-value = 1.98), and the performance deteriorates as we look at more distant holding periods.

Similar to Table 4, we identify estimates that are statistically significantly different from zero at the 5% level under the Benjamini-Hochberg-Yekutieli method with the + sign. Adjusted for multiple comparisons, Panel A of Table 6 shows that both total and transitory firm characteristics significantly predict the cross section of returns up to five years after portfolio formation; permanent firm characteristics do not predict returns at any horizon.

In Panel B of Table 6 we report estimates from regressions in which the dependent variable is one of the three strategies and the right-hand side factors are the market factor and one of the other strategies. The first column, for example, takes the strategy based on total firm characteristics and regresses it against the market factor and the strategy based on persistent firm characteristics. The alphas from these spanning regressions measure the incremental information content of the left-hand side strategy over the right-hand side factors and strategies.[25] The statistically significant one-month alpha in the first column, for example, indicates that an investor trading the market and the strategy based on persistent firm characteristics could have earned a statistically significantly higher Sharpe ratio over the sample period by tilting the portfolio towards the total-characteristics strategy.

The spanning regressions in Panel B show that transitory characteristics contain all information about the cross section of returns: (1) total firm characteristics are incrementally informative over persistent components, but not over transitory components; (2) transitory firm characteristics are incrementally informative over both total and persistent characteristics; and (3) persistent characteristics

[25]See Huberman and Kandel (1987) and Barillas and Shanken (2017).

are not incrementally informative over total or transitory characteristics. Although some individual estimates in columns 2 and 4 are statistically significant under classical hypothesis testing, the Benjamini-Hochberg-Yekutieli method attributes these estimates to chance.

We would like to highlight the ability of transitory characteristics to subsume the predictive power of total characteristics. In principle, the persistent component in characteristics could be economically so small that the sorts on the total and transitory characteristics would be nearly identical. If so, each sort would subsume the other. This is not what we see in the data: transitory characteristics subsume total characteristics, but not vice versa.

We have performed the decomposition in as simple of a way as possible to avoid any concerns of data dredging. Instead of using the same model in equation (6) for every characteristic, we could let the model vary depending on how persistent each characteristic is, and the horizon over which it predicts the cross section of returns. In this sense, the results of our decomposition can be viewed as conservative estimates of the predictive power of transitory characteristics.

The result that only transitory firm characteristics predict returns is consistent with the conjecture that differences in expected returns do not persist. If a characteristic such as gross profitability is more persistent than the differences in returns it explains, the persistent component of this characteristic cannot possibly explain return differences. Table 6 shows that this intuition extends to all predictors.

8 Conclusions

We show that cross-sectional differences in stocks' expected returns converge to zero in five years. Using different combinations of 55 return predictors, we are unable to identify any differences in long-term returns beyond this point. We devise a characteristics-free bootstrapping procedure that measures our ability to detect persistent cross-sectional differences in expected returns, and find no evidence of such differences. If differences in expected returns persisted, past average stock returns would positively and significantly predict the cross section of stock returns. In the data, this sign is negative.

Our results are consistent with firms' risks changing over time or, alternatively, with most cross-sectional differences in expected returns emanating from mispricing. We show that in production-based asset pricing models such as Berk, Green, and Naik (1999) and Zhang (2005), firms' expected returns converge toward the mean the same way they do in the actual data. We further show that the convergence in expected returns can generate long-term reversals, that is, the negative association between past and future returns. Long-term reversals are the inevitable consequence of a rapid convergence in discount rates; they are thus not necessarily an indicator of the markets overreacting to new information (De Bondt and Thaler 1985).

The rapid convergence of discount rates prompts us to revisit the relation between firm characteristics and expected returns. We find that characteristics are more persistent than the differences in expected returns they predict. Motivated by this finding, we decompose characteristics into permanent and transitory components and find that transitory components contain all the information to predict the cross section of stock returns. Persistent differences in characteristics do not associate with any differences in average returns.

