The Performance of Characteristic-Sorted Portfolios: Evaluating the Past and Predicting the Future

AYDOĞAN ALTI, TRAVIS L. JOHNSON, and SHERIDAN TITMAN*

February 2024

Abstract

We study the fluctuating performance of characteristic-sorted portfolios through the lens of a statistical model that allows for persistent variation in expected returns. The model provides a simple formula for adjusting the standard errors of unconditional (or long-run) expected return estimates; with plausible persistence parameters, the standard errors double. Maximum likelihood tests of the model show that the magnitude of persistent variation is estimated very imprecisely: the historical data are consistent with both i.i.d. and highly persistent return processes. Finally, our Bayesian analysis of the model shows that with relatively agnostic prior beliefs about the persistence parameters, posterior estimates of expected returns exhibit large fluctuations over time.

*All three authors are with the Department of Finance, the University of Texas at Austin. Contact information: aydogan.alti@mccombs.utexas.edu, travis.johnson@mccombs.utexas.edu, and sheridan.titman@mccombs.utexas.edu. We thank Svetlana Bryzgalova (discussant), Scott Cederburg, Eugene Fama, Kenneth French, Amit Goyal (discussant), Xindi He (discussant), Narasimhan Jegadeesh, Lubos Pastor, Allan Timmermann, and seminar participants at the American Finance Association (AFA) Meetings, Asian Bureau of Finance and Economic Research (ABFER) Conference, Financial Intermediation Research Society (FIRS) Conference, UGA Fall Finance Conference, Baylor University, Chinese University of Hong Kong, Hong Kong University, HKUST, Peking University, Thammasat University, and the University of Texas at Austin for helpful comments, and the Q Group for awarding us the 2022 Jack Treynor prize.

A large and growing literature links firm characteristics, such as valuation ratios, to expected stock returns. While the evidence documented in this literature convincingly rejects the CAPM, going beyond this rejection and interpreting alternatives has proven to be challenging. In particular, it is an open question whether the historical links between characteristics and returns represent permanent economic forces that will continue to shape returns in the future, or transitory forces that will gradually dissipate.

A case in point is the value premium. The historical tendency of value stocks to outperform growth stocks, shown by Fama and French (1992) and others, has weakened in the most recent period. Yet as Fama and French (2021) point out, we cannot confidently conclude that the value premium post-1991 is different than the value premium pre-1991. Although the value premium is quite high in the early period, and is not statistically significant in the latter period, the returns are sufficiently noisy that we cannot reject the hypothesis that they are drawn from identical distributions.[1]

This paper examines the returns of characteristic-sorted portfolios through the lens of a statistical model that allows expected returns to fluctuate over time. We use this model to analyze how persistent fluctuations influence our inferences about portfolio returns. As we show, accounting for such fluctuations have important implications for how we interpret the historical performance of characteristic-sorted portfolios, as well as how to optimally forecast future returns using information from historical returns.

Our analysis provides three main contributions. First, we derive a simple formula that adjusts the standard errors of *unconditional* expected return estimates to account for persistent variation in *conditional* expected returns. We show that, given plausible persistence parameters, this adjustment can double the standard errors, casting doubt on the size and significance of long-run return premia associated with characteristics. Second, we estimate the full model, including the persistence parameters, using maximum likelihood. We find that the inferences about persistent variation in expected returns are very imprecise: the

[1] Another example of long-term fluctuations in characteristic-sorted portfolio returns is the strengthening profitability effect, documented by Novy-Marx (2013).

historical data are consistent with both i.i.d. and highly persistent parameterizations of the model. Third, we analyze our model's implications within a Bayesian framework. We find that, given relatively agnostic prior beliefs about persistence parameters, Bayesian investors infer large and persistent fluctuations in conditional expected returns and thus chase characteristic performance trends. We also show that the Bayesian application of the model generates improved out-of-sample return forecasts of characteristic-sorted portfolios relative to forecasts based on historical averages.

The premise of our paper – that expected returns fluctuate over time – should not be controversial; most existing rational and behavioral asset pricing theories would predict such fluctuations to occur as the underlying forces that link returns to characteristics change over time.[2] However, whether such fluctuations have a material impact on statistical inference and return forecasts depends on their persistence. For instance, large but relatively short-lived fluctuations that can be exploited with momentum and reversal strategies are unlikely to affect estimates of how a characteristic will perform over the next decade. Conversely, a process with small but highly persistent shocks that last for most of the sample period would be nearly impossible to distinguish from i.i.d. returns. In this case, the historical average return may provide as good a predictor of future performance as any. As our analysis shows, time variation in expected returns influences return forecasts the most when it exhibits intermediate levels of persistence, ranging from business-cycle frequencies (e.g., expected return shocks with a half-life of about one year) to somewhat longer frequencies that may capture the timing of structural economic changes (e.g., a half-life of five to ten years). In this case, the optimal forecast of future performance overweights performance in recent years.

The main focus of our empirical analysis is on the returns of the value, investment, profitability, and size portfolios studied in Fama and French (2015).[3] As a first step we estimate

[2]We provide further motivation for the theoretical underpinnings of our statistical model in Section 1.1.

[3]Throughout the paper, we analyze returns of market-neutral portfolios that are constructed using value-weighted quintiles of the individual characteristics.

predictive regressions of the quarterly returns of these portfolios on their past returns cumulated over various horizons. We find that for all four portfolios, the past year's return positively predicts the next quarter's return, and while only the size portfolio's coefficient estimate is significant on its own, joint tests of the four portfolios strongly reject the null hypothesis of no predictability. In contrast, past returns cumulated over longer horizons do not significantly predict future returns. These findings are consistent with characteristic portfolio returns exhibiting intermediate levels of persistence; however, the estimates are sufficiently imprecise that they are also consistent with a wide variety of return processes.

We then use our model to assess how persistent variation affects statistical inferences about expected returns. Specifically, we derive a closed-form formula for adjusting the standard errors of unconditional expected return estimates as a function of the model-implied first-order return autocorrelation and its decay rate over time. This adjustment results in higher standard errors when returns exhibit more persistent variation – that is, when the first-order autocorrelation is more positive and decays more slowly. Intuitively, there are effectively fewer independent observations with persistent variation relative to the i.i.d. case, resulting in less precise expected return estimates.

As our estimates illustrate, different assumptions about the magnitude of persistent variation in characteristic-sorted portfolio returns generate substantially different inferences about unconditional expected returns. If returns are assumed to be i.i.d., meaning they have no persistence, standard errors are low and the hypothesis of zero unconditional expected returns is strongly rejected. However, if returns are assumed to exhibit plausible degrees of persistence, the model-adjusted standard errors are substantially higher and the hypothesis is often not rejected. For example, standard errors nearly double, relative to the case where returns are i.i.d., when return autocorrelation is 5% at a one-quarter lag and decays with a half-life of five years.

We also show that the commonly-used Newey and West (1987) procedure adjusts standard errors too little when return data are generated by our model. For the intermediate

levels of persistence discussed above, the under-adjustment is substantial. This is because the number of lags that capture these persistent autocorrelation patterns is too large a fraction of the length of the historical sample for the asymptotic results in Newey and West (1987) to hold.

In addition to adjusting standard errors using assumed parameters, we use maximum likelihood to estimate the full set of model parameters, including those that determine the magnitude of persistent variation in returns. The maximum likelihood estimates allow us to test competing hypotheses about expected returns. Consistent with the highly noisy coefficient estimates in our reduced-form regressions, we find that the magnitude of persistent variation is imprecisely estimated. Hypothesis tests fail to reject a wide variety of plausible explanations for historical returns of the value, investment, and profitability portfolios, ranging from no persistence to highly persistent but ultimately mean-reverting expected returns.[4] The exception is the size portfolio, which exhibits strong autocorrelation in returns that decay with a half-life of one to two years. Due to these return patterns, the i.i.d. hypothesis – and thus the CAPM benchmark – is strongly rejected for the size portfolio, despite its near-zero average historical performance. Traditional tests, which focus solely on average returns, may thus generate false negatives by ignoring significant variation in conditional expected returns, as well as false positives about unconditional expected returns.

The results discussed so far, which obtain with frequentist estimation methods, suggest that the historical return data on characteristic-sorted portfolios generally provide imprecise inferences about the magnitude of persistent variation in expected returns. As Kandel and Stambaugh (1996) and others have pointed out, however, investors may incorporate information from data to generate forecasts and choose portfolios even if estimates from data are not statistically significant. This point naturally leads to a question that is inherently Bayesian: how do investors with different prior beliefs incorporate historical data into

[4]Similarly, the evidence against the null hypothesis of zero unconditional expected returns becomes weaker when expected returns are allowed to be time varying. With time-variation, p-values for the null of zero unconditional expected returns are 10.4%, 1.2%, and 1.4% for the value, investment, and profitability portfolios, respectively. Without time-variation, p-values are 5.5%, 0.1%, and 0.1%.

their forecasts? To answer this question, we study the Bayesian inference problem of an investor who combines prior beliefs about model parameters with the observed return data on characteristic-sorted portfolios. We analyze a number of economically plausible prior beliefs. For example, an investor guided by predictions of the CAPM may have the prior belief that the expected returns of (market-neutral) characteristic-sorted portfolios cannot deviate substantially from zero. In contrast, an investor with less confidence in the CAPM may allow for larger and more persistent fluctuations in expected returns.

Our Bayesian analysis shows that prior beliefs about persistence substantially affect investors' posterior beliefs about unconditional expected returns. If investors have strong priors that expected returns fluctuate very little, then their posterior beliefs about unconditional expected returns tend to be relatively tight and not highly sensitive to priors about unconditional expected returns. However, if investors' priors put more weight on the possibility of persistent fluctuations, their posteriors about unconditional expected returns become more diffuse and more sensitive to their priors. Thus, investors learn less from data and rely more on their priors when they expect returns to exhibit more persistence.

We also use our Bayesian approach to estimate conditional expected returns at each point in time in our sample. We find that Bayesian investors with relatively agnostic priors about persistence (e.g., beliefs that allow for both i.i.d. and persistent processes) estimate fluctuating conditional expected returns that can differ substantially from unconditional expected returns. For example, because the value portfolio performed particularly poorly towards the end of the sample, the conditional value premium in 2022, measured as the mean of the posterior, is around a quarter of the unconditional value premium, and is close to zero in many specifications. The profitability portfolio exhibits the opposite pattern: because returns were stronger in recent decades, the posteriors for the conditional expected return in 2022 are higher than the posteriors for the unconditional expected return. Also of interest are the estimates for the size portfolio. Despite its near-zero average performance, the size portfolio exhibits the largest fluctuations in conditional expected returns among the four

portfolios we analyze, with conditional Sharpe ratios ranging between -0.75 and one.

Our final analysis contributes to the ongoing debate about what Cochrane (2011) termed the 'factor zoo' by applying our approach to a broad set of 174 characteristic-sorted portfolios described in Chen and Zimmermann (2021). We find that admitting the possibility of time-variation in conditional expected returns significantly reduces the number of these portfolios that exhibit strong evidence against zero unconditional expected returns. Specifically, only 10 (52) of the 174 portfolio have p-values below 0.1% (1%) using our maximum likelihood estimator, compared to 58 (81) using the standard OLS approach. Failure to meet these lower p-value thresholds is particularly relevant since p-values between 1% and 5% can arise because of data mining (Harvey, Liu, and Zhu, 2016).

We also use the broader set of 174 portfolios to show that out-of-sample (OOS) forecasts for future returns based on our statistical model outperform a variety of benchmarks. To do so, we repeat the Bayesian estimation procedure at the end of each calendar year using only prior observations to form posterior beliefs and compute forecasts of average returns in the next year. We find that OOS forecasts from our model strongly predict future realized returns, and that these forecasts subsume the information from average past returns.

Our analysis builds on a well-established literature that analyzes time-variation in expected returns within the context of the aggregate market portfolio. For example, Ferson, Sarkissian, and Simin (2003) consider the predictability of aggregate market returns using predictor variables such as price/dividend ratios. They present simulations that show that OLS regressions can overstate the significance of such relationships in finite samples when expected returns are persistent, even when the Newey and West (1987) standard errors adjustment is used. Although our application is different, our framework is similar and we also show that the Newey and West (1987) correction is not effective in dealing with the problem. In addition, we explore frequentist remedies, such as OLS standard error corrections when the level of persistence in expected returns is known, and maximum likelihood when the level of persistence is estimated.

Pástor and Stambaugh (2009, 2012) and Avramov, Cederburg, and Lučivjanská (2018) also conduct Bayesian analyses of time-varying expected returns and, like us, find that the priors about the return generating process substantially affect the posteriors about expected returns. Our study differs from these papers in both application and focus. Specifically, our analysis concerns characteristic-sorted portfolios rather than the aggregate market portfolio, and we put greater emphasis on prior beliefs about persistence.[5]

We are aware of only two studies that use Bayesian methods to study time variation in characteristic-sorted portfolio returns: Pástor (2000) and Smith and Timmermann (2022). In contrast to our analysis, persistence plays no role in Pástor (2000) as expected returns are assumed to be constant. Smith and Timmermann (2022) assume that there are occasional structural breaks where return premia of all characteristic-sorted portfolios, as well as the market factor, change at the same time and then remain constant until the next structural break occurs. Our approach complements Smith and Timmermann (2022) by examining continuous variations in return premia that are specific to an individual characteristic-sorted portfolio. Continuous variations permit standard error corrections for in-sample inference as well as OOS forecasting – in contrast, structural breaks can only be identified in hindsight.

Finally, our analysis is related to the growing literature on factor momentum (see, for instance, Lewellen (2002), McLean and Pontiff (2016), Avramov et al. (2017), Gupta and Kelly (2019), Arnott et al. (2021a), and Ehsani and Linnainmaa (2022)). Studies in this literature examine dynamic trading strategies involving several characteristic-sorted portfolios that are designed to exploit relatively short-lived variations in these portfolios' conditional expected returns. In contrast, our primary focus is longer-lived fluctuations in portfolio performance. From a methodological perspective, we contribute to this literature by estimating a model of persistent variation in returns.

The remainder of the paper is organized as follows. Section 1 describes our statis-

[5]Other papers in the literature that employ Bayesian methods to study aggregate market returns include Kandel and Stambaugh (1996), Barberis (2000), Wachter and Warusawitharana (2009), and Johannes, Korteweg, and Polson (2014).

tical model of the return-generating process. Section 2 documents the historical return performance and reports the results of reduced-form predictive regressions for the four characteristic-sorted portfolios we study. Section 3 presents the frequentist estimations of the model. Section 4 presents the Bayesian analysis. Section 5 provides a summary set of results for the broader set of 174 characteristic-sorted portfolios. Section 6 concludes.

1. Statistical Model

Our empirical analyses of characteristic-sorted portfolios apply a statistical model in which conditional expected returns exhibit persistent fluctuations. In this section, we first provide a brief discussion of the theoretical motivation for our model. We then present the model specification and describe the return patterns that the model generates.

1.1. Model Motivation

In traditional asset pricing tests, such as tests of the CAPM, the null hypothesis is that expected excess returns are *always* zero – that is, at each point in time – not just on average. Thus, under the null, returns are serially uncorrelated. As we discuss in the Introduction, these tests, which appropriately assume independent residuals, convincingly reject the CAPM.

The traditional tests provide less guidance, however, if we want to distinguish between various alternatives to the CAPM null. For instance, leading behavioral explanations of the value premium are based on the idea that investor overconfidence (Daniel, Hirsleifer, and Subrahmanyam, 1998) or optimism in evaluating new technologies (Shiller, 2000) cause growth stocks to be overpriced. However, as modeled in Altı and Titman (2019), behavioral biases can also generate persistent cycles whereby the value effect is positive in some episodes and negative in others. More generally, characteristic-sorted portfolio returns may exhibit both unconditional and conditional deviations from the CAPM benchmark.[6]

[6]Rational theories of characteristic-based return predictability can also generate persistent fluctuations in expected returns. For instance, Gârleanu, Kogan, and Panageas (2012) and Kogan, Papanikolaou, and Stoff-

The statistical model described below allows for not just deviations in unconditional expected returns from the CAPM benchmark, but also allows for time-variation in expected returns. Estimates of the model thus capture a broader set of alternatives to the CAPM.

1.2. Model Specification

We assume that the time-series of the zero-cost portfolio returns r_t satisfies:

$$r_{t+1} = \mu_t + \epsilon_{t+1}, \tag{1}$$

$$\mu_{t+1} = \mu + \lambda(\mu_t - \mu + \delta_{t+1}), \tag{2}$$

where μ_t and μ are the conditional and unconditional expected returns, respectively. The shocks to unexpected returns ϵ_t and the shocks to expected returns δ_t are i.i.d. and follow a joint normal distribution with variances σ_ϵ and σ_δ, respectively, and correlation $\rho \in (-1, 1)$.[7] We expect ρ to be negative, since positive shocks to expected returns, ceteris paribus, reduce an investment's value.[8]

The parameter $\lambda \geq 0$ in Equation (2) determines the persistence of shocks to μ_t. To facilitate interpretation of economic magnitudes, we express λ in our empirical analyses in terms of the annualized half-life of shocks to expected returns, H:

$$H = \frac{\log(0.5)}{\log(\lambda)} \frac{1}{N}, \tag{3}$$

where N is the number of periods per year (e.g. four for quarterly data).

The econometrician does not observe μ_t, but can estimate it – along with other model parameters – from the observed return realizations $\mathbf{R} = [r_1, r_2, \ldots, r_T]'$. Given parameters

man (2020) provide theories where rational investors hold lower-returning growth stocks to hedge technology shocks. One might expect the economic fundamentals that drive these rational explanations to fluctuate over time, causing time-variation in the value premium.