If, as our results suggest, firms' expected returns rapidly converge to the mean, analysts should use about the same discount rates to value stocks. Our back-of-the-envelope computation suggests that the differences in implied long-term discount rates are far smaller than those in short-term rates. Consider, for example, bottom- and top-decile firms whose one-year-ahead discount rates are 1.4% and 10.6%, respectively. The implied long-term discount rates for these firms—defined here as the constant discount rates that value each firm correctly—are 5.5% and 6.5%, respectively. Put differently, the amount of cross-sectional dispersion in *short-term* discount rates far exceeds that embedded in firm valuations.

The relevance of the short-term differences in expected returns depends on the lifespan of a firm's projects. If the average project generates long-term cash flows, the role of initial differences in discount rates will be small. Dechow, Sloan, and Soliman (2004) estimate that the average equity duration of publicly traded U.S. firms is 15.1 years. This estimate together with ours suggests the effective long-term discount rates of most firms must be close to each other.

The finding that differences in unconditional expected returns are small goes against the conventional wisdom that discount rates should vary across firms and even across projects within the same firm. For example, corporate finance textbooks allow for considerable variation in the cost of equity, and rarely consider the idea that firms' discount rates would converge over time.[26]

Readers suggest our results are surprising in light of Proposition II of Modigliani and Miller (1958): if two firms' *unlevered* rates of return on equity are the same, but their debt-to-equity ratios differ, their expected levered rates of returns differ as well. The assumption of similar unlevered rates of return on equity is reasonable. High rates of returns attract competition, which can be expected to drive the rates of return on investment toward equality (Stigler 1963; Fama and French 2000). Debt-to-equity ratios are also known to differ considerably across firms. Although leverage ratios display convergence toward the mean, economically significant differences in leverage persist for long periods of time (Lemmon, Roberts, and Zender 2008).

Why do the data then not support the prediction that expected *levered* rates of return differ across firms?[27] The idea that equity risk is increasing in leverage relies on the assumption that markets are frictionless, which leads firms' investment and financing decisions to be independent of one another. George and Hwang (2010) argue financial distress costs affect firms' capital structure decisions, induc-

[26] Welch (2017) is a rare exception to this tradition. His view reflects that of Levi and Welch (2017), who recommend that the "cost-of-capital estimates should be shrunk far more than is common practice." See Levi and Welch for a review of the cost-of-capital prescriptions of corporate finance textbooks and academic literature.

[27] For example, Linnainmaa and Roberts (2018, Table 5) document that leverage is not a significant predictor of returns. Welch (2018) finds that increases in leverage, if anything, lower average returns and increase volatility.

ing low-leverage firms to assume greater exposures to systematic risk than high-leverage firms. This neutralizes the mechanical effect leverage has on equity risk. Johnson et al. (2011) draw similar conclusions from a generalized version of the George and Hwang (2010) model. The puzzle can thus be explained by rational models, albeit ones with market frictions.

Our results also have asset pricing implications. Reliably extracting the risk-return relationship from the data may be difficult even when one exists. Suppose, for example, the true data-generating process is a conditional CAPM. Our results would then imply firms' betas must change rapidly to match the changes in discount rates. Lin and Zhang (2013) make an analogous argument by simulating data from Zhang's (2005) model and showing the resulting data appear to support the characteristics-based model of Daniel and Titman (1997) and not the risk-based model. Our results on the changes in average returns, betas, and characteristics point toward the same conclusion. To give asset pricing models a fair chance, econometric methods for testing them must accommodate rapid changes in firm risks.

REFERENCES

Ang, A. and J. Liu (2004). How to discount cashflows with time-varying expected returns. *Journal of Finance 59*(6), 2745–2783.

Barillas, F. and J. Shanken (2017). Which alpha? *Review of Financial Studies 30*(4), 1316–1338.

Barth, M. E., Y. Konchitchki, and W. R. Landsman (2013). Cost of capital and earnings transparency. *Journal of Accounting and Economics 55*(2-3), 206–224.

Berk, J. B., R. C. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *Journal of Finance 54*(5), 1553–1607.