[7] Conrad and Kaul (1988) use a similar specification but assume $\rho = 0$.

[8] In Equation (2) we multiply the shock δ_{t+1} by λ so that returns are i.i.d. when $\lambda = 0$. With δ_{t+1} outside the parenthesis and $\lambda = 0$, $cov(r_{i,t}, r_{i,t-1}) = \rho\sigma_\delta^2$, meaning that one would also need to assume $\rho = 0$ for returns to be independent across time.

$\Omega = [\mu, \lambda, \sigma_\epsilon, \sigma_\delta, \rho,]$, $\mathbf{R}$ has the following mean and covariance matrix:

$$\mathbb{E}(\mathbf{R}|\Omega) = \mu, \tag{4}$$

$$\text{Cov}\,(\mathbf{R}|\Omega) = \mathbf{\Sigma}(\Omega), \qquad \mathbf{\Sigma}(\Omega)_{i,j} = \begin{cases} \frac{\lambda^2\sigma_\delta^2}{1-\lambda^2} + \sigma_\epsilon^2 & \text{if } i = j \\ \lambda^{|i-j|}\left(\frac{\lambda^2\sigma_\delta^2}{1-\lambda^2} + \rho\sigma_\delta\sigma_\epsilon\right) & \text{if } i \neq j \end{cases}. \tag{5}$$

Equation (5) shows that shocks to both the expected and unexpected returns contribute to the volatility of returns (the terms σ_δ^2 and σ_ϵ^2, respectively). Note also that the covariance between r_i and r_j decays at a constant rate λ as $|i - j|$ grows.

1.3. Identification

The model is over-parameterized in the sense that multiple values of Ω lead to the same predicted moments $\mathbb{E}(\mathbf{R}|\Omega) = \mu$ and $\text{Cov}(\mathbf{R}|\Omega) = \mathbf{\Sigma}(\Omega)$. To see this, consider the variance and one-lag autocorrelation of returns:

$$\sigma_r^2(\Omega) = \text{Var}(r_t) = \frac{\lambda^2\sigma_\delta^2}{1-\lambda^2} + \sigma_\epsilon^2, \tag{6}$$

$$\gamma(\Omega) = \text{Corr}(r_{t+1}, r_t) = \lambda\frac{\lambda^2\sigma_\delta^2 + (1-\lambda^2)\rho\sigma_\delta\sigma_\epsilon}{\lambda^2\sigma_\delta^2 + (1-\lambda^2)\sigma_\epsilon^2}. \tag{7}$$

Using this alternative notation, the covariance matrix becomes:

$$\mathbf{\Sigma}(\Omega)_{i,j} = \begin{cases} \sigma_r^2 & \text{if } i = j, \\ \lambda^{|i-j|-1}\gamma\sigma_r^2 & \text{if } i \neq j. \end{cases} \tag{8}$$

Inspecting Equation (8), we see that any two parameterizations Ω and $\tilde{\Omega}$ which yield the same λ, σ_r, and γ will result in the same covariance matrix $\mathbf{\Sigma}$ for returns.

To better understand identification in our model, note that we can use the sample mean of returns to estimate the unconditional expected return μ (Equation (4)), and the average rate

at which the covariance between r_i and r_j decays as $|i-j|$ grows to estimate the persistence parameter λ (Equation (5)). The identification problem arises because we have three other model parameters to be identified (σ_ϵ, σ_δ, and ρ), but only two other moments that can be estimated: the variance of returns σ_r^2 in Equation (6) and the one-lag autocorrelation γ in Equation (7).[9] Intuitively, the identification problem arises because one cannot distinguish between different channels that generate return variance and autocorrelation. An increase in the volatility of expected return shocks σ_δ increases both return variance and autocorrelation, but the same increases can also be generated from increases in the volatility of unexpected return shocks σ_ϵ and the correlation parameter ρ.

We address this identification problem in our frequentist analysis by estimating the four moments $\theta = [\mu, \lambda, \sigma_r, \gamma]$, which we can identify, rather than the full set of underlying parameters Ω. In doing so, we apply the constraint that there must be a parameterization Ω which is consistent with θ and satisfies $\sigma_\epsilon > 0$, $\sigma_\delta > 0$, $\lambda \geq 0$, and $\rho \in (-1, 1)$. The identification problem does not arise in our Bayesian analysis because we compute a posterior distribution for parameter values, which is unique given a set of priors and observed data, rather than a single point estimate, which is not unique.

1.4. The Sign of Return Autocorrelations

Although our model can generate both positive and negative return autocorrelations, negative autocorrelations of material magnitude tend to occur only when expected returns exhibit very little persistence. Time variation in expected returns is a source of positive autocorrelation because current and recent past returns have similar conditional expectations. A negative return autocorrelation requires a negative correlation between realized returns and *changes* in expected returns. However, unless shocks to expected returns are quickly mean-reverting, the information in past returns about a single period's change in expected return is less relevant than the information about the ongoing level of expected returns.

[9] Note from Equation (8) that return covariances at longer lags do not provide any additional information about the model parameters, because all these covariance terms are scaled by $\gamma\sigma_r^2$.

We can see this formally using Equation (7), which specifies the first-order return autocorrelation γ as a function of λ, ρ, σ_ϵ, and σ_δ. The first term in the numerator, $\lambda^2\sigma_\delta^2$, represents the effect of persistent expected return variations and is always positive. The second term in the numerator, $(1-\lambda^2)\rho\sigma_\delta\sigma_\epsilon$, represents the effect of the correlation between shocks to expected and unexpected returns and is negative when $\rho < 0$. Comparing these two terms, we see that the sign of γ is determined by the relative magnitudes of λ and $|\rho\sigma_\epsilon/\sigma_\delta|$. In particular, when λ is close to one, as is the case for persistent variations that motivate our analysis, $\gamma < 0$ only if $\sigma_\epsilon/\sigma_\delta$ is sufficiently large – that is, if shocks to expected returns exhibit very little variation relative to shocks to unexpected returns.

In Appendix Table 1, we quantify how small λ (or, equivalently, H) needs to be for γ to be negative. We assume $\rho = -1$ to provide an upper bound on the prevalence of negative autocorrelations, and consider various combinations of H and $\sigma_\epsilon/\sigma_\delta$. We find that materially negative γ values obtain only when H is half a year or less.

2. Characteristic-Sorted Portfolios

We apply our statistical model to study the portfolios that are formed by sorting stocks based on value, investment, profitability, and size. We focus on these four characteristic-sorted portfolios because they are the basis of the Fama and French (2015) five-factor model and have been extensively studied elsewhere. In Section 5 we extend our analysis to a broader set of 174 portfolios that have also been analyzed in previous literature.

2.1. Data and Characteristic Definitions

We use data on the historical returns of characteristic-sorted portfolios from Ken French's website.[10] Each portfolio combines a long position in a value-weighted portfolio of firms in one extreme quintile of the characteristic with a short position in the other extreme.

The characteristics are defined following Fama and French (2015). *Value* is the ratio of

[10] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

the book value of equity ($B_{i,y}$) to the market value of equity ($M_{i,y}$) as of the end of the prior fiscal year y. *Investment* is the growth rate in the book value of assets ($\text{Assets}_{i,y}/\text{Assets}_{i,y-1}$). *Profitability* is revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses in year y divided by book equity in year $y-1$. *Size* is $M_{i,y}$.

In contrast to most of the literature, which examines the monthly returns of characteristic sorted portfolios, we analyze quarterly returns. As we show in Appendix A, the monthly returns of three of the four portfolios we study exhibit strong positive first-order autocorrelations.[11] While these short-term autocorrelations are consistent with time-varying expected returns, they could also be driven by lead-lag effects and other short-term microstructure effects. To focus our analysis on longer-term autocorrelations driven by persistent variations in expected returns, we estimate our model using quarterly returns.

Since the market risk premium is not our focus, we use market-neutral versions of each portfolio's returns throughout, calculated as:

$$r_{i,t}^{\beta=0} = r_{i,t} - \hat{\beta}_i(r_{m,t} - r_{f,t}), \tag{9}$$

where $r_{i,t}$ is the quarterly return of the long-short portfolio i, $r_{m,t}-r_{f,t}$ is the quarterly excess market return, and $\hat{\beta}_i$ is the full-sample market beta. The average return and Sharpe ratio of this market-neutral portfolio are equivalent to the alpha and information ratio, respectively, of the underlying long-short portfolio.

2.2. Historical Performance of Characteristic-Sorted Portfolios

Table 1 summarizes the historical performance of the characteristic-sorted portfolios. The value, investment, and profitability portfolios have annualized mean returns above 4% and Sharpe ratios between 0.3 and 0.5, while the size portfolio's mean return is close to zero. Based on standard errors calculated under the assumption of i.i.d. returns, we strongly reject

[11] This is consistent with the evidence in Gupta and Kelly (2019) that 47 of 65 characteristic-based portfolios have significantly positive first-order autocorrelation.

the null hypothesis of zero expected returns for the first three portfolios.[12]

We also examine the historical performance of the portfolios in the first and second halves of our sample period, 1963–1992 and 1993–2021. As discussed in McLean and Pontiff (2016), Linnainmaa and Roberts (2018), and Fama and French (2021), the value strategy's average returns are smaller in the second half of the sample and are not statistically different from zero. However, as emphasized by Fama and French (2021), the difference between the two halves of the sample is not statistically significant.[13]

The investment, profitability, and size portfolios each show different return patterns across subsamples. The investment portfolio's returns are largely consistent over time and statistically significant in both halves of the sample. The profitability portfolio follows the opposite pattern as the value portfolio, performing better in the second half of the sample than the first, though again the difference is statistically insignificant. The size portfolio has small and statistically insignificant returns in both halves of the sample.[14]

2.3. Predictive Regressions with Model-Simulated and Historical Data

As discussed in the Introduction, a number of studies document short-term autocorrelation in characteristic-sorted portfolio returns, or more generally analyze portfolio timing strategies that are premised upon time-variation in expected returns. Before we estimate our model, we present similar reduced-form regression analyses using both model-simulated return data and historical returns of the four characteristic-sorted portfolios that we analyze.

Estimates of the autocorrelation structure for individual portfolios are inherently imprecise because realized returns are quite volatile relative to the plausible variations in expected returns. Furthermore, autocorrelation estimates have a well-document downward bias in

[12]We compute the i.i.d. standard errors by taking the standard deviation across simulated samples formed by re-sampling historical data with replacement.

[13]Recent studies provide two potential explanations for the decline of the value premium: book value may have worsened as a proxy for the value of assets in place (Choi, So, and Wang, 2021; Eisfeldt, Kim, and Papanikolaou, 2022; Goncalves and Leonard, 2023), or a series of shocks may have widened the difference in multiples between growth and value stocks (Israel, Laursen, and Richardson, 2020; Arnott et al., 2021b). Both imply time variation in expected returns and thus are consistent with our model.

[14]The absence of a significant size effect in the early part of the sample is in part due to our use of value weights in portfolio construction.

small samples (Kendall, 1954; Marriott and Pope, 1954). To illustrate these issues, we first present autocorrelation estimates that we generate with model-simulated data samples under a variety of assumptions about H and γ. In these simulations, we fix the model parameters and generate 50,000 samples with 234 observations, which is the number of quarters in our empirical sample.[15] For each simulated return series, we estimate return autocorrelations using regressions of quarterly returns on averages of returns over the previous L quarters:

$$r_t = a + b_L \left(\frac{1}{L} \sum_{l=1}^{L} r_{t-l} \right) + \epsilon_t. \tag{10}$$

Panel A of Table 2 presents average values, as well as 95% confidence intervals, for the regression coefficient $\hat{b}_L$ across samples that are simulated under a variety of parametric assumptions. Specifically, we consider (H, γ) combinations where the half-life of shocks H equals 2.5, five, and ten years, and the first-order return autocorrelation γ equals 2.5%, 5%, and 10%. For each parameterization we also compute $\hat{\sigma}_{sr}$, which denotes the volatility of the annualized Sharpe ratio conditional on past returns. More specifically, $\hat{\sigma}_{sr}$ is the volatility of the conditional Sharpe ratio computed by an investor who knows the model parameters and observes an infinite history of past return realizations, but not the realizations of the conditional expected returns μ_t. As Table 2 shows, the volatilities of conditional Sharpe ratios generated by these parameterizations are roughly comparable to the historical Sharpe ratios of characteristic-sorted portfolios reported in Table 1.

The first row of Panel A in Table 2 reports the autocorrelation estimates when $H = 0$ and thus returns are distributed i.i.d. As expected, the average estimated autocorrelation coefficients are negative due to the aforementioned downward bias: because the in-sample mean is used to calculate the autocorrelation coefficient, the data appears to be mean-reverting even when there is no true mean reversion. The bias is stronger the longer the past

[15] The remaining model parameters we use for the simulations are $\mu = 0$ and $\sigma_r = 7.57\%$, which is the full-sample return standard deviation for the value portfolio. These choices have no effect on the results we report in this section as autocorrelation estimates from regressions with intercepts are invariant to linear transformations of r_t.

return window because there is a smaller effective sample size.

The other rows in Panel A of Table 2 show autocorrelation estimates when expected returns exhibit persistent variation, i.e., $H > 0$ and $\gamma > 0$. Two observations emerge from the reported estimates. First, the downward bias continues to affect autocorrelation estimates even when expected returns are persistent, especially with longer past return windows. Second and more importantly, the confidence intervals for the autocorrelation estimates are quite wide and include negative values in every parameterization – even those with large γ.

Panel B of Table 2 presents estimates of Equation (10) for the historical samples of the quarterly returns of the value, investment, profitability, and size portfolios. The first column shows that all four portfolios have positive b_4, indicating that past-year returns positively predict next-quarter returns. The second and third columns show that longer past-return windows have point estimates with varying signs.

We do not report standard errors or bias corrections in Panel B of Table 2 because Panel A shows that both depend heavily on the magnitude and persistence of the variations in expected returns. In most cases, the confidence intervals for different parameterizations in Panel A include all of the point estimates in Panel B, which means that one cannot reject any of the posited autocorrelation structures. Even economically large regression coefficients, such as $\hat{b}_{20} = -0.30$ for the investment portfolio, lie in the 95% confidence interval for every parameterization from i.i.d. to $H = 10$ and $\gamma = 5\%$.

For three out of the four characteristic-sorted portfolios, the autocorrelation estimates do not reject any reasonable parameterization of our model. The size portfolio is the exception; we convincingly reject the i.i.d. null in the regression with both the one-year and five-year lags. In addition, in a pooled regression that combines the returns of the four portfolios, the i.i.d. null hypothesis is rejected for one-year and five-year returns with p-values of 0.0% and 2.2%, respectively.[16] In a pooled regression with only value, investment, and profitability, we reject the i.i.d null with the one year lag with a p-value of 2.8%. The reason we can

[16] We conduct this test comparing the sum of the individual estimates from observed data to the distribution of this sum in samples simulated under the i.i.d. assumption.

jointly but not individually reject the null is that the coefficients for the value, investment, and profitability portfolios are each above the i.i.d. benchmark, but not by enough to reject on an individual basis.

3. OLS and Maximum Likelihood Estimations

As the previous section shows, the autocorrelation patterns observed in the historical returns of characteristic-sorted portfolios are consistent with a range of assumptions about the magnitude of persistent variation in conditional expected returns. In this section, we formally analyze the impact of such variation within the context of our model using OLS regressions and maximum likelihood estimations.

3.1. OLS Estimation With Model-Corrected Standard Errors

We start by estimating the unconditional expected return μ and its standard error using OLS regressions of observed returns r_t on a constant. The OLS estimate $\hat{\mu}^{OLS}$ is a consistent estimator of μ, even when conditional expected returns vary, as in our model. The correct standard errors for $\hat{\mu}^{OLS}$ depend on the covariance matrix of the residuals $\psi_t = r_t - \mu$.[17]

The typical approaches used to adjust standard errors are the White (1980) correction for potential heteroskedasticity and Newey and West (1987), which corrects for both heteroskedasticity and autocorrelations up to a small number of lags. When expected returns are time-varying and persistent, both of the standard approaches produce understated standard errors. The reason is that persistent variations in expected returns generate small but long-lasting correlations in ψ_t that extend beyond the windows considered by Newey and West (1987). Formally, our model generates the residuals

$$\psi_t = \mu_{t-1} - \mu + \epsilon_t, \tag{11}$$

[17] We use 'correct standard errors' as an informal shorthand for the correct specification of the asymptotic distribution of $\hat{\mu}$.

which implies that ψ_t has long-lasting autocorrelation due to persistent variations in μ_t. As we show below, even if we extend the number of lags in Newey and West (1987) to match or exceed the half-lives of shocks to expected returns, standard errors remain underestimated because the number of lags is too large a fraction of the observed time series for the asymptotic results in Newey and West (1987) to hold.

If the first-order return autocorrelation γ and the half-life of shocks H are known, we can correct the OLS standard errors for the resulting autocorrelation in the residuals ψ_t using the structure of our model.[18] Specifically, the standard error of $\hat{\mu}^{OLS}$ in this case is

$$\begin{aligned} \text{SE}(\hat{\mu}^{OLS}) =& \sqrt{\frac{\mathbf{1'\Sigma 1}}{T^2}} \\ =& \left(\frac{\sigma_r}{\sqrt{T}}\right)\sqrt{1+2\gamma\left[\frac{\lambda^T + T(1-\lambda) - 1}{T(1-\lambda)^2}\right]}. \end{aligned} \tag{12}$$

In Equation (12), T is the number of observations, $\mathbf{1}$ is a $T \times 1$ vector of ones, and $\mathbf{\Sigma}$ is the covariance matrix of returns. The first line is the general formula for computing standard errors with a known covariance matrix (see Section 4.5 of Cameron and Trivedi (2005)). In our model, $\mathbf{\Sigma}$ is fully specified by γ and λ, as shown in Equation (8). Substituting from Equation (8) results in the formula in the second line of Equation (12) after some algebraic manipulation. Note that the first term in this formula, $\sigma_r/\sqrt{T}$, is the unadjusted OLS standard error. Thus, the second term in square roots is the adjustment factor, which is a function of T, γ, and λ. For large T, Equation (12) approximates to:

$$\text{SE}(\hat{\mu}^{OLS}) \approx \left(\frac{\sigma_r}{\sqrt{T}}\right)\sqrt{1+\frac{2\gamma}{1-\lambda}}. \tag{13}$$

The standard error adjustments in Equations (12) and (13) can also be stated in terms of the annualized half-life H instead of λ by substituting $\lambda = 0.5^{\frac{1}{HN}}$, which follows from Equation

[18]Note that the model structure can also be used to obtain Generalized Least Squares (GLS) estimates. As detailed in Appendix C, we find that the GLS estimates differ only slightly from the OLS estimates for the four characteristic-sorted portfolios we study.

(3), where N is the number of return observation periods per year.

Table 3 shows how different assumptions about γ and H affect the standard errors of unconditional expected return estimates. For comparison, we also report standard errors that are adjusted using the Newey and West (1987) procedure with the number of lags matching the half-life of shocks H. For the value, investment, and profitability portfolios, we find that Newey-West standard errors differ very little from unadjusted standard errors.[19] Newey-West standard errors are around 50% larger for the size portfolio, which is a reflection of the stronger evidence of persistence for the size portfolio presented in Table 2.

In contrast to the Newey-West adjustment, model-implied standard errors that are computed using Equation (12) tend to be larger for all four portfolios. The magnitude of this adjustment depends on both H and γ. The adjustments to standard errors are relatively small when expected return shocks are assumed to be quickly mean-reverting ($H = 0.5$ years). With more persistent shocks (H between 2.5 and 10 years), however, Table 3 shows that model-implied standard errors are much larger. For instance, standard errors approximately double relative to their unadjusted counterparts when when $H = 5$ years and $\gamma = 5\%$. Overall, the results in Table 3 indicate that accounting for plausible magnitudes of persistent variation in expected returns results in inferences about unconditional expected returns that are materially less precise than inferences under the assumption of i.i.d. returns.

3.2. Maximum Likelihood Tests

This subsection describes maximum likelihood estimates of our model. In contrast to the ordinary least squares regressions described in the previous subsection, maximum likelihood estimation requires distributional assumptions, but allows us to estimate all of the model parameters. In particular, we estimate, rather than assume, the model parameters that determine the magnitude of persistent variations in expected returns.

[19] In Appendix B, we show that Newey and West (1987) standard errors are downward-biased in samples simulated using our statistical model. We also show that this bias is primarily due to the downward small-sample bias in autocorrelation estimates.

3.2.1. Hypothesis tests for μ

We test the hypothesis that the unconditional expected return μ equals zero under a variety of assumptions about the structure of time-variation in conditional expected returns. For each assumption, we estimate the model using maximum likelihood twice; first with no restrictions on μ and then under the restriction that $\mu = 0$. In each model estimation we restrict H to be less than or equal to 10 years because there is no statistical power to detect values of H that are large relative to the length of the sample period. Using these estimates, we compute the p-value for the $\mu = 0$ null using a likelihood ratio test.

The first three columns in Panel A of Table 4 test the hypothesis of $\mu = 0$ using model estimates that assume no time variation in expected returns (i.e., $H = 0$ and thus returns are i.i.d.). Testing $\mu = 0$ under this assumption is analogous to using OLS with no standard error correction. Not surprisingly, the likelihood-ratio p-values strongly reject the hypothesis for the investment and profitability portfolios, and reject, with a p-value of 5.5%, for the value portfolio.

The next five columns in Panel A of Table 4 relax the i.i.d. assumption and estimate the values of H and γ that maximize the likelihood of observing the historical data.[20] Imposing the restriction $\mu = 0$ in this case results in economically large estimates of H and/or γ. These parameter estimates explain the historical data as arising from relatively high realizations of *conditional* expected returns that reverse over time. Given that this explanation is plausible, the p-values for the hypothesis that $\mu = 0$ is somewhat higher than in the i.i.d. case, especially for the value, investment, and profitability portfolios. For instance, while the likelihood ratio tests still reject the hypothesis that $\mu = 0$ for the investment and profitability portfolios, the p-values of 1.2% and 1.4% respectively are an order of magnitude larger than the p-value of 0.1% in the i.i.d. case.

[20] As discussed in Section 1.3, we estimate $\theta = [\mu, H, \sigma_r, \gamma]$ rather than the underlying parameters $\Omega = [\mu, \lambda, \sigma_\epsilon, \sigma_\delta, \rho,]$ because the latter are not fully identified. We restrict the admissible values of θ to the range for which there exists at least one possible Ω yielding the same covariance matrix of returns as θ.

3.2.2. Hypothesis tests for H and γ

Next, we use maximum likelihood estimates to assess the plausibility of the various assumptions about H and γ that we examined in our OLS analyses in Section 3. For each assumption, we re-estimate the model by restricting H and γ to their assumed values, and calculate likelihood ratios relative to the unrestricted model to test whether we can reject the hypothesized restriction. We also use likelihood ratios to test whether we can reject the hypothesis that $\mu = 0$ given the restrictions on H and γ.

Panel B of Table 4 presents the results. For the value, investment, and profitability portfolios, we cannot reject *any* of the hypotheses on time-variation in expected returns at the 5% level, including no variation (i.e., i.i.d. returns) as well as large and persistent variation ($H = 10$ years, $\gamma = 5\%$). Similar to the results reported in Table 3, Panel B of Table 4 also shows that these alternative assumptions have a material impact on our inferences about $\mu = 0$. For instance, p-values for the hypothesis that $\mu = 0$ for the investment and profitability portfolios increase from 0.1% in the i.i.d. case to 9.8% and 5.1%, respectively, when $H = 5$ years and $\gamma = 5\%$.

While the results for the value, investment, and profitability portfolios suggest that the historical data offer little guidance about time-variation in expected returns, the results for the size portfolio are quite different. Consistent with the reduced-form evidence in Table 2, the estimates for the size portfolio in Table 4 point to strong positive autocorrelation in returns that decay with a half-life between one and two years. As a result, the maximum-likelihood tests strongly reject both the i.i.d. and $H = 10$ hypotheses. Importantly, the rejection of the i.i.d. hypothesis implies that the size portfolio returns deviate significantly from the CAPM benchmark. In contrast, traditional asset pricing tests, which solely focus on average returns and ignore time variation, fail to reject the CAPM for the size portfolio, since the portfolio's historical average performance is close to zero (see Table 1). Thus, by ignoring time variation in expected returns, traditional tests may generate not only false positives (rejecting $\mu = 0$ when in fact returns exhibit persistent variation around zero means), but

also false negatives (failing to reject the CAPM) too often.

3.2.3. Sample Size and Test Power

The results in Table 4 indicate that inferences from historical data about H and γ are generally quite imprecise; furthermore, the precision of inferences about μ crucially depend on H and γ. A natural question then is how large the sample of returns needs to be to obtain more precise parameter estimates. This subsection presents simulations that address this question.

Specifically, we simulate random samples of varying lengths for two different parameterizations of our model: one where returns are i.i.d., and the other a persistent process where $H = 5$ years and $\gamma = 5\%$. In both parameterizations we assume that the unconditional expected return $\mu = 5\%$ and the annualized return volatility $\sigma_r = 15\%$. For each simulated sample, we estimate the model with maximum likelihood and test the hypothesis that $\mu = 0$. We also test the 'opposite' hypothesis about persistence (i.e., the hypothesis that $H = 5$ years and $\gamma = 5\%$ when the true data generating process is i.i.d., and the i.i.d. hypothesis when the true data generating process has $H = 5$ and $\gamma = 5\%$). In these hypothesis tests using likelihood ratios, the econometrician does not know the true data generating process and has to estimate the model exactly as we do in Table 4. The results are summarized in Figure 1, which plots the fraction of simulated samples in which the tested hypothesis is rejected with a p-value of 5% or less.

Panel A of Figure 1 shows the rejection rates for the $\mu = 0$ hypothesis. With 50-year samples, which is close to the length of the historical sample, the rejection rate for $\mu = 0$ is around 70% when the true data generating process is i.i.d. Thus, there is a 30% chance that one would fail to reject $\mu = 0$ with 50 years of data despite the fact that $\mu = 5\%$. When the return data are generated by the persistent process, the failure to reject $\mu = 0$ becomes even more acute, with only about 37% of samples rejecting. Longer samples allow i.i.d. samples to converge to the true values relatively more quickly: with 200 years, more than 95% of the

samples reject $\mu = 0$. In contrast, reaching a similar rejection rate requires 800 years when returns are generated by the persistent process.

Panel B of Figure 1 shows the rejection rates for hypotheses about the magnitude of persistent variation in returns. The i.i.d. and the persistent return processes that generate the simulated data are quite different from each other in terms of the economic interpretation of the model. Yet there is little power to distinguish between them with 50 years of data: only about 40% (20%) of the samples reject the persistent (i.i.d.) process when returns are in fact i.i.d. (persistent). As the figure shows, reliably distinguishing between these alternative return generating processes requires at least 400 years of data.

4. Bayesian Analysis

Our results in the previous section suggest that given the possibility of persistent variation in expected returns, reaching precise conclusions about model parameters may require a longer time series than is available. This does not imply, however, that data from shorter samples are uninformative. As Kandel and Stambaugh (1996) and others have observed, Bayesian investors will in fact use data, in combination with their priors, to inform their beliefs about the return generating process even if statistical tests of data do not reject null hypotheses at conventional significance levels. Therefore, a natural next step for our analysis is to ask how investors with different prior beliefs interpret historical data, and how this in turn influences their investment decisions. We examine these normative questions in this section. Specifically, we conduct a Bayesian analysis that specifies prior likelihoods of different model parameterizations and uses the observed data to calculate posterior likelihoods. We also compute posteriors for moments that affect investors' portfolio timing decisions, such as the conditional Sharpe ratios at each point in time.

4.1. Prior Beliefs About Model Parameters

To facilitate economic intuition, our Bayesian analysis specifies prior beliefs on a transformation of the model parameters expressed in terms of Sharpe ratios rather than expected returns.[21] Specifically, we specify priors over μ_{sr}, the unconditional Sharpe ratio of portfolio returns; σ_r, the volatility of portfolio returns; H, the half-life of shocks to expected returns; σ_{sr}, the volatility of conditional Sharpe ratios; and ρ, the correlation between unexpected and expected return shocks. The mapping between these annualized parameters in annualized terms and the underlying model parameters $\Omega = [\mu, \lambda, \sigma_\epsilon, \sigma_\delta, \rho,]$ is given by the following equations, where $N = 4$ is the number of periods per year:

$$\mu_{sr} = \frac{\mu}{\sigma_\epsilon}\sqrt{N}, \qquad \sigma_r = \sqrt{\left(\frac{\lambda^2\sigma_\delta^2}{1-\lambda^2} + \sigma_\epsilon^2\right)N}, \tag{14}$$

$$H = \frac{\log(0.5)}{\log(\lambda)}\frac{1}{N}, \qquad \sigma_{sr} = \sqrt{\left(\frac{\lambda^2\sigma_\delta^2/(1-\lambda^2)}{\sigma_\epsilon^2}\right)N}, \qquad \rho = \rho.$$

We consider a variety of priors on μ_{sr} and H, which are summarized in Panel A of Table 5. For μ_{sr}, we first consider normal prior distributions centered at -0.4, 0, and 0.4, all with a standard deviation of 0.4.[22] For the value portfolio, these three investors can be viewed as having growth, neutral, or value inclinations. We also examine an uninformative prior where μ_{sr} is distributed uniformly between -2 and 2.[23]

To illustrate how prior beliefs about H influence Bayesian inferences about expected returns, we consider three dogmatic priors and one agnostic prior. The dogmatic priors assert that $H = 0$ (making returns i.i.d.), $H = 2.5$ years, or $H = 5$ years with certainty. The

[21] Another advantage of this transformation is that the same Sharpe ratio priors can be applied across all characteristic-sorted portfolios regardless of their volatility levels.

[22] We use ± 0.4 as the center for the unconditional Sharpe ratio distributions to roughly match the US equity market's estimated Sharpe ratio.

[23] Uniform distributions over wider supports give nearly-identical results because the data strongly reject unconditional Sharpe ratios above 2 or below -2 for for the portfolios we study.

agnostic prior views H as unknown and uniformly distributed between 0 and 10 years.

We consider uniform priors over wide ranges for the remaining parameters. The prior for the volatility of annual returns, σ_r, is uniformly distributed between 10% and 20%. The prior for the volatility of conditional Sharpe ratios, σ_{sr}, is uniformly distributed between 0 and 1. The correlation between unexpected and expected return shocks, ρ, is likely to be negative given the inverse relation between prices and expected returns. In contrast to aggregate market returns, however, characteristic-sorted portfolio returns should be driven primarily by cash-flow news rather than discount rate news. Based on these observations, we specify the prior on ρ to be uniformly distributed in the interval $[-0.5, 0]$.

To provide intuition for economic magnitudes, Panel B of Table 5 reports summary statistics for a number of moments that are implied by the priors specified in Panel A. We calculate these moments by simulating 50,000 draws from each prior and calculating the value of each moment implied by each parameter draw. The first set of columns shows that μ, the unconditional expected return, has a prior mean that equals -5.3%, zero, or 5.3%, depending on the prior specification. The middle set of columns show that the prior mean values of one-lag return autocorrelation, γ, are positive in all cases, indicating that the positive effect due to persistence of expected returns typically outweighs the negative effect due to $\rho < 0$, although the 95% confidence intervals do include negative values.

The parameter σ_{sr} governs the volatility of the true conditional Sharpe ratio, defined as the conditional expected return μ_t in Equation (2) divided by the volatility of unexpected return shocks σ_ϵ. This Sharpe ratio is unlikely to be perfectly observable to investors in practice. If an investor forecasts expected returns based on past return realizations, combined with his beliefs about the model parameters, then the relevant measure of the time variation in conditional Sharpe ratios is $\hat{\sigma}_{sr}$, defined earlier in Section 2.3. The last set of columns in Panel B in Table 5 report summary statistics for $\hat{\sigma}_{sr}$ for each prior specification we consider. Note that, while the Sharpe ratio conditional on past returns can be quite volatile, it is not as volatile as the unobserved conditional Sharpe ratio. In particular, both the means and

the 95% confidence intervals for $\hat{\sigma}_{sr}$ are about half as large as those for σ_{sr}.

4.2. Posteriors for Unconditional Expected Returns and Sharpe Ratios

We compute posterior distributions given each (H, μ_{sr}) prior specification and each of the four characteristic-based portfolios, resulting in 64 prior-data pairs. For each pair, we characterize the posterior distributions by generating 50,000 posterior draws using the M.C.M.C. procedure detailed in Appendix D. We compute the posterior distributions for the full parameter vector Ω, but for brevity we report the results only for a subset of parameters and moments that are of economic interest.

The results are summarized in Table 6, which reports summary statistics for the posterior distributions of a number of model parameters and moments, and Figure 2, which plots summary statistics for posterior distributions of μ_{sr} in the black lines on the left. As the table and the figure show, different priors about H result in substantially different posterior beliefs about unconditional expected returns μ and Sharpe ratios μ_{sr}. For instance, the 95% confidence intervals for μ and μ_{sr} become wider as priors for H increase, making the possibility of zero or negative unconditional expected returns much more likely. The intuition for this result is the same as in the frequentist analysis above: the data do not strongly reject the possibility that the historical return performance of characteristic-sorted portfolios is explained by persistent but dissipating positive shocks to conditional expected returns.

The Bayesian analysis also produces an insight that is distinct from the frequentist analysis: the extent to which priors about μ_{sr} affect posteriors about μ and μ_{sr} depends on the prior about H. Bayesian investors who believe $H = 0$ have similar posteriors about μ and μ_{sr} despite large differences in priors. On the other hand, Bayesian investors whose priors are that H is, or might be, large have substantial differences in their posterior beliefs about μ and μ_{sr} despite observing 58 years of data. For example, means of posteriors about the value portfolio's μ_{sr} are clustered between 0.19 and 0.27 for the $H = 0$ prior, but vary from 0.11 to 0.27 for the $H \sim U(0, 10)$ prior.

The reason the sensitivity of posteriors to priors depends on H is that Bayesian investors 'shrink' observed in-sample average returns towards the mean of their priors, and the extent of this adjustment depends on H. If an investor believes H equals zero and thus returns are i.i.d., the data are more informative about unconditional expected returns and thus the posterior hews closer to the in-sample average. If the investor believes H is or may be larger than zero, the return data are less informative about unconditional expected returns and so the posterior depends more on the prior.