Binsbergen, J. H. and C. C. Opp (2019). Real anomalies. *Journal of Finance 74*(4), 1659–1706.

Black, F. (1986). Noise. *Journal of Finance 41*(3), 528–543.

Chordia, T., A. Goyal, and A. Saretto (2020). Anomalies and false rejections. *Review of Financial Studies 33*(5), 2134–2179.

Christiano, L. J. and M. Eichenbaum (1992). Current real-business-cycle theories and aggregate labor-market fluctuations. *American Economic Review 82*(3), 430–450.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Conrad, J. and G. Kaul (1998). An anatomy of trading strategies. *Review of Financial Studies 11*, 489–519.

Cooper, I. (2006). Asset pricing implications of nonconvex adjustment costs and irreversibility of investment. *Journal of Finance 61*(1), 139–170.

Cvitanić, J., A. Lazrak, L. Martellini, and F. Zapatero (2006). Dynamic portfolio choice with parameter uncertainty and the economic value of analysts' recommendations. *Review of Financial Studies 19*(4), 1113–1156.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

De Bondt, W. F. M. and R. H. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), 379–395.

Dechow, P. M., R. G. Sloan, and M. T. Soliman (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies 9*(2–3), 197–228.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (1997). Industry costs of equity. *Journal of Financial Economics 43*(2), 153–193.

Fama, E. F. and K. R. French (1999). The corporate cost of capital and the return on corporate investment. *Journal of Finance 54*(6), 1939–1967.

Fama, E. F. and K. R. French (2000). Forecasting profitability and earnings. *Journal of Business 73*(2), 161–175.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Francis, J., D. Nanda, and P. Olsson (2008). Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research 46*(1), 53–99.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001). Toward an implied cost of capital. *Journal of Accounting Research 39*(1), 135–176.

George, T. J. and C.-Y. Hwang (2010). A resolution of the distress risk and leverage puzzles in the cross section of stock returns. *Journal of Financial Economics 96*(1), 56–79.

Gomes, J., L. Kogan, and L. Zhang (2003). Equilibrium cross section of returns. *Journal of Political Economy 111*(4), 693–732.

Gu, L., D. Hackbarth, and T. Johnson (2017). Inflexibility and stock returns. *Review of Financial Studies 31*(1), 278–321.

Hackbarth, D. and T. Johnson (2015). Real options and risk dynamics. *Review of Economic Studies 82*(4), 1449–1482.

Hann, R. N., M. Ogneva, and O. Ozbas (2013). Corporate diversification and the cost of capital. *Journal of Finance 68*(5), 1961–1999.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ...And the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis 22*(3), 329–343.

Hou, K., M. A. Van Dijk, and Y. Zhang (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics 53*(3), 504–526.

Huberman, G. and S. Kandel (1987). Mean-variance spanning. *Journal of Finance 42*(4), 873–888.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance 56*(2), 699–720.

Johnson, T. C., T. Chebonenko, I. Cunha, F. D'Almeida, and X. Spencer (2011). Endogenous leverage and expected stock returns. *Finance Research Letters 8*(3), 132–145.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica 50*(6), 1345–1370.

Lemmon, M. L., M. R. Roberts, and J. F. Zender (2008). Back to the beginning: Persistence and the cross-section of corporate capital structure. *Journal of Finance 63*(4), 1575–1608.

Levi, Y. and I. Welch (2017). Best practice for cost-of-capital estimates. *Journal of Financial and Quantitative Analysis 52*(2), 427–463.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. *Review of Financial Studies 15*(2), 533–563.

Lin, X. and L. Zhang (2013). The investment manifesto. *Journal of Monetary Economics 60*(3), 351–366.

Linnainmaa, J. T. and M. R. Roberts (2018). The history of the cross section of stock returns. *Review of Financial Studies 31*(7), 2606–2649.

Lo, A. W. and A. C. MacKinlay (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies 3*(2), 175–205.

MacKinlay, A. G. and L. Pastor (2000). Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies 13*(4), 883–916.