Table 6 also shows that the posterior distributions for H, γ, and $\hat{\sigma}_{sr}$ for the value, investment, and profitability portfolios differ very little from the corresponding prior distributions reported in Table 5. This finding is consistent with the evidence in Tables 2 and 4 that the data offer little guidance on the autocorrelation patterns of these portfolios. The size portfolio, by contrast, has posterior distributions for H with means well below the prior mean of five years, indicating that the data push Bayesian investors towards lower H. Furthermore, unlike the other three portfolios, the posteriors for size portfolio's γ and $\hat{\sigma}_{sr}$ have much higher means compared to priors and confidence intervals that exclude zero, suggesting that the evidence tilts in favor of positive autocorrelation. Still, posteriors for the size portfolio's H, γ, and $\hat{\sigma}_{sr}$ are quite wide, leaving room for many potential interpretations of the data.

Just as priors about persistence affect posteriors about unconditional expected returns, priors about unconditional expected returns also affect posteriors about persistence. For example, investors with negative priors on μ_{sr} are more amenable to interpreting positive historical returns as arising from large and persistent variations in conditional expected returns – i.e., higher posteriors for H, γ, and $\hat{\sigma}_{sr}$ – relative to investors with positive priors on μ_{sr}. Table 6 shows this is indeed the case for the value, investment, and profitability portfolios, which exhibit positive average return performance during our sample period.

4.3. Posteriors for Conditional Sharpe Ratios

In addition to unconditional expected returns and Sharpe ratios, the Bayesian analysis allows us to compute posterior distributions of conditional Sharpe ratios at each point in time during our sample period.[24] Figure 3 plots the time series of posterior means of these conditional Sharpe ratios that obtain with the 'agnostic' prior specification with $\mu_{sr} \sim U(-2, 2)$ and $H \sim U(0, 10)$ for the four characteristic-sorted portfolios we study. As the figure shows, the fluctuations in conditional Sharpe ratios are large, generally varying between zero and 0.8 on an annualized basis. Despite having unconditional Sharpe ratio estimates near zero, the conditional Sharpe ratios of the size portfolio are especially volatile, varying from -0.75 to above one. These relatively larger fluctuations are a reflection of the more positive return autocorrelations that the size portfolio exhibits during our sample period.

We also compute posterior distributions of forward-looking conditional Sharpe ratios for the quarter following the end of our sample period (Q1 of 2022). These posteriors differ from the unconditional posteriors because they rely on recent trends to extrapolate future performance. When $\gamma > 0$, the extrapolation is positive, meaning that conditional expected returns are higher (lower) than unconditional expected returns when recent returns are higher (lower) than the full-sample average. When $\gamma < 0$, the extrapolation is negative, meaning that conditional expectations are inversely related to recent trends.

The grey lines on the right in Figure 2 present the means and 95% confidence intervals for posterior beliefs about the 2022 conditional Sharpe ratio of each portfolio. As the first rows show, conditional and unconditional Sharpe ratios are by definition the same when $H = 0$. When $H > 0$, however, both the value and investment portfolios have smaller conditional Sharpe ratios compared to their unconditional Sharpe ratios. For the value portfolio, which had particularly poor recent performance, the pessimistic or neutral Bayesian

[24]The conditional Sharpe ratio distributions reported in this subsection are the posterior beliefs that obtain after observing the full historical sample of return data. Thus, the magnitudes should be interpreted as measuring in-sample economic significance rather than informing real-time portfolio choices. We consider out-of-sample portfolio choices in Section 5.3.

investors believe that the conditional Sharpe ratios in 2022 are centered near zero and could even be quite negative. Because the profitability portfolio performed better in recent years than earlier in the sample, we find the opposite effect in Panel C of Figure 2: posteriors about 2022 Sharpe ratios are *higher* than posteriors about the unconditional Sharpe ratio. As with the bearish view of value and investment, the bullish view for profitability is stronger for larger values of H than smaller ones.

5. Broader Analysis

Our analysis up to this point focuses on the four prominent examples of characteristic-sorted portfolios. As we show, estimates of the unconditional expected returns of these portfolios become less precise once we account for the possibility of persistent variation in conditional expected returns. We also find that there is very little power to estimate the magnitude of persistent variation for the value, investment and profitability portfolios, while there is stronger evidence of persistent variation for the size portfolio.

In this section we explore these issues by applying our model to a broader set of 174 long-short equity portfolios studied in the prior literature. In addition to the four portfolios we have considered so far, these portfolios include all strategies described in the March 2022 update of the "Open Source Cross-Sectional Asset Pricing" dataset (Chen and Zimmermann, 2021) with at least 120 months of historical data. Our analysis continues to focus on quarterly market-neutral returns using value-weighted quintiles.

5.1. Frequentist Analysis of the "Factor Zoo"

The analysis of Chen and Zimmermann (2021) illustrates the apparent ease with which research uncovers significant relationships between firm characteristics and average stock returns, a phenomenon Cochrane (2011) labels the "factor zoo". As Harvey, Liu, and Zhu (2016) and others discuss, this evidence may be partly due to p-hacking; i.e., researchers testing many different characteristics and reporting only those that turn out to have a

statistically-significant relation to future returns. Harvey, Liu, and Zhu (2016) suggests addressing p-hacking by requiring higher t-statistic (or lower p-value) thresholds for rejecting the no-predictability null hypothesis. As Chen and Zimmermann (2021) show, however, higher significance thresholds do not necessarily eliminate the anomaly. For instance, 58 of the 174 portfolios in our sample have p-values below 0.1%, which is extremely unlikely under the null, even with p-hacking.

We offer an alternative interpretation of the factor zoo: by ignoring the possibility of persistent variation in expected returns, the published analyses overstates the statistical significance of estimates of unconditional expected returns. Figure 4, which summarizes our main findings in this regard,[25] shows that relative to OLS estimates that assume i.i.d. returns, rejections of the $\mu = 0$ null hypothesis are weaker in our maximum likelihood estimates that allow for time-variation in expected returns. The gap between the MLE and OLS tests are large for p-values of 0.1% and 1%, which MLE achieve for only 10 and 52 portfolios, respectively, compared to 58 and 81 using OLS.

5.2. Bayesian Analysis of Time-Variation in Conditional Sharpe Ratios

Our next set of results focus on the magnitude of persistent return variation for the broader set of 174 portfolios. To facilitate this analysis, we employ the full-sample Bayesian approach described in Section 4. We report results that obtain with the neutral unconditional Sharpe ratio prior $\mu_{sr} \sim N(0, 0.4)$ and the agnostic half-life prior $H \sim U(0, 10)$.[26]

Panel A of Figure 6 presents the frequency distribution of the posterior mean for $\hat{\sigma}_{sr}$, the volatility of the Sharpe ratio conditional on past returns. We find substantial differences across portfolios in the time-variation in conditional Sharpe ratios. While the posterior means for $\hat{\sigma}_{sr}$ for many of the portfolios are similar to value, investment, and profitability and cluster near or below the center of the prior distribution (0.23, reported in Table 5), a substantial minority are similar to size and have posterior means that are much larger

[25]The complete results for our frequentist and Bayesian analyses of the 174 portfolios are reported in Appendix Table 4.

[26]Results using other priors are available upon request.

than the prior mean. In unreported analysis, we also find that portfolios that exhibit higher (absolute) return correlation with the size portfolio tend to have higher posterior means for $\hat{\sigma}_{sr}$. Examples include portfolios sorted on illiquidity, share price, and idiosyncratic volatility. This result indicates that understanding what drives variations in conditional size premia may also help explain time-variations in a variety of other portfolios.

Panel B of Figure 6 presents the frequency distribution of the difference between the posterior means of 2022 conditional Sharpe ratios and their unconditional Sharpe ratios. The main result from this panel is that this difference seems to be centered close to zero: if anything, the difference between 2022 Sharpe ratios and unconditional Sharpe ratios is slightly positive on average, which may seem surprising given the evidence in McLean and Pontiff (2016) and Smith and Timmermann (2022) that return anomalies have weakened over time. There are two reasons for this result. The first reason is that the Bayesian posteriors are influenced by prior beliefs, which are centered around zero.[27] The second reason is that the average portfolio in our sample performed well toward the end of the sample period, which receives the most weight in Bayesian forecasts of 2022 returns.[28]

5.3. Out-of-Sample Return Forecasts

In our final analysis, we consider the out-of-sample (OSS) forecasting problem for the 174 characteristic-sorted portfolios from the perspective of an investor whose beliefs about the distribution of future returns are influenced by his prior beliefs as well as past return data. To facilitate this analysis, we repeat the Bayesian estimation procedure at the end of each calendar year in our sample, using only past data available up to that point in time. This expanding-window OOS approach allows us to quantify how different prior beliefs about persistence and unconditional expected returns affect Bayesian investors' portfolio choices through the sample period.

For each calendar year y in our sample, starting 10 years after data become available for

[27]The across-portfolio average of posterior means for unconditional Sharpe ratios is 0.27, while the average in-sample Sharpe ratio is 0.33.

[28]For instance, realized returns in 2021 averaged 7.54% compared to the full-sample average of 4.14%.

the given characteristic, we calculate posterior belief distributions about model parameters based on the quarterly data available through y and given a subset of the prior beliefs described in Section 4.1.[29] Next, we generate 50,000 random draws from the posterior parameter distribution and for each parameter draw, generate 10 random observations from the implied return distribution for the next-year. Given these draws, we compute OOS forecasts for average returns $\hat{r}_{i,t} \equiv E_t(\tilde{r}_{i,t+1\to t+4})$

Figure 5 illustrates how different priors about H affect inferences about $\hat{r}_{i,t}$ for the value, investment, profitability, and size portfolios. The solid black and gray lines plot the forecasts made by an investor with a bullish prior ($\mu_{sr} \sim N(0.4, 0.4)$) and an investor with a neutral prior ($\mu_{sr} \sim N(0.0, 0.4)$), respectively, where both investors believe $H = 0$. The dotted lines in Figure 5 show the forecasts made by an investor who accounts for the possibility of persistent variations in returns ($H \sim U(0, 10)$) and has a neutral priors about μ.

The main result in Figure 5 is that the investor with $H \sim U(0, 10)$ priors (dotted line) uses recent past returns to aggressively 'time' their forecast. This investor's forecasts for annualized returns range between approximately -2% and 8% for value, investment, and profitability, and -12% and 10% for size. Variations of this magnitude would result in large changes in optimal portfolio weights given standard utility functions.

Figure 5 also shows that differences in priors about H (comparing the gray and dotted lines) have a larger impact on return forecasts than differences in priors about μ (comparing the gray and black lines). While the bullish investor initially forecasts a larger return than the neutral investor for all four portfolios, this difference gradually dissipates over time as more data becomes available and the investors place less weight on their priors. The forecast differences associated with different priors about H, by contrast, are larger and do not dissipate over time for these portfolios because, as discussed above, the data offer very little guidance on the true value of H.

[29] We repeat the exercise once per year instead of each quarter to save computational time. Given the slow-moving changes in expected returns we are interested in, we would not expect large quarter-to-quarter variations in conditional distributions of either model parameters or returns.

We next assess how the Bayesian forecasts $\hat{r}_{i,t}$ using our most-agnostic prior ($\mu_{sr} \sim N(0, 0.4)$, $H \sim U(0, 10)$) relate to realized next-year returns. To maximize statistical power, we use a Fama and MacBeth (1973) regression approach on the full set of 174 portfolios. At the end of each calendar year, we use the procedure above to compute $\hat{r}_{i,t}$ for all characteristic-sorted portfolios with at least 10 years of past returns, which is 132 to 173 portfolios depending on the year. We then use a cross-sectional regression of future realized annual returns on Bayesian $\hat{r}_{i,t}$ and alternative forecasts $\hat{r}^{\text{alt}}_{i,t}$:

$$r_{i,t+1} = a_t + b_t \cdot (\text{Bayesian } \hat{r}_{i,t}) + c_t \cdot \hat{r}^{\text{alt}}_{i,t} + \epsilon_{i,t+1}. \tag{15}$$

We report the time-series averages of b_t and c_t in Table 7, along with their standard errors.

Column (1) of Table 7 shows that our model's forecasts strongly predict future realized returns. The reported standard errors imply a t-statistic of 4.9 for the no-predictability null and a t-statistic of 1.8 for the one-for-one predictability null. Columns (2) through (4) indicate that a simple equally-weighted averages of past returns is also a strong predictor, particularly when using the full past sample.[30] Column (5), however, shows that including the Bayesian $\hat{r}_{i,t}$ absorbs nearly all of this predictability, indicating that our model's forecasts successfully combine most of the relevant information in past returns.[31]

In summary, our OOS analysis shows that Bayesian investors with priors that allow for time-variation have forecasts for next-period returns that do in fact vary substantially over time. Moreover, these forecasts provide incremental predictors of future realized returns. This success validates that, for the typical characteristic-sorted portfolio, expected returns do in fact vary over time in a manner captured by our statistical model.

[30]The full past sample result, and the past-year result, are similar to the findings in McLean and Pontiff (2016) and Gupta and Kelly (2019), respectively.

[31]The exception is there is some negative incremental predictability from the past five year average $\overline{r}$. This suggests the Bayesian $\hat{r}_{i,t}$ overweights returns in this window relative to the optimal forecast. We confirm this possibility in Columns (6) and (7), which show that past five year $\overline{r}_i$ is a slightly-negative incremental predictor of realized returns but a positive predictor of Bayesian $\hat{r}_{i,t}$.

6. Conclusion

The literature on characteristic-sorted portfolio returns is large and growing. The main motivation for early studies in this literature was to test the CAPM, which predicts market-neutral portfolio returns to be mean-zero in every period and thus uncorrelated over time. Traditional tests in the literature, which are designed to reject the CAPM null, are thus appropriate. Having established strong rejections of the CAPM, however, the natural next step is to evaluate alternative return generating processes more closely. For instance, leading behavioral explanations of the value anomaly highlight reasons for value stocks to beat growth stocks on average, but do not require the value premium to be constant over time, or even necessarily positive every period. It is therefore natural to develop new tests that reflect the possibility that the expected returns of characteristic-sorted portfolios vary over time. We propose a statistical model that admits this possibility and conduct both frequentist and Bayesian analyses based on the model.

There are two main takeaways that emerge from our analyses. The first takeaway is that accounting for time variation in expected returns materially affects statistical inferences about characteristic-sorted portfolio returns. For the value, investment and profitability portfolios the historical data are broadly consistent with a wide variety of return generating processes. For these portfolios, one cannot reject that returns are i.i.d.; yet one can also not reject that they are generated by persistent processes with mean-reverting expected returns. Our findings for these portfolios indicate that the precision of unconditional expected return estimates obtained in traditional tests may be overstated. Furthermore, by focusing on average returns, traditional tests may fail to detect significant variations in conditional expected returns. Our analysis of the size portfolio does in fact reveal significant evidence of persistent variation in conditional expected returns, and as such rejects the CAPM, despite the size portfolio's near-zero average performance.

Our second takeaway relates to how investors should interpret the historical data. The

lack of precise statistical inferences from historical data does not imply that the data are not informative for investors. As our Bayesian analyses indicate, investors with relatively agnostic priors, who are thus open to the possibility of persistent variation in returns, will form posterior beliefs about the conditional expected returns of characteristic-sorted portfolios that fluctuate substantially over time and that increase with portfolios' recent returns. These results have implications for practitioners as well as academics who are interested in characteristic-sorted portfolios. Financial institutions now offer a multitude of relatively passive investment products, such as ETFs and mutual funds, that focus on the long-term links between returns and characteristics. At the same time, there exist active hedge funds that attempt to time short-term variations in characteristic return premia. Our analysis offers guidance for both groups as to how much they should use past performance while forming their performance forecasts for these portfolios.

References

Altı, Aydoğan and Sheridan Titman, 2019, A dynamic model of characteristic-based return predictability, *Journal of Finance* 74, 3187–3216.

Arnott, Robert D, Mark Clements, Vitali Kalesnik, and Juhani T Linnainmaa, 2021a, Factor momentum, Available at SSRN 3116974 .

Arnott, Robert D, Campbell R Harvey, Vitali Kalesnik, and Juhani T Linnainmaa, 2021b, Reports of values death may be greatly exaggerated, *Financial Analysts Journal* 77, 44–67.

Avramov, Doron, Scott Cederburg, and Katarína Lučivjanská, 2018, Are stocks riskier over the long run? taking cues from economic theory, *Review of Financial Studies* 31, 556–594.

Avramov, Doron, Si Cheng, Amnon Schreiber, and Koby Shemer, 2017, Scaling up market anomalies, *Journal of Investing* 26, 89–105.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Cameron, A Colin and Pravin K Trivedi, *Microeconometrics: methods and applications* (Cambridge University Press 2005).

Chen, Andrew Y and Tom Zimmermann, 2021, Open source cross-sectional asset pricing, *Critical Finance Review, Forthcoming* .

Choi, Ki-Soon, Eric C So, and Charles CY Wang, 2021, Going by the book: Valuation ratios and stock returns, SSRN working paper 3854022 .

Cochrane, John H, 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Conrad, Jennifer and Gautam Kaul, 1988, Time-variation in expected returns, *Journal of Business* 409–425.

Daniel, Kent, David Hirsleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.

Ehsani, Sina and Juhani T Linnainmaa, 2022, Factor momentum and the momentum factor, *Journal of Finance* 77, 1877–1919.

Eisfeldt, Andrea L, Edward Kim, and Dimitris Papanikolaou, 2022, Intangible value, *Critical Finance Review* 11, 299–332.

Fama, Eugene F and Kenneth R French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

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

Fama, Eugene F and Kenneth R French, 2021, The value premium, *Review of Asset Pricing Studies* 11, 105–121.

Fama, Eugene F and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Ferson, Wayne E, Sergei Sarkissian, and Timothy T Simin, 2003, Spurious regressions in financial economics?, *Journal of Finance* 58, 1393–1413.