Martin, I. and C. Wagner (2019). What is the expected return on a stock? *Journal of Finance 74*(4), 1887–1929.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *Journal of Finance 71*(1), 5–32.

Modigliani, F. and M. H. Miller (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review 48*(3), 261–297.

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

Pástor, L. and R. F. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *Journal of Finance 64*(4), 1583–1628.

Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance 52*(1), 327–340.

Shumway, T. and V. A. Warther (1999). The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *Journal of Finance 54*(6), 2361–2379.

Stigler, G. J. (1963). Competition and the rate of return. In *Capital and Rates of Return in Manufacturing Industries*, Chapter 3, pp. 54–71. Princeton University Press.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *Journal of Finance 57*(1), 233–264.

Welch, I. (2017). *Corporate Finance, 4th edition.* United States: Lulu.

Welch, I. (2018). Leverage-and cash-based tests of risk and reward with improved identification. UCLA working paper.

Zhang, L. (2005). The value premium. *Journal of Finance 60*(1), 67–103.

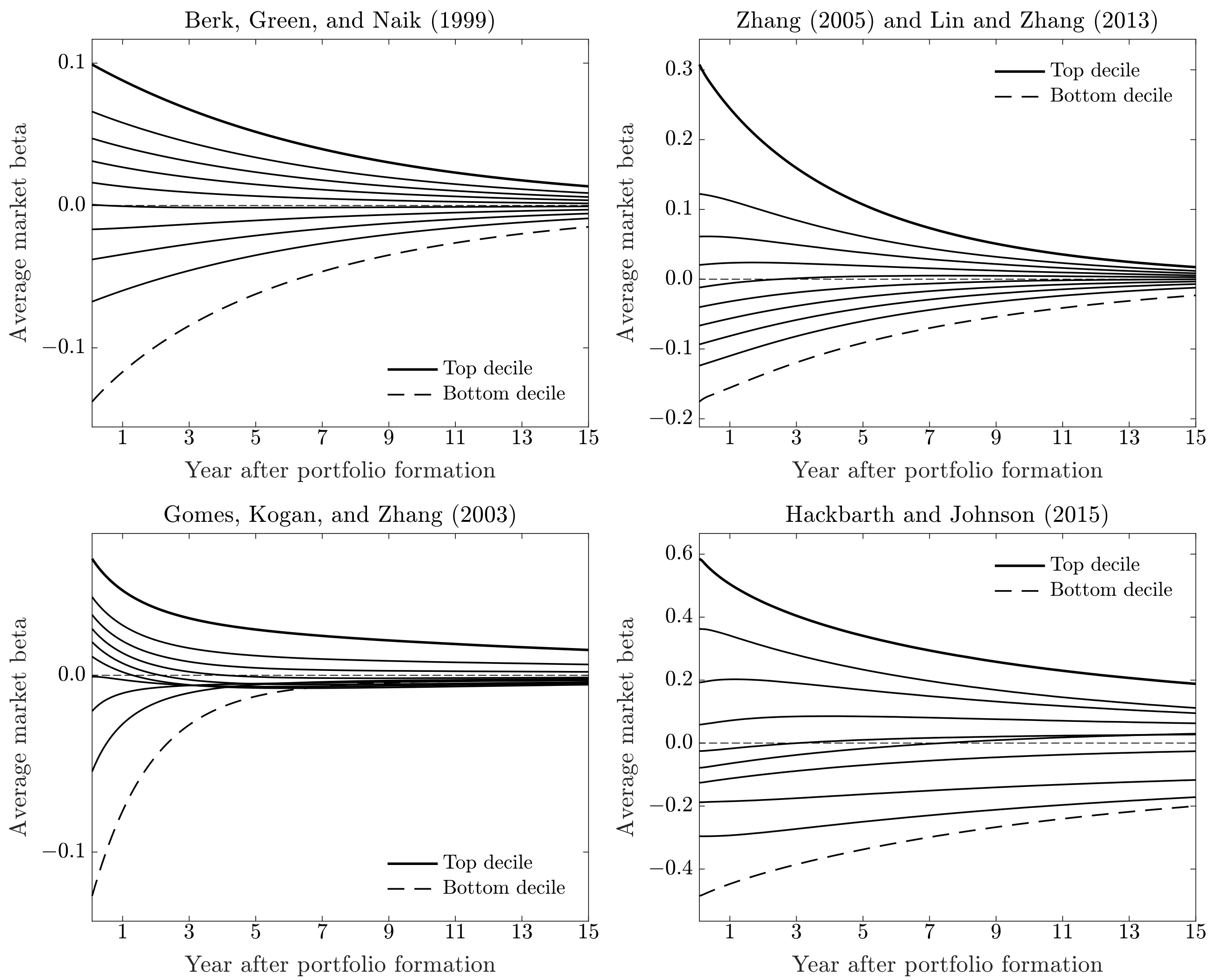

Figure A1: **Average market betas on stocks sorted by expected returns.** We simulate 1,000 months of return data from the four models described in Section 2. We run these simulations using the same parameters as those used in the original studies. We discard the first 400 months and then begin ranking stocks into deciles based on expected returns. We report average cross-sectionally demeaned market betas for these deciles over the next 15 years after portfolio formation.

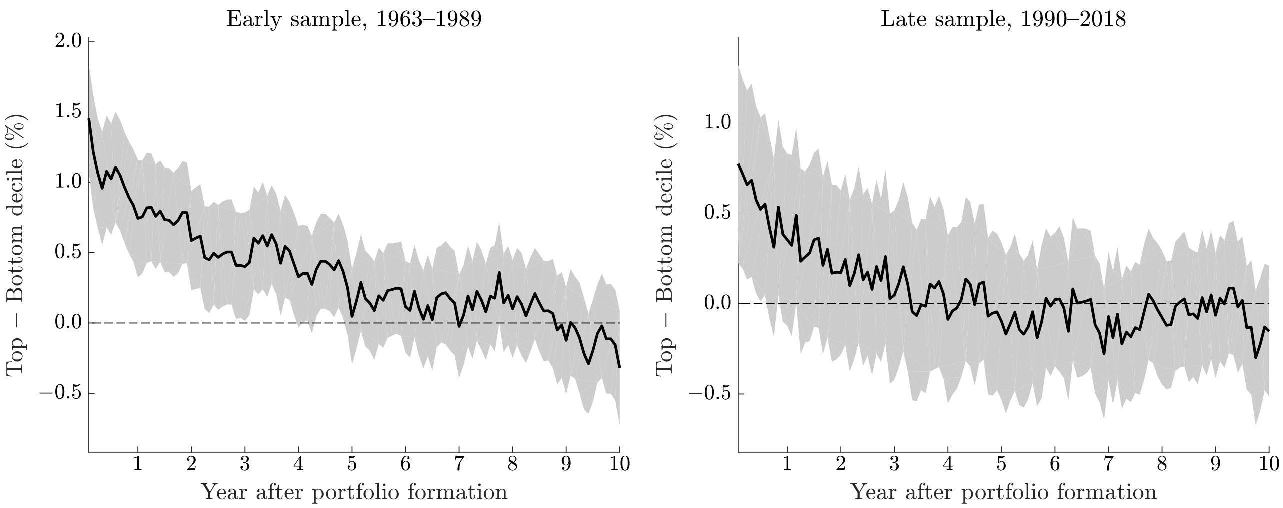

Figure A2: **Average monthly returns on long-short strategies in months 1–120 after portfolio formation: Subsamples.** We sort stocks into decile portfolios using 55 return predictors listed in Table 3. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We construct value-weighted portfolios each month and hold these portfolios for up to ten years. We report differences in average returns between the top and bottom deciles using the 1963–1989 (left panel) and 1990–2018 samples (right panel). The average return in month t is only the month t return, not the average return from today to month t. The shaded areas indicate 95% confidence intervals.

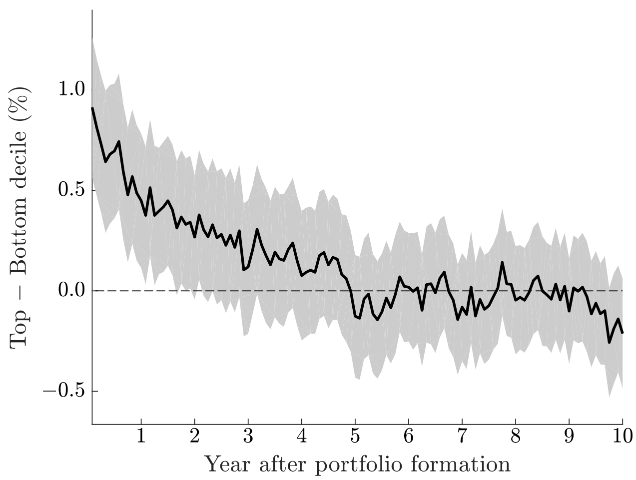

Figure A3: **Average monthly returns on long-short strategies in months 1–120 after portfolio formation: Conditional on survival.** We sort stocks into decile portfolios using 55 return predictors listed in Table 3. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We construct value-weighted portfolios each month and hold these portfolios for up to ten years. We report differences in average returns between the top and bottom deciles. This table differs from Figure 3 in that we condition on firm survival: each month the sample includes only those firms that will go on to survive for at least 10 years. The average return in month t is only the month t return, not the average return from today to month t. The shaded areas indicate 95% confidence intervals.