Gârleanu, Nicolae, Leonid Kogan, and Stavros Panageas, 2012, Displacement risk and asset returns, *Journal of Financial Economics* 105, 491–510.

Goncalves, Andrei and Gregory Leonard, 2023, The fundamental-to-market ratio and the value premium decline, *Journal of Financial Economics* 147, 382–405.

Gupta, Tarun and Bryan Kelly, 2019, Factor momentum everywhere, *Journal of Portfolio Management* 45, 13–36.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hodrick, Robert J, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357–386.

Israel, Ronen, Kristoffer Laursen, and Scott Richardson, 2020, Is (systematic) value investing dead?, *Journal of Portfolio Management* 47, 38–62.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *Journal of Finance* 69, 611–644.

Kandel, Shmuel and Robert F Stambaugh, 1996, On the predictability of stock returns: an asset-allocation perspective, *Journal of Finance* 51, 385–424.

Kendall, Maurice G, 1954, Note on bias in the estimation of autocorrelation, *Biometrika* 41, 403–404.

Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman, 2020, Left behind: Creative destruction, inequality, and the stock market, *Journal of Political Economy* 128, 855–906.

Lewellen, Jonathan, 2002, Momentum and autocorrelation in stock returns, *Review of Financial Studies* 15, 533–564.

Linnainmaa, Juhani T and Michael R Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606–2649.

Marriott, FHC and JA Pope, 1954, Bias in the estimation of autocorrelations, *Biometrika* 41, 390–402.

McLean, R David and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Newey, Whitney K and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

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

Pástor, L'uboš, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.

Pástor, L'uboš and Robert F Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

Pástor, L'uboš and Robert F Stambaugh, 2012, Are stocks really less volatile in the long run?, *Journal of Finance* 67, 431–478.

Shiller, Robert J, *Irrational exuberance* (Princeton university press 2000).

Smith, Simon and Allan Timmermann, 2022, Have risk premia vanished?, *Journal of Financial Economics* 145, 553–576.

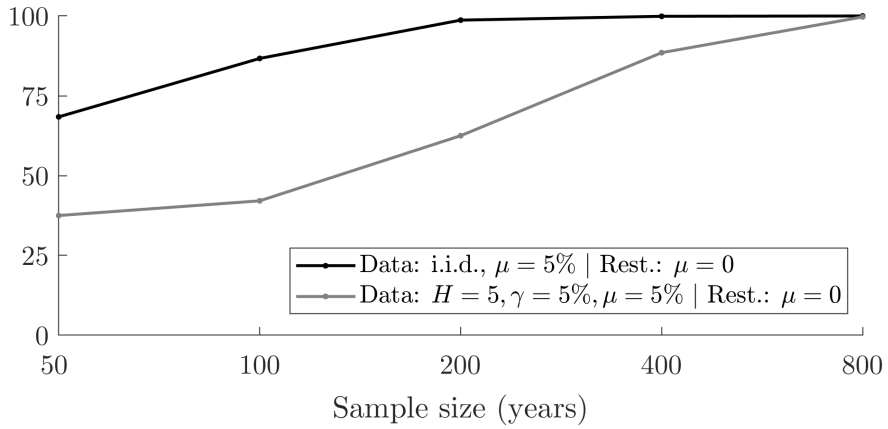
Wachter, Jessica A and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 817–838.

Figure 1: Maximum Likelihood Hypothesis Tests in Simulated Samples

This figure reports rejection rates for parameter restrictions that obtain with maximum likelihood ratio tests on simulated samples of varying lengths. Each sample consists of quarterly data simulated under a specific parameterization of our model. For each parameterization of each length we simulate 2,500 samples. In all simulated samples, the annualized return volatility $\sigma_r = 0.15$. The parameter μ is the annualized unconditional expected return, H is the annualized half-life of shocks to expected return, and γ is the one-quarter autocorrelation in returns. The parameters of the process that generate the simulated data are given in the "Data" label for each line. The parameter restrictions for the tested hypothesis are given in the "Rest." label for each line, with any unspecified parameter left unrestricted. The lines in the figures report the fraction of simulated samples, in percent, for which a likelihood ratio test rejects the restriction at the 5% level.

Panel A: % of simulated samples rejecting restrictions on μ



Panel B: % of simulated samples rejecting restrictions on H and γ

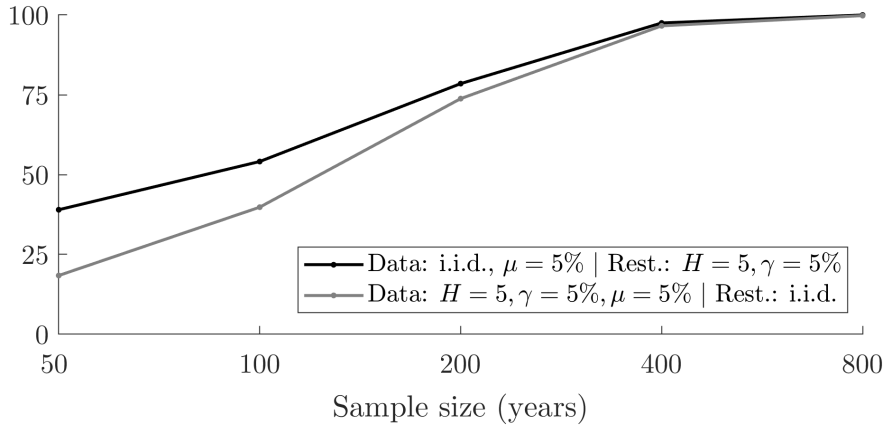


Figure 2: Posterior Sharpe Ratios

This figure reports the means and 95% confidence intervals for posterior distributions of Sharpe ratios for value, investment, profitability, and size portfolios, as defined in the header of Table 1. Posteriors for the unconditional Sharpe ratio are in black on the left, and the conditional Sharpe ratio as of Q1 of 2022 are in grey on the right. Each panel shows the posteriors for 16 different (H, μ_{sr}) combinations of prior beliefs, where H is the half-life of shocks to expected returns in years, and μ_{sr} is the unconditional Sharpe ratio. All Sharpe ratios are annualized. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

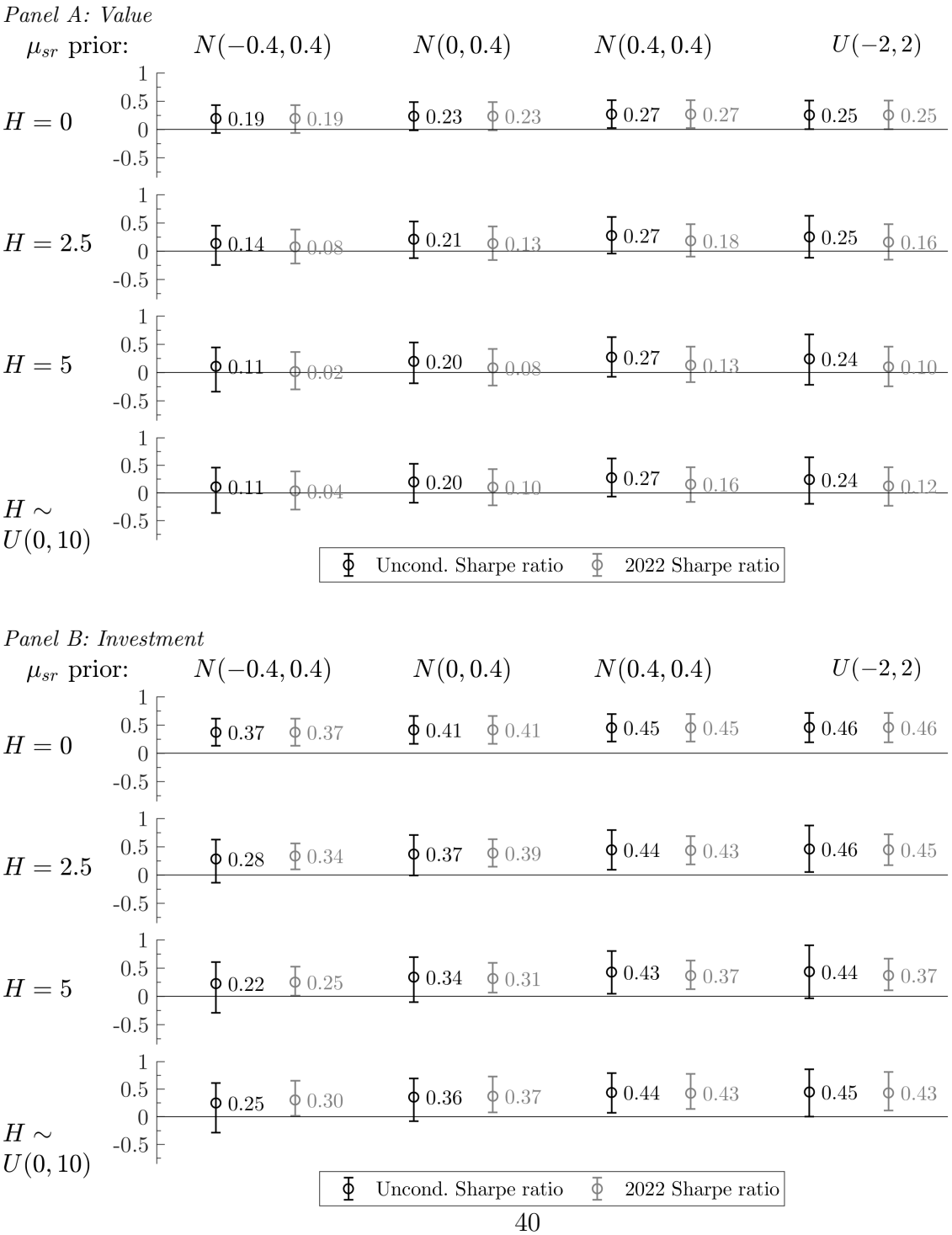
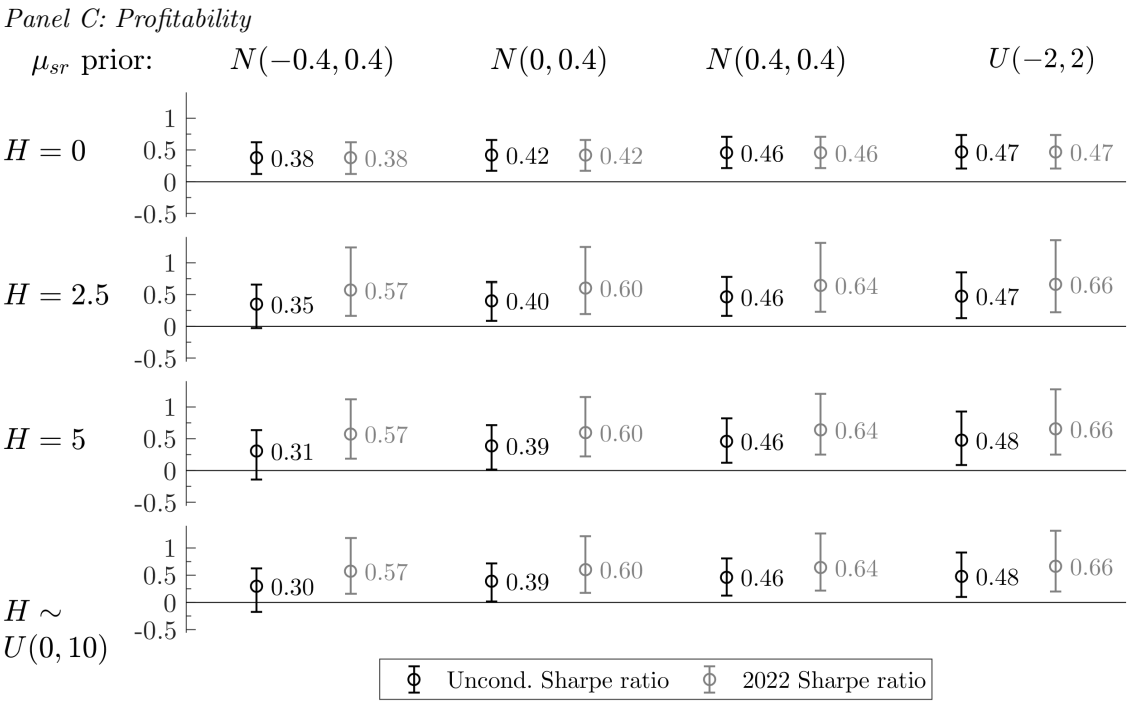


Figure 2: Posterior Sharpe Ratios (continued)



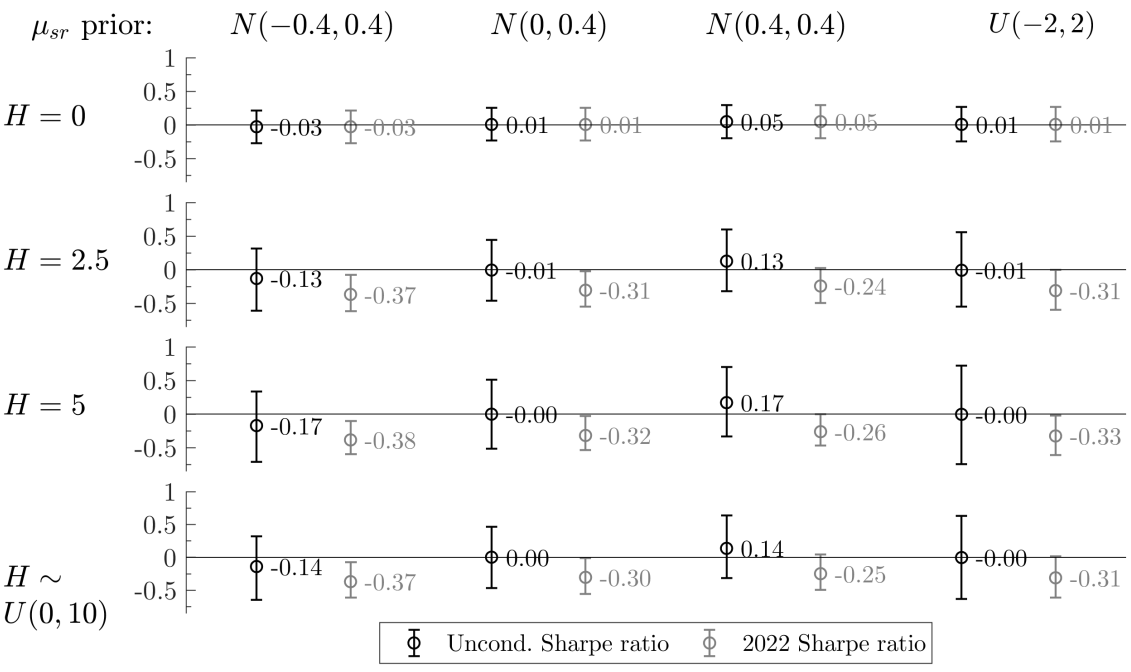
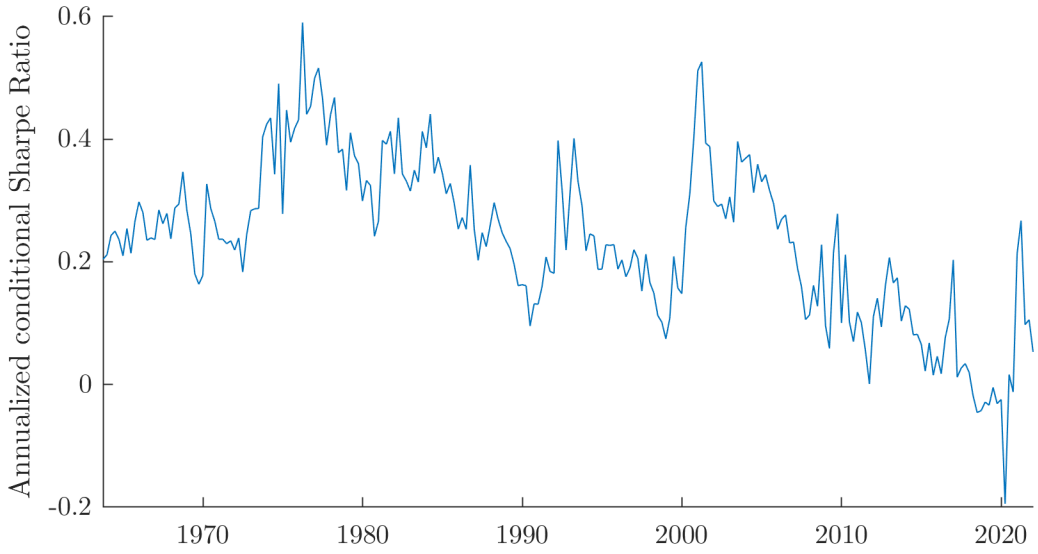


Figure 3: Posteriors on Conditional Sharpe Ratios Across Time

This figure shows means of posterior belief distributions for conditional Sharpe ratios of value, investment, profitability, and size portfolios, as defined in the header of Table 1. Posterior distributions are computed using the full sample of historical data and the prior beliefs that the unconditional Sharpe ratio μ_{sr} is distributed $N(0, 0.4)$ and the half-life of shocks to expected returns H is distributed $U(0, 10)$. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Value



Panel B: Investment

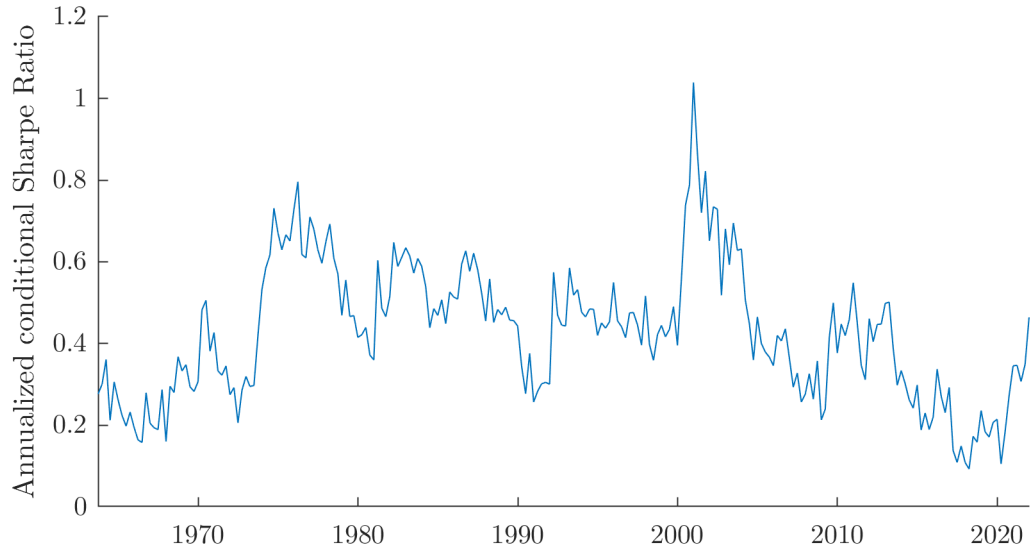
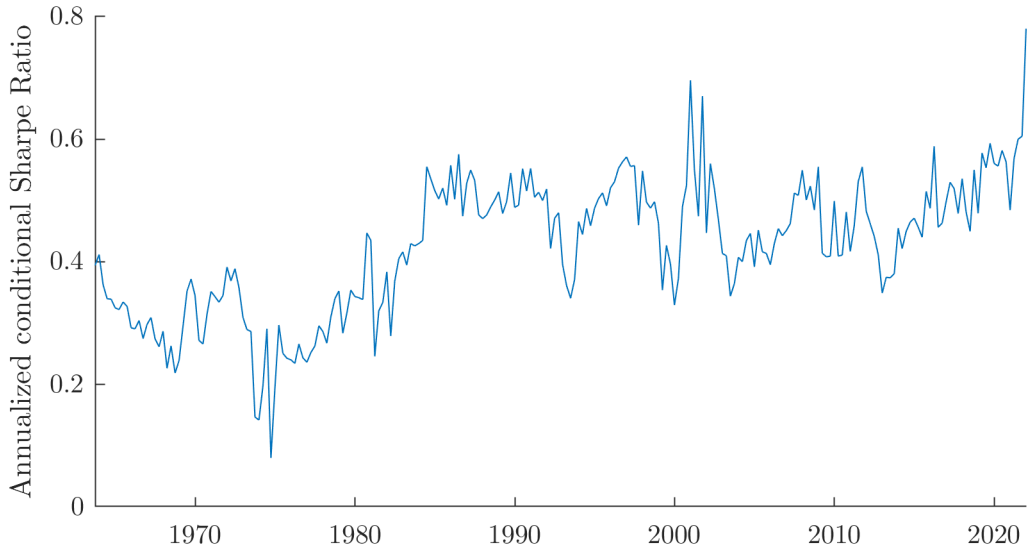


Figure 3: Posteriors on Conditional Sharpe Ratios Across Time (continued)

Panel C: Profitability



Panel D: Size

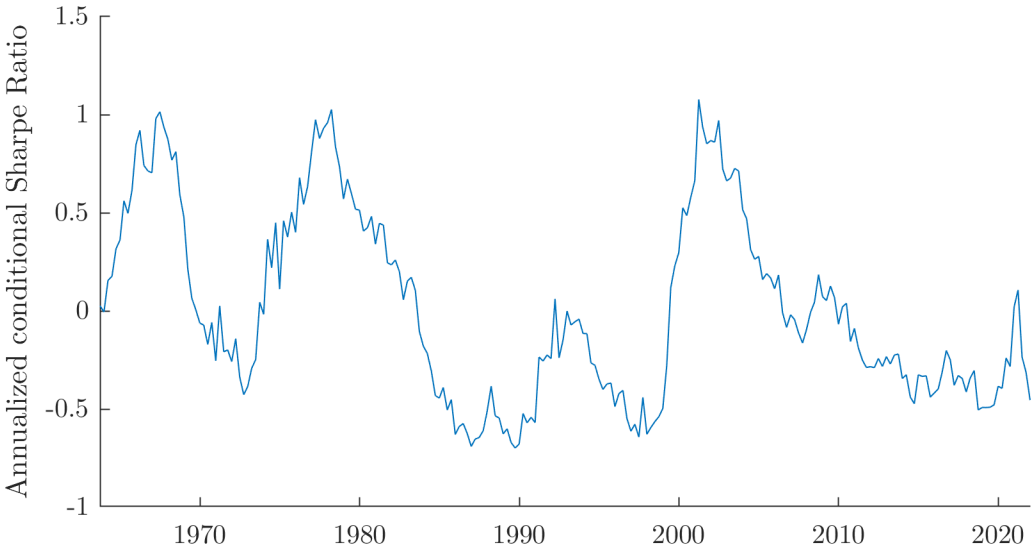


Figure 4: Unconditional Expected Return Inferences for 174 Portfolios

This figure reports the number of characteristic-sorted portfolios for which the null hypothesis that the unconditional expected return μ is zero is rejected at different p-value thresholds. The blue bars use OLS with White (1980) standard errors. The red bars use likelihood ratio tests based on the maximum likelihood estimates of our statistical model, as described in Section 3.2.1. We conduct these tests for quarterly returns of 174 long-short equity portfolios, which include the four portfolios that we focus on in our main analysis and 170 portfolios analyzed by Chen and Zimmermann (2021). The sample period varies across portfolios and consists of at most 234 quarterly observations from Q3 1963 through Q4 2021.

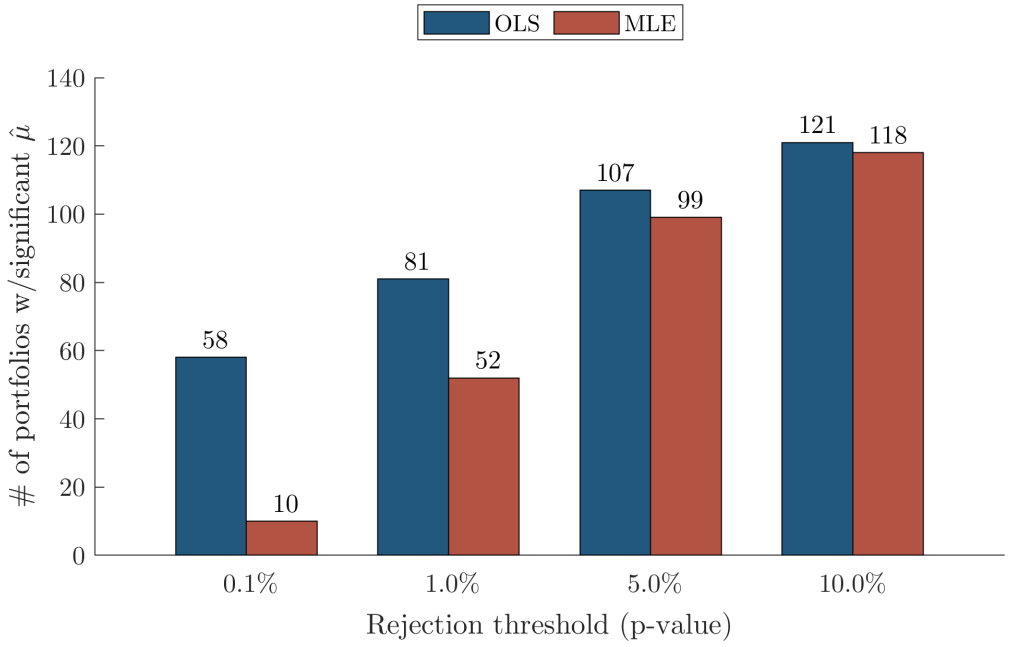
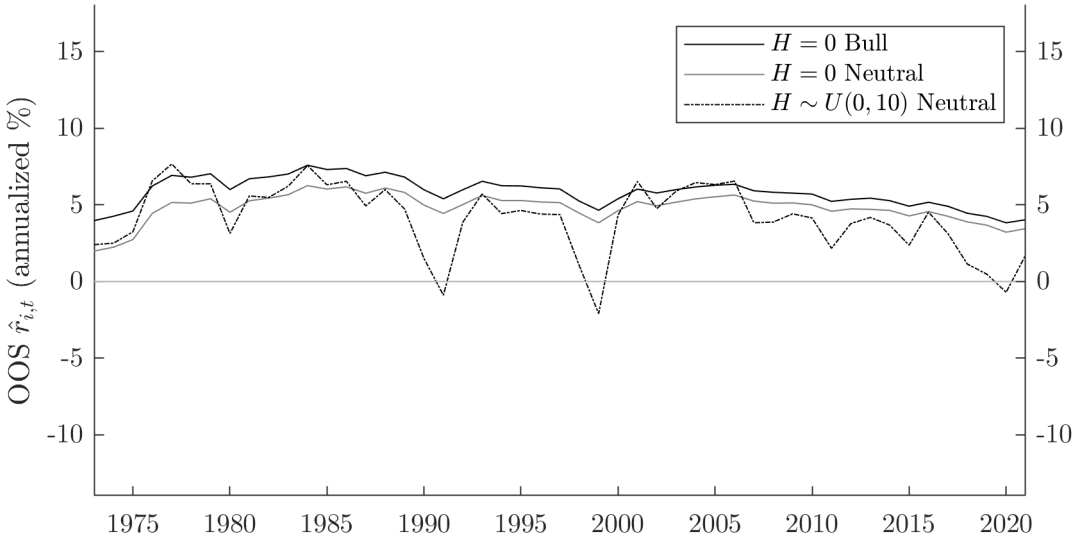


Figure 5: Out-of-Sample Expected Return Forecasts

This figure presents conditional expected return forecasts $\hat{r}_{i,t}$ computed by Bayesian investors with different priors who calculate posteriors using only past data available at each point in time. The solid black line shows $\hat{r}_{i,t}$ for an investor with priors that the half-life of shocks to expected returns H equals zero and that the unconditional Sharpe ratio μ_{sr} is distributed $N(0.4, 0.4)$ ('Bull'). The solid gray line shows $\hat{r}_{i,t}$ for an investor with priors that $H = 0$ and μ_{sr} is distributed $N(0, 0.4)$ ('Neutral'). The dotted black line shows $\hat{r}_{i,t}$ for an investor with priors that $H \sim U(0, 10)$ and $\mu_{sr} \sim N(0, 0.4)$. Return forecasts are for market-neutral returns in annualized percents. We compute forecasts at the beginning of each calendar year from 1973–2021 based on quarterly return data starting in 1963.

Panel A: Value



Panel B: Investment

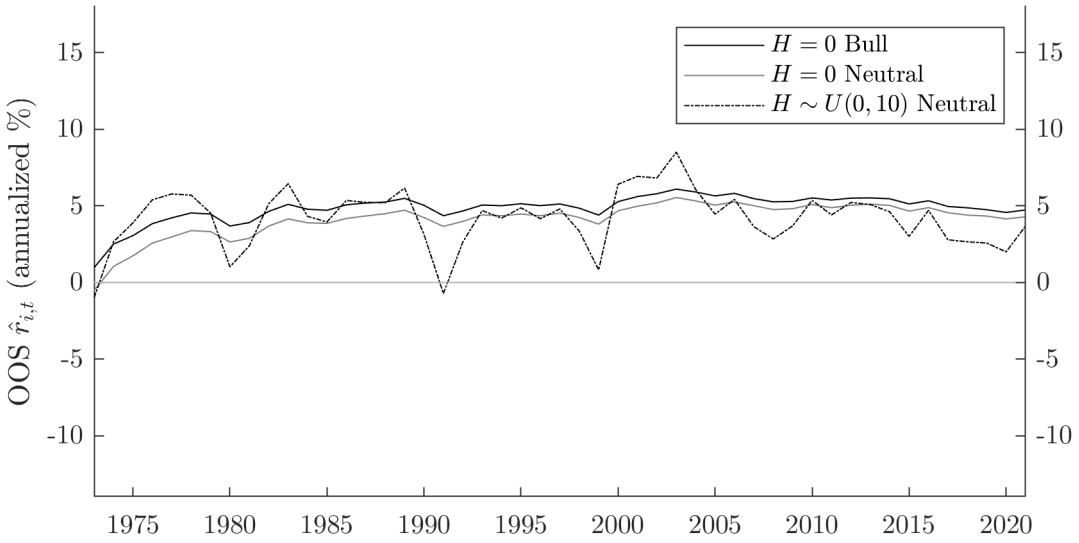
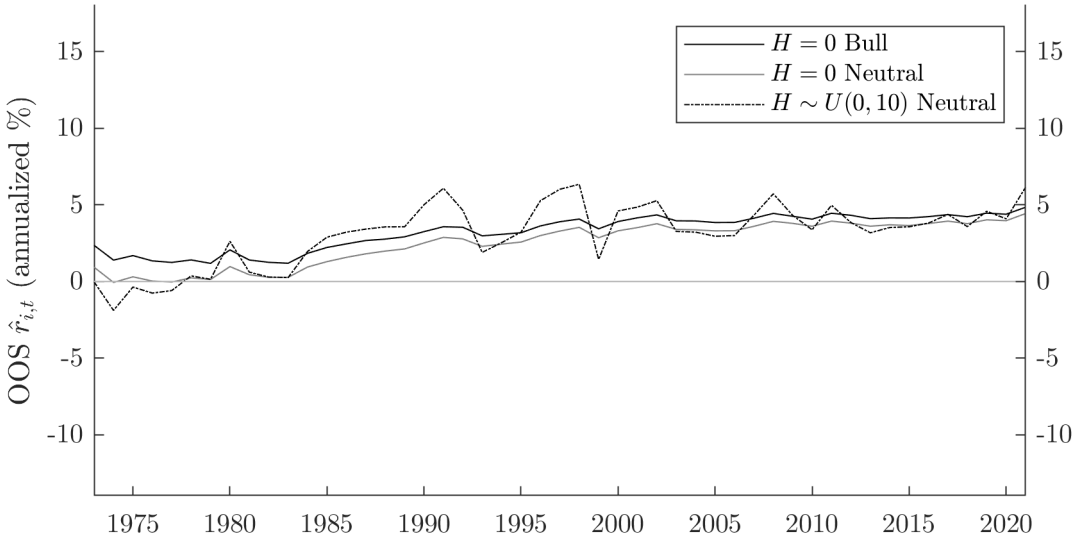


Figure 5: Out-of-Sample Expected Return Forecasts (continued)

Panel C: Profitability

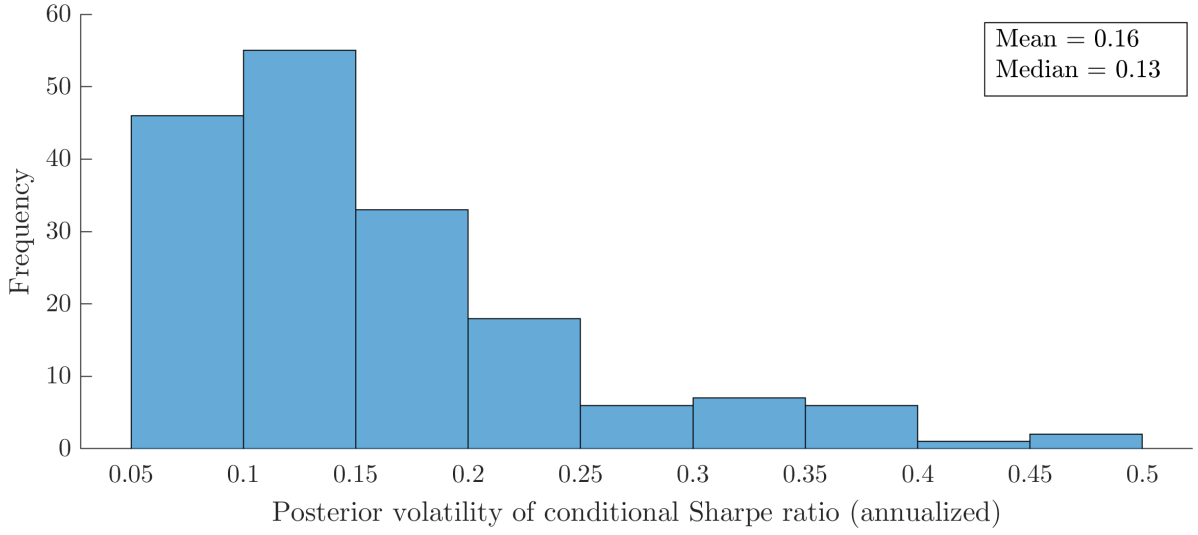


Panel D: Size

Figure 6: Bayesian Posteriors for 174 Portfolios

This figure presents frequency distributions that summarize the Bayesian inferences about the conditional Sharpe ratios of 174 long-short equity portfolios, which include the four portfolios that we focus on in our main analysis and 170 portfolios analyzed by Chen and Zimmermann (2021). Panel A plots the frequency distribution of the posterior mean for $\hat{\sigma}_{sr}$, the volatility of the Sharpe ratio conditional on past returns. Panel B plots the frequency distribution of the difference between the posterior means for the 2022 conditional Sharpe ratio and the unconditional Sharpe ratio. All Sharpe ratios are annualized. The sample period varies across portfolios and consists of at most 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Volatility of conditional Sharpe ratios



Panel B: Conditional Sharpe ratios in 2022 relative to unconditional Sharpe ratios

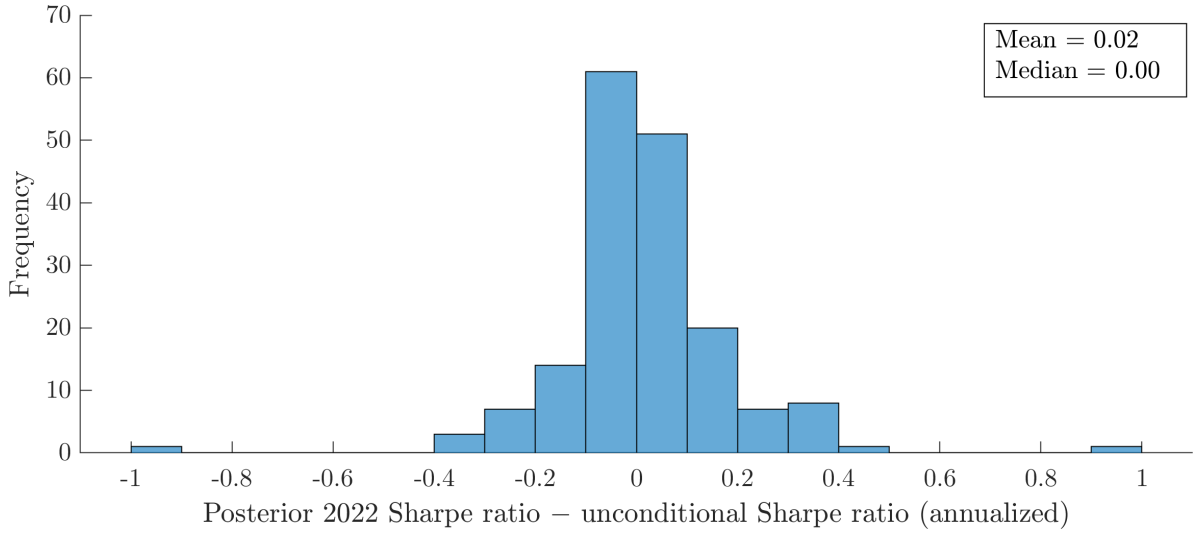


Table 1: Historical Performance of Characteristic-Sorted Portfolios

This table presents statistics summarizing the historical performance of value-weighted quintile portfolios formed on value, investment, profitability, and size characteristics. The value portfolio is based on sorting firms by their book-to-market ratios, the investment portfolio on sorting by the annual growth rate of total assets, the profitability portfolio on sorting by operating profits divided by book equity, and the size portfolio on sorting by market capitalization, as in Fama and French (2015). The investment portfolio is long firms in the lowest quintile and short firms in the highest quintile, while the other three portfolios are long the highest quintile and short the lowest. For each portfolio, we compute market-neutral returns by hedging out market risk using the full-sample market β. We report the mean annualized quarterly return in percentage terms and the annualized Sharpe ratio for the full sample, two subsamples, and the difference between the subsamples. t-statistics based on iid re-sampling of the calendar quarters in each sample are in parenthesis. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

	Mean (annualized %)				Sharpe ratio (annualized)			
	All	1963–1992	1993–2021	Diff	All	1963–1992	1993–2021	Diff
Value	3.80	6.09	1.47	-4.62	0.25	0.45	0.09	-0.36
	(1.92)	(2.19)	(0.52)	-(1.17)	(1.91)	(2.40)	(0.47)	-(1.35)
Investment	4.75	4.81	4.68	-0.12	0.45	0.48	0.41	-0.07
	(3.39)	(2.46)	(2.35)	-(0.04)	(3.36)	(2.59)	(2.18)	-(0.26)
Profitability	4.91	3.40	6.45	3.05	0.45	0.36	0.54	0.18
	(3.47)	(1.71)	(3.20)	(1.08)	(3.45)	(1.92)	(2.86)	(0.69)
Size	0.16	1.97	-1.68	-3.65	0.01	0.12	-0.12	-0.24
	(0.08)	(0.69)	-(0.58)	-(0.90)	(0.08)	(0.64)	-(0.62)	-(0.89)

Table 2: Autocorrelation Estimates with Simulated and Historical Data

This table presents statistics summarizing autocorrelation estimates with model-simulated and historical portfolio returns. For each portfolio, we run overlapping time-series regressions of quarterly returns r_t on a constant and rolling averages of past returns from quarters $t-L$ through $t-1$:

$$r_t = a + b_L \left(\frac{1}{L} \sum_{l=1}^{L} r_{t-l} \right) + \epsilon_t. \tag{16}$$

Panel A presents average coefficients b_L and 95% confidence intervals from 50,000 simulated 234-quarter samples under various parameterizations of the model. The parameter H is the half-life of shocks to expected returns denoted in years, γ is the first-order autocorrelation of quarterly returns, and $\hat{\sigma}_{sr}$ is the volatility of the annualized Sharpe ratio conditional on past realized returns. Panel B show estimates for the value, investment, profitability, and size portfolios, as defined in the header of Table 1. Standard errors based on iid re-sampling of the calendar months in our sample are in parenthesis. The joint significance row presents the fraction of i.i.d. simulations for which the sum of the $\hat{b}$ across the four portfolios exceeds the sum in observed data. The historical sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Model Simulations

Parameterization			Average coefficient [95% confidence interval]		
H	γ	$\hat{\sigma}_{sr}$	b_4	b_{10}	b_{20}
IID		0.00	-0.03	-0.09	-0.20
			[-0.31,0.22]	[-0.59,0.30]	[-1.04,0.38]
2.5	2.5%	0.12	0.04	0.06	0.01
			[-0.24,0.29]	[-0.44,0.43]	[-0.76,0.52]
2.5	5%	0.22	0.11	0.16	0.14
			[-0.18,0.36]	[-0.32,0.52]	[-0.59,0.61]
2.5	10%	0.38	0.22	0.32	0.31
			[-0.07,0.47]	[-0.13,0.63]	[-0.32,0.71]
5	2.5%	0.15	0.04	0.06	0.03
			[-0.25,0.29]	[-0.45,0.44]	[-0.76,0.55]
5	5%	0.26	0.10	0.16	0.18
			[-0.19,0.36]	[-0.33,0.54]	[-0.57,0.66]
10	2.5%	0.18	0.03	0.03	0.01
			[-0.26,0.28]	[-0.48,0.43]	[-0.81,0.56]
10	5%	0.30	0.08	0.13	0.15
			[-0.22,0.35]	[-0.38,0.53]	[-0.64,0.66]

Panel B: Estimates for historical data

	b_4	b_{10}	b_{20}
Value	0.12	-0.07	0.11
Investment	0.21	0.07	-0.24
Profitability	0.15	-0.18	-0.30
Size	0.42	0.48	0.18
Pooled	0.25	0.20	0.07
iid p-value	0.0%	2.2%	16.7%
Pooled (without size)	0.15	-0.05	-0.07
iid p-value	2.8%	48.9%	40.4%

Table 3: OLS with Time-Varying Expected Returns

This table presents estimates of unconditional expected returns of the value, investment, profitability, and size portfolios, which are defined in the header of Table 1, along with standard errors calculated using a variety of approaches. The first row reports the point estimate of the unconditional expected return, $\hat{\mu}$, in annualized percentage terms. The 'Unadjusted SE' row reports the typical OLS standard error, which is calculated under the assumption of independently distributed returns. The next three rows report standard errors calculated using the Newey and West (1987) adjustment with 10, 20, or 40 quarterly lags. The remaining rows report OLS standard errors corrected for autocorrelation using Equation (12) with different values of H, the half-life of shocks to expected returns in years, and γ, the first-order autocorrelation of quarterly returns. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

	Value	Investment	Profitability	Size
$\hat{\mu}$ (annualized %)	3.80	4.75	4.91	0.16
Unadjusted SE	(1.97)	(1.39)	(1.41)	(2.02)
Newey-West SE (10 lags)	(2.04)	(1.58)	(1.42)	(2.89)
Newey-West SE (20 lags)	(2.03)	(1.48)	(1.29)	(3.08)
Newey-West SE (40 lags)	(2.01)	(1.36)	(1.36)	(2.78)
Model-based SE ($H = 0.5$, $\gamma = 2.5\%$)	(2.14)	(1.50)	(1.53)	(2.19)
Model-based SE ($H = 0.5$, $\gamma = 5\%$)	(2.29)	(1.61)	(1.63)	(2.34)
Model-based SE ($H = 0.5$, $\gamma = 10\%$)	(2.56)	(1.80)	(1.83)	(2.62)
Model-based SE ($H = 2.5$, $\gamma = 2.5\%$)	(2.58)	(1.81)	(1.84)	(2.64)
Model-based SE ($H = 2.5$, $\gamma = 5\%$)	(3.06)	(2.15)	(2.19)	(3.13)
Model-based SE ($H = 2.5$, $\gamma = 10\%$)	(3.85)	(2.71)	(2.75)	(3.94)
Model-based SE ($H = 5$, $\gamma = 2.5\%$)	(2.99)	(2.10)	(2.13)	(3.06)
Model-based SE ($H = 5$, $\gamma = 5\%$)	(3.74)	(2.62)	(2.67)	(3.82)
Model-based SE ($H = 10$, $\gamma = 2.5\%$)	(3.54)	(2.48)	(2.52)	(3.62)
Model-based SE ($H = 10$, $\gamma = 5\%$)	(4.60)	(3.23)	(3.28)	(4.70)

Table 4: Maximum Likelihood Hypothesis Tests

This table presents parameter estimates and hypothesis tests based on maximum-likelihood estimations of our model for the value, investment, profitability, and size portfolios, which are defined in the header of Table 1. Panel A reports estimates of μ, the unconditional expected return in annualized percentage terms; σ_r, the volatility of returns in annualized percentage terms; H, the half-life of shocks to expected returns in years; and γ, the first-order autocorrelation of quarterly returns. The rows labelled $\mu = 0$ re-estimate the model with μ restricted to be zero, and report the p-value for this restriction based on a likelihood ratio test. Panel B presents a variety of hypothesis tests for different restrictions on H and γ. In each case, we report the likelihood ratio p-value of the (H, γ) restriction (Rest. p-value), and the likelihood ratio p-value of the $\mu = 0$ restriction given the (H, γ) restriction. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Tests for $\mu = 0$

	IID			Time-varying means				
	μ (%)	σ_r (%)	$\mu = 0$ p-value	μ (%)	σ_r (%)	H (years)	γ (%)	$\mu = 0$ p-value
Value	3.80	15.10		3.77	15.10	0.14	11.82	
$\mu = 0$	0.00	15.22	5.5%	0.00	15.22	0.17	13.32	10.4%
Investment	4.75	10.60		4.77	10.60	0.34	9.19	
$\mu = 0$	0.00	10.87	0.1%	0.00	10.80	10.00	4.80	1.2%
Profitability	4.91	10.77		4.97	10.78	0.21	7.63	
$\mu = 0$	0.00	11.05	0.1%	0.00	10.98	10.00	4.29	1.4%
Size	0.16	15.43		-0.11	15.43	1.23	14.02	
$\mu = 0$	0.00	15.43	93.5%	0.00	15.43	1.23	14.02	97.5%

Panel B: Restrictions on H and γ

Restrictions	H (years):		2.5			5			10		
	γ (%):	IID	-	2.5	5	-	2.5	5	-	2.5	5
Value	Rest. p-value (%)	18.5	6.7	15.6	10.2	6.6	13.3	7.8	6.9	12.1	7.3
	$\mu = 0$ p-value (%)	5.5	12.4	15.4	23.9	15.4	24.1	36.8	18.2	35.1	50.8
Investment	Rest. p-value (%)	26.1	11.0	27.2	22.5	13.8	18.8	12.8	22.2	14.3	9.3
	$\mu = 0$ p-value (%)	0.1	3.6	1.1	3.3	3.2	3.4	9.8	2.8	8.8	21.5
Profitability	Rest. p-value (%)	48.7	23.3	33.3	19.5	23.1	30.5	16.5	25.2	29.6	16.9
	$\mu = 0$ p-value (%)	0.1	0.7	0.7	2.0	1.4	1.8	5.1	2.9	4.5	11.0
Size	Rest. p-value (%)	0.5	25.4	4.2	13.2	4.6	2.1	4.4	1.0	1.0	1.5
	$\mu = 0$ p-value (%)	93.5	96.7	96.0	97.9	96.7	95.8	97.6	96.9	94.9	96.6

Table 5: Prior Beliefs about Model Parameters

This table presents summary statistics for the prior distributions that we use for our Bayesian analyses. Panel A lists the possible priors we consider for μ_{sr}, the annualized unconditional Sharpe ratio; H, the half-life of shocks to expected returns in years; σ_r, the volatility of returns in annualized percentage terms; σ_{sr}, the volatility of the annualized conditional Sharpe ratio; and ρ, the correlation between shocks to unexpected and expected returns. The term $N(\textit{mean}, \textit{standard deviation})$ indicates a normal distribution, $U(\textit{lower bound}, \textit{upper bound})$ indicates a uniform distribution, and a number stated without a distribution indicates a dogmatic prior that the parameter value equals that number. Panel B presents the means and 95% confidence intervals implied by each (H, μ_{sr}) prior parameterization for prior distributions of μ, the unconditional expected return in annualized percentage terms; γ, the first-order autocorrelation of quarterly returns; and $\hat{\sigma}_{sr}$, the volatility of the annualized Sharpe ratio conditional on past realized returns.

Panel A: Priors on Transformed Parameters

μ_{sr}	H (years)	σ_r (%)	σ_{sr}	ρ
$N(-0.4, 0.4)$	0	$U(10, 20)$	$U(0, 1)$	$U(-0.5, 0)$
$N(0, 0.4)$	2.5			
$N(0.4, 0.4)$	5			
$U(-2, 2)$	$U(0, 10)$			

Panel B: Moments of Priors

		μ (%)		γ (%)		$\hat{\sigma}_{sr}$	
H	μ_{sr}	mean	95% CI	mean	95% CI	mean	95% CI
0	$N(-0.4, 0.4)$	-5.30	[-15.90,5.16]	-	-	-	-
0	$N(0, 0.4)$	-0.01	[-10.47,10.42]	-	-	-	-
0	$N(0.4, 0.4)$	5.32	[-5.10,16.05]	-	-	-	-
0	$U(-2, 2)$	-0.02	[-25.84,25.70]	-	-	-	-
2.5	$N(-0.4, 0.4)$	-5.34	[-16.07,5.07]	4.81	[-0.58,15.44]	0.19	[0.00,0.52]
2.5	$N(0, 0.4)$	-0.04	[-10.42,10.37]	4.76	[-0.59,15.41]	0.19	[0.00,0.52]
2.5	$N(0.4, 0.4)$	5.29	[-5.10,15.98]	4.78	[-0.58,15.43]	0.19	[0.00,0.52]
2.5	$U(-2, 2)$	-0.05	[-25.88,25.72]	4.76	[-0.58,15.45]	0.19	[0.00,0.52]
5	$N(-0.4, 0.4)$	-5.30	[-15.94,5.11]	5.58	[-0.25,16.49]	0.25	[0.00,0.61]
5	$N(0, 0.4)$	0.01	[-10.47,10.47]	5.60	[-0.26,16.47]	0.25	[0.00,0.61]
5	$N(0.4, 0.4)$	5.23	[-5.18,15.94]	5.61	[-0.26,16.54]	0.25	[0.00,0.61]
5	$U(-2, 2)$	0.00	[-25.75,25.93]	5.56	[-0.25,16.41]	0.25	[0.00,0.61]
$U(0, 10)$	$N(-0.4, 0.4)$	-5.32	[-16.03,4.99]	4.91	[-1.91,16.37]	0.23	[0.00,0.63]
$U(0, 10)$	$N(0, 0.4)$	-0.02	[-10.48,10.38]	4.90	[-1.79,16.37]	0.23	[0.00,0.63]
$U(0, 10)$	$N(0.4, 0.4)$	5.29	[-5.14,15.98]	4.92	[-1.76,16.40]	0.23	[0.00,0.63]
$U(0, 10)$	$U(-2, 2)$	-0.01	[-25.81,25.82]	4.90	[-1.83,16.38]	0.23	[0.00,0.63]

Table 6: Posterior Beliefs

This table presents summary statistics for posterior distributions that obtain given the historical return data for the value, investment, profitability, and size portfolios, as described in the header of Table 1. The first two columns describe the prior belief specification (H, μ_{sr}), where H is the half-life of shocks to expected returns and μ_{sr} is the unconditional Sharpe ratio. The remaining columns report the means and 95% confidence intervals for the posterior distributions of μ, the unconditional expected return; H; γ, the first-order autocorrelation of quarterly returns; and $\hat{\sigma}_{sr}$, the volatility of the Sharpe ratio conditional on past realized returns. The variables μ, H, and $\hat{\sigma}_{sr}$ are annualized. We compute the posterior beliefs using an M.C.M.C. procedure described in Appendix D. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Value

Prior		μ (%)		H (years)		γ (%)		$\hat{\sigma}_{sr}$	
H	μ_{sr}	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
0	$N(-0.4, 0.4)$	2.86	[-0.94,6.40]	-	-	-	-	-	-
0	$N(0, 0.4)$	3.46	[-0.23,7.13]	-	-	-	-	-	-
0	$N(0.4, 0.4)$	3.99	[0.41,7.64]	-	-	-	-	-	-
0	$U(-2, 2)$	3.76	[0.05,7.53]	-	-	-	-	-	-
2.5	$N(-0.4, 0.4)$	2.02	[-3.53,6.67]	-	-	3.04	[-0.60,12.07]	0.13	[0.00,0.44]
2.5	$N(0, 0.4)$	3.11	[-1.85,7.65]	-	-	2.77	[-0.59,11.73]	0.12	[0.00,0.43]
2.5	$N(0.4, 0.4)$	4.05	[-0.59,8.74]	-	-	2.67	[-0.60,11.51]	0.12	[0.00,0.42]
2.5	$U(-2, 2)$	3.72	[-1.72,9.10]	-	-	2.99	[-0.61,12.26]	0.13	[0.00,0.44]
5	$N(-0.4, 0.4)$	1.62	[-4.94,6.56]	-	-	3.12	[-0.28,12.62]	0.16	[0.00,0.51]
5	$N(0, 0.4)$	2.91	[-2.82,7.85]	-	-	2.70	[-0.29,11.85]	0.14	[0.00,0.49]
5	$N(0.4, 0.4)$	4.02	[-1.17,9.15]	-	-	2.70	[-0.28,11.78]	0.14	[0.00,0.49]
5	$U(-2, 2)$	3.57	[-3.21,9.78]	-	-	3.00	[-0.29,12.52]	0.15	[0.00,0.51]
$U(0, 10)$	$N(-0.4, 0.4)$	1.63	[-5.25,6.73]	4.83	[0.33,9.73]	3.29	[-0.68,12.68]	0.16	[0.00,0.51]
$U(0, 10)$	$N(0, 0.4)$	2.93	[-2.59,7.80]	4.65	[0.30,9.73]	2.81	[-0.87,11.70]	0.14	[0.00,0.47]
$U(0, 10)$	$N(0.4, 0.4)$	4.05	[-0.93,9.11]	4.64	[0.28,9.72]	2.72	[-0.87,11.71]	0.13	[0.00,0.47]
$U(0, 10)$	$U(-2, 2)$	3.56	[-2.84,9.35]	4.79	[0.34,9.76]	3.00	[-0.63,12.24]	0.15	[0.00,0.49]

Table 6: Posterior Beliefs (continued)

Panel B: Investment

Prior		μ (%)		H (years)		γ (%)		$\hat{\sigma}_{sr}$	
H	μ_{sr}	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
0	$N(-0.4, 0.4)$	3.86	[1.42,6.29]	-	-	-	-	-	-
0	$N(0, 0.4)$	4.31	[1.74,6.84]	-	-	-	-	-	-
0	$N(0.4, 0.4)$	4.69	[2.12,7.22]	-	-	-	-	-	-
0	$U(-2, 2)$	4.75	[2.06,7.38]	-	-	-	-	-	-
2.5	$N(-0.4, 0.4)$	2.91	[-1.38,6.42]	-	-	4.65	[-0.52,13.96]	0.19	[0.00,0.49]
2.5	$N(0, 0.4)$	3.79	[-0.13,7.19]	-	-	4.08	[-0.52,13.47]	0.17	[0.00,0.47]
2.5	$N(0.4, 0.4)$	4.56	[0.94,8.03]	-	-	3.98	[-0.53,13.50]	0.17	[0.00,0.47]
2.5	$U(-2, 2)$	4.70	[0.53,8.79]	-	-	4.09	[-0.54,13.51]	0.17	[0.00,0.47]
5	$N(-0.4, 0.4)$	2.34	[-2.91,6.31]	-	-	4.50	[-0.22,14.64]	0.21	[0.00,0.56]
5	$N(0, 0.4)$	3.52	[-1.06,7.19]	-	-	3.62	[-0.27,13.79]	0.18	[0.00,0.54]
5	$N(0.4, 0.4)$	4.44	[0.48,8.16]	-	-	3.29	[-0.30,13.41]	0.16	[0.00,0.53]
5	$U(-2, 2)$	4.51	[-0.38,9.24]	-	-	3.61	[-0.26,13.79]	0.18	[0.00,0.54]
$U(0, 10)$	$N(-0.4, 0.4)$	2.57	[-2.95,6.25]	4.75	[0.38,9.72]	4.46	[-0.44,14.14]	0.20	[0.00,0.55]
$U(0, 10)$	$N(0, 0.4)$	3.65	[-0.83,7.11]	4.33	[0.36,9.71]	3.86	[-0.62,13.41]	0.17	[0.00,0.52]
$U(0, 10)$	$N(0.4, 0.4)$	4.51	[0.71,8.00]	4.20	[0.33,9.64]	3.60	[-0.65,12.81]	0.16	[0.00,0.50]
$U(0, 10)$	$U(-2, 2)$	4.58	[0.02,8.65]	4.19	[0.30,9.65]	3.77	[-0.66,13.63]	0.17	[0.00,0.52]

Panel C: Profitability

Prior		μ (%)		H (years)		γ (%)		$\hat{\sigma}_{sr}$	
H	μ_{sr}	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
0	$N(-0.4, 0.4)$	4.01	[1.31,6.49]	-	-	-	-	-	-
0	$N(0, 0.4)$	4.44	[1.82,6.97]	-	-	-	-	-	-
0	$N(0.4, 0.4)$	4.83	[2.25,7.45]	-	-	-	-	-	-
0	$U(-2, 2)$	4.93	[2.21,7.71]	-	-	-	-	-	-
2.5	$N(-0.4, 0.4)$	3.70	[-0.24,6.90]	-	-	2.38	[-0.62,11.69]	0.11	[0.00,0.43]
2.5	$N(0, 0.4)$	4.25	[0.92,7.35]	-	-	2.28	[-0.62,11.02]	0.10	[0.00,0.41]
2.5	$N(0.4, 0.4)$	4.90	[1.72,8.12]	-	-	2.19	[-0.64,11.00]	0.10	[0.00,0.41]
2.5	$U(-2, 2)$	5.01	[1.31,8.78]	-	-	2.35	[-0.63,11.46]	0.11	[0.00,0.42]
5	$N(-0.4, 0.4)$	3.26	[-1.42,6.80]	-	-	2.73	[-0.31,11.67]	0.14	[0.00,0.48]
5	$N(0, 0.4)$	4.11	[0.19,7.49]	-	-	2.37	[-0.30,11.11]	0.13	[0.00,0.47]
5	$N(0.4, 0.4)$	4.88	[1.30,8.61]	-	-	2.34	[-0.30,10.99]	0.13	[0.00,0.46]
5	$U(-2, 2)$	5.05	[0.91,9.58]	-	-	2.45	[-0.30,11.45]	0.13	[0.00,0.48]
$U(0, 10)$	$N(-0.4, 0.4)$	3.15	[-1.73,6.58]	5.01	[0.29,9.78]	2.81	[-0.79,12.25]	0.14	[0.00,0.51]
$U(0, 10)$	$N(0, 0.4)$	4.12	[0.19,7.52]	4.74	[0.27,9.77]	2.41	[-1.12,11.45]	0.12	[0.00,0.47]
$U(0, 10)$	$N(0.4, 0.4)$	4.85	[1.34,8.44]	4.74	[0.27,9.72]	2.32	[-1.20,11.01]	0.12	[0.00,0.46]
$U(0, 10)$	$U(-2, 2)$	5.03	[1.06,9.45]	4.71	[0.29,9.72]	2.50	[-0.99,11.48]	0.13	[0.00,0.48]

Table 6: Posterior Beliefs (continued)

Panel D: Size

Prior		μ (%)		H (years)		γ (%)		$\hat{\sigma}_{sr}$	
H	μ_{sr}	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
0	$N(-0.4, 0.4)$	-0.42	[-4.16,3.28]	-	-	-	-	-	-
0	$N(0, 0.4)$	0.10	[-3.55,3.84]	-	-	-	-	-	-
0	$N(0.4, 0.4)$	0.73	[-3.04,4.50]	-	-	-	-	-	-
0	$U(-2, 2)$	0.11	[-3.70,4.02]	-	-	-	-	-	-
2.5	$N(-0.4, 0.4)$	-1.86	[-8.62,4.57]	-	-	10.85	[3.20,17.08]	0.40	[0.15,0.56]
2.5	$N(0, 0.4)$	-0.09	[-6.56,6.25]	-	-	10.66	[3.04,16.96]	0.39	[0.14,0.56]
2.5	$N(0.4, 0.4)$	1.81	[-4.57,8.45]	-	-	10.70	[3.03,16.85]	0.39	[0.14,0.56]
2.5	$U(-2, 2)$	-0.12	[-7.83,7.88]	-	-	10.93	[3.23,16.95]	0.40	[0.15,0.56]
5	$N(-0.4, 0.4)$	-2.46	[-10.12,4.84]	-	-	10.90	[1.88,17.67]	0.45	[0.12,0.64]
5	$N(0, 0.4)$	-0.03	[-7.41,7.28]	-	-	10.63	[1.40,17.59]	0.44	[0.09,0.64]
5	$N(0.4, 0.4)$	2.44	[-4.87,9.92]	-	-	10.86	[1.82,17.59]	0.45	[0.12,0.64]
5	$U(-2, 2)$	-0.07	[-10.66,10.28]	-	-	11.00	[1.67,17.70]	0.45	[0.11,0.64]
$U(0, 10)$	$N(-0.4, 0.4)$	-2.00	[-9.14,4.55]	3.39	[0.78,8.98]	10.49	[2.27,17.05]	0.39	[0.11,0.62]
$U(0, 10)$	$N(0, 0.4)$	0.04	[-6.71,6.65]	3.32	[0.74,8.99]	10.31	[2.00,17.08]	0.38	[0.10,0.62]
$U(0, 10)$	$N(0.4, 0.4)$	1.96	[-4.53,9.16]	3.47	[0.83,9.13]	10.47	[2.08,17.05]	0.39	[0.10,0.63]
$U(0, 10)$	$U(-2, 2)$	-0.04	[-9.04,9.00]	3.51	[0.78,9.12]	10.64	[2.35,17.30]	0.40	[0.11,0.63]

Table 7: Out of Sample Bayesian Forecasts

This table presents evidence on the out-of-sample forecasting performance of Bayesian forecasts of characteristic-sorted portfolio returns, along with several benchmarks. Our main tests, in columns (1) through (6), use Fama-MacBeth regressions of next-quarter portfolio returns ($r_{i,t+1}$) on forecasts derived from our Bayesian framework (Bayesian $\hat{r}_{i,t}$), the average past quarterly return in the full prior sample (All past $\overline{r}_i$), the average quarterly return in the past five years (Past five years $\overline{r}_i$), and the average quarterly return in the past year (Past year $\overline{r}_i$). Each of these forecasts uses only past quarterly return data for the portfolio in question, made market neutral using past quarterly market portfolio returns. Realized next-quarter returns are net of the past beta times the realized market return. In column (7), we regress the Bayesian forecast on the three contemporaneous past return averages. Standard errors are in parenthesis. The sample period is from Q3 of 1973 through Q4 of 2021, with the number of available characteristic-sorted portfolios varying from 132 to 173.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LHS	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	$r_{i,t+1}$	Bayesian $\hat{r}_{i,t}$
Bayesian $\hat{r}_{i,t}$	0.73***	-	-	-	0.71**	-	-
	(0.15)				(0.32)	-	
All past $\overline{r}_i$	-	0.65***	-	-	0.26	0.60***	0.47***
		(0.11)			(0.17)	(0.11)	(0.01)
Past five years $\overline{r}_i$	-	-	0.28***	-	-0.24*	-0.10	0.23***
			(0.09)		(0.13)	(0.12)	(0.01)
Past-year $\overline{r}_i$	-	-	-	0.16***	0.06	0.11*	0.05***
				(0.05)	(0.06)	(0.06)	(0.00)

Appendix A. Autocorrelations in Monthly Returns

As described in Section 2.1, we analyze quarterly returns of characteristic-sorted portfolios rather than the monthly returns that are typically studied in the literature. We do so because monthly returns exhibit strong first-order autocorrelations, which may be caused by lead-lag effects, under-reaction, or some other transitory source of persistence. While interesting on its own, this form of autocorrelation differs from the main focus of this paper, namely, slow-moving but persistent variations in expected returns. To avoid biasing our estimates towards large but quickly-reverting variations, we use quarterly rather than monthly data.

Appendix Figure 1 illustrates the magnitude of the autocorrelations at monthly lags $l = 1$ through $l = 60$ for the four portfolios we study, as well as the autocorrelations estimated in a pooled regression including all four portfolios. The first-order autocorrelation is the single largest coefficient for any of the 60 months for value, investment, and profitability, as well as in pooled regressions. Quarterly data do not exhibit a strong first-order autocorrelation (see Table 2) because the two- and three-month autocorrelations in Appendix Figure 1 are much smaller and statistically insignificant.

Appendix B. Small-Sample Bias in Newey and West (1987) Standard Errors

As described in Section 3.1 and presented in Table 3, Newey and West (1987) standard errors differ very little from unadjusted standard errors in our setting. In this Appendix, we show that this pattern can be explained by a small-sample bias in Newey and West (1987) when expected returns have persistent variations.[32]

To illustrate the small-sample performance of Newey and West (1987), we simulate samples using the statistical model presented in Section 1 under a variety of parameterizations for γ, the first-order autocorrelation of returns, and H, the half-life of shocks to expected returns.[33] For each parameterization, we simulate 50,000 samples with 234 quarterly returns, and regress realized returns on a constant. We compute standard errors using Newey and West (1987) with lags equal to $2H$ and compare these to simulated standard errors calculated as the standard deviation across simulations.

Panel A of Appendix Table 2 shows that Newey and West (1987) standard errors exhibit a downward bias in this setting. The bias is stronger when autocorrelation is larger (higher γ) or more persistent (higher H), with Newey and West (1987) standard errors often 25%–50% too small relative to simulated standard errors. This leads the Newey and West (1987) procedure to reject the null too frequently, with t-statistics above the 5% critical value in 15%-40% of simulations. Interestingly, there is some bias even when the true autocorrelation γ is zero. The reason is that Newey and West (1987) embeds a small-sample estimate of autocorrelation, which is downward biased. This effect worsens as H increases because of the increase in the number of lags used in the Newey and West (1987) adjustment.

[32]Hodrick (1992) documents a similar small-sample bias for Newey and West (1987) inferences in a time-series return predictability setting.

[33]We also assume the unconditional expected return, μ, is zero, and the return volatility is equal the to in-sample volatility of the value portfolio.

A potential fix for the downward bias is to increase the number of lags used in Newey and West (1987) to reflect that autocorrelation persists beyond $2H$ periods in our statistical model. Panel B of Appendix Table 2 shows this remedy does not help. Increasing the number of lags slightly reduces the over-rejection initially but eventually has the opposite effect, and the rejection rate never approaches 5%.

To better understand the source of the downward bias, we sort simulated samples by the estimate for $\hat{b}_{20}$ from Equation 10, a summary measure of how much estimated autocorrelation there is in the sample. Panel C of Table 2 shows that the estimate of $\hat{b}_{20}$ is downward biased, averaging around 0.20 below the true value in the two parameterizations we report.[34] The estimates vary substantially across simulations with both parameterizations as well, with interquartile ranges of 1.0 and 0.8, respectively. These variations are closely related to the downward bias in Newey and West (1987), which is much larger in the lower $\hat{b}_{20}$ quintiles than the higher ones.

Appendix C. Generalized Least Squares (GLS) Estimations

In addition to the OLS standard error correction described in Section 3.1, we estimate unconditional expected returns μ in GLS regressions that use the covariance matrix Σ implied by γ and H. These GLS estimations adjust both the point estimates $\hat{\mu}$ and the standard errors. While the OLS estimates are based on an equally-weighted average of the returns in the sample, the GLS estimates utilize an average weighted by the amount of orthogonal information each observation contains about the unconditional expected return. When conditional expected returns exhibit persistent variations, the observations in the middle of the sample are relatively more redundant because they 'over-sample' the same epoch of conditional expected returns. As illustrated by Appendix Figure 2, GLS therefore over-weights observations at the beginning and end of the sample. This effect is larger for larger values of H, and reverses when γ is negative.

Appendix Table 3 shows the GLS point estimates and t-statistics, along with the corresponding OLS statistics, for a variety of assumptions about γ and H. We find that the GLS corrections to point estimates are generally small, but somewhat more negative in some specifications for the value and the investment portfolios. The reason for this negative effect is that these portfolios had unusually low returns at the beginning and/or the end of the sample, which GLS infers as providing more independent information relative to the observations in the middle. Overall, the main conclusion from the OLS analysis, that t-statistics can be half as large for reasonable values of H and γ relative to uncorrected t-statistics, remains unchanged when using GLS.

[34] The true value for a given parameterization is $b_{20} = \frac{\mathrm{Cov}\left(r_t, \frac{1}{20}\sum_{l=1}^{20} r_{t-l}\right)}{\mathrm{Var}\left(\frac{1}{20}\sum_{l=1}^{20} r_{t-l}\right)} = 20\frac{\sum_{l=1}^{20}\Sigma_{1,1+l}}{\sum_{l=1}^{20}\sum_{m=1}^{20}\Sigma_{l,m}}$, where $\Sigma_{i,j}$ is the covariance matrix defined in Equation (8).

Appendix D. Sampling Bayesian Posteriors

We draw samples of $N = 50,000$ observations from the posterior distribution of model parameters Ω^{post} using the following procedure:

1. Draw N observations Ω_i^{prior}, $i \in [1, N]$ from the prior distribution