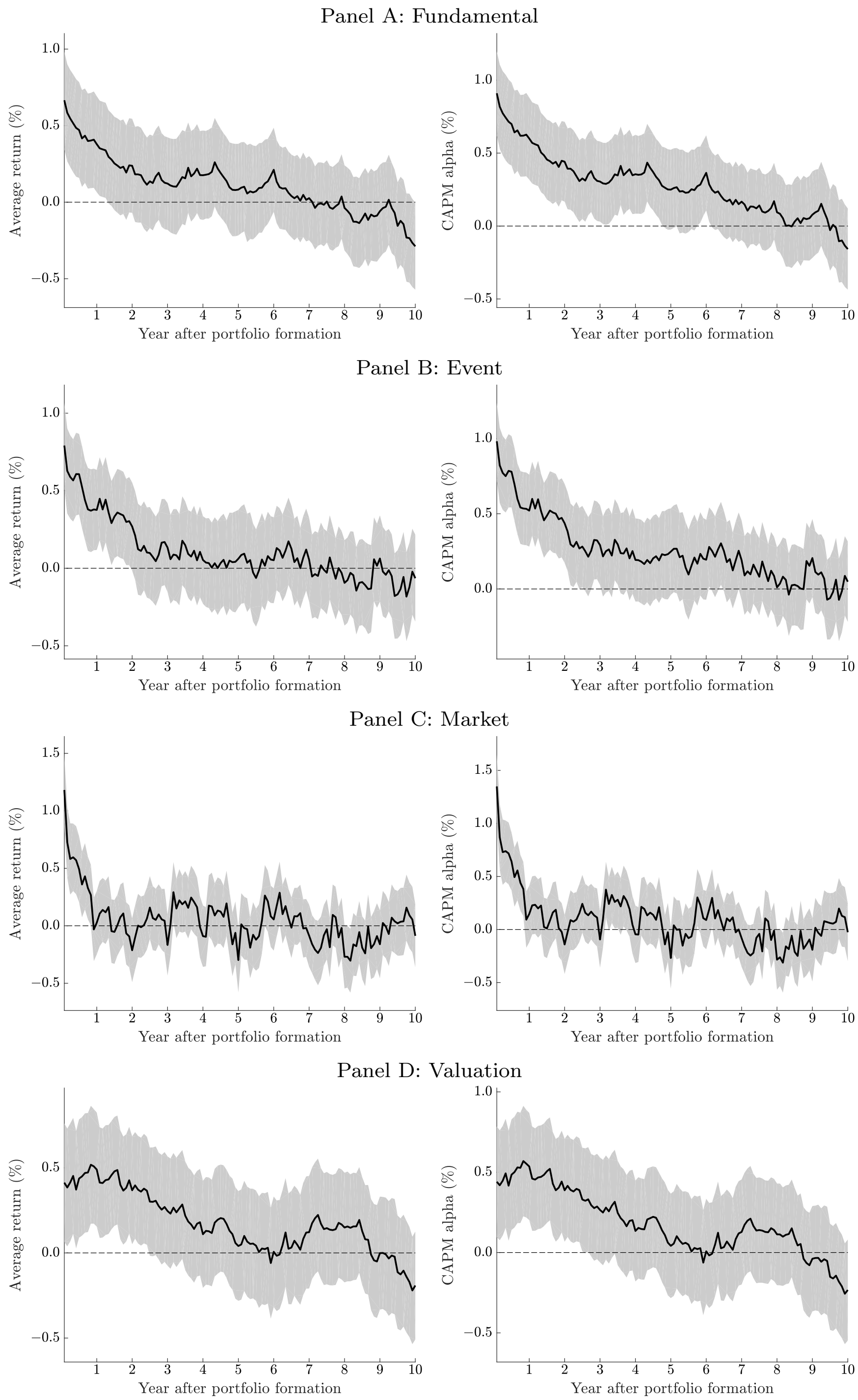

Figure A4: **Average monthly returns and CAPM alphas on long-short strategies in months 1–120 after portfolio formation.** This figure reports average returns and CAPM alphas for long-short strategies. The computations are the same as those in Figure 5 except that, instead of using all 55 predictors, each panel uses a different subset of predictors.

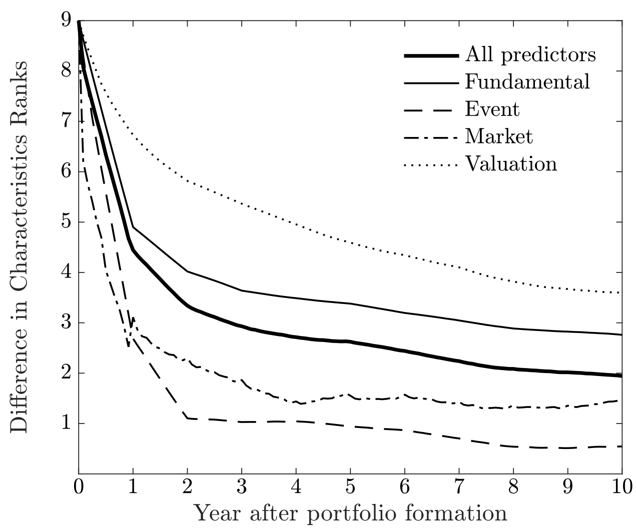

Figure A5: **Changes in firm characteristics by predictor category.** This figure reports changes in firm-characteristic ranks for 10 years after portfolio formation. We assign firms into deciles each month by a combination of 55 return predictors or one of the following subsets of predictors: fundamental, event, market, and valuation. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. A firm's characteristic rank (decile) is determined by the average of its non-missing percentile ranks. This figure reports the average difference in the characteristic ranks between the top and bottom deciles.

Table A1: Annual post-formation spread portfolio returns: Equal-weighted portfolios

We sort stocks into decile portfolios using 55 return predictors ("all") or four predictor subsets: (1) fundamental, (2) event, (3) market, and (4) valuation. We convert each predictor into a cross-sectional percentile rank using NYSE-listed stocks as the reference distribution. Each stock's predictor is the average of its non-missing percentile ranks. We form portfolios each month and compute equal-weighted returns for these portfolios for up to ten years. We report average monthly returns and CAPM alphas for high-minus-low portfolios for different horizons following portfolio formation. The holding periods are nonoverlapping: year 3, for example, is the average return over just the third year after portfolio formation. We adjust standard errors for overlapping returns using the Jegadeesh and Titman (1993) method. Estimates that are statistically significantly different from zero at the FDR $\leq$ 5% level under the Benjamini-Hochberg-Yekutieli method are denoted with the + sign.

		Subset of predictors					Subset of predictors			
Horizon	All	(1)	(2)	(3)	(4)	All	(1)	(2)	(3)	(4)
	Average monthly returns (%)					Monthly CAPM alphas (%)				
1 month	1.89$^+$	0.85$^+$	1.37$^+$	2.33$^+$	1.02$^+$	2.14$^+$	1.02$^+$	1.51$^+$	2.54$^+$	1.16$^+$
Year 1	1.06$^+$	0.62$^+$	0.95$^+$	0.51$^+$	1.04$^+$	1.31$^+$	0.78$^+$	1.08$^+$	0.73$^+$	1.20$^+$
Year 2	0.55$^+$	0.35	0.51$^+$	−0.02	0.83$^+$	0.77$^+$	0.49$^+$	0.63$^+$	0.15	0.98$^+$
Year 3	0.41$^+$	0.23	0.37$^+$	0.01	0.64$^+$	0.60$^+$	0.37$^+$	0.48$^+$	0.16	0.76$^+$
Year 4	0.39$^+$	0.25	0.27$^+$	0.09	0.56$^+$	0.58$^+$	0.38$^+$	0.38$^+$	0.22$^+$	0.68$^+$
Year 5	0.24	0.12	0.12	0.00	0.44$^+$	0.41$^+$	0.24	0.24$^+$	0.11	0.55$^+$
Year 6	0.18	0.15	0.09	0.00	0.22	0.34$^+$	0.25	0.19$^+$	0.11	0.32$^+$
Year 7	0.13	0.09	0.09	−0.05	0.20	0.27$^+$	0.19	0.18$^+$	0.05	0.30$^+$
Year 8	0.09	0.09	−0.01	−0.10	0.24	0.23$^+$	0.20	0.08	−0.01	0.33$^+$
Year 9	0.03	0.05	−0.05	−0.06	0.05	0.16	0.15	0.02	0.02	0.11
Year 10	0.01	0.04	0.08	0.10	−0.10	0.14	0.13	0.16$^+$	0.20$^+$	−0.03
	t-values					t-values				
1 month	10.82$^+$	5.16$^+$	13.56$^+$	14.50$^+$	5.93$^+$	13.74$^+$	6.50$^+$	16.73$^+$	17.30$^+$	7.21$^+$
Year 1	6.35$^+$	3.83$^+$	10.69$^+$	3.83$^+$	6.63$^+$	8.85$^+$	5.12$^+$	13.94$^+$	6.40$^+$	8.17$^+$
Year 2	3.45$^+$	2.22	6.14$^+$	−0.14	5.88$^+$	5.46$^+$	3.34$^+$	8.53$^+$	1.48	7.41$^+$
Year 3	2.69$^+$	1.52	4.34$^+$	0.14	4.78$^+$	4.49$^+$	2.54$^+$	6.25$^+$	1.84	6.01$^+$
Year 4	2.78$^+$	1.77	3.25$^+$	0.94	4.15$^+$	4.58$^+$	2.80$^+$	5.12$^+$	2.74$^+$	5.31$^+$
Year 5	1.81	0.86	1.52	−0.01	3.51$^+$	3.50$^+$	1.87	3.36$^+$	1.36	4.56$^+$
Year 6	1.48	1.15	1.09	−0.02	1.82	3.10$^+$	2.06	2.71$^+$	1.46	2.80$^+$
Year 7	1.07	0.69	1.16	−0.57	1.71	2.51$^+$	1.55	2.57$^+$	0.67	2.69$^+$
Year 8	0.76	0.75	−0.17	−1.27	2.10	2.20$^+$	1.67	1.15	−0.17	3.00$^+$
Year 9	0.33	0.44	−0.68	−0.71	0.48	1.71	1.30	0.37	0.30	1.10
Year 10	0.07	0.32	0.99	1.13	−1.01	1.42	1.11	2.31$^+$	2.68$^+$	−0.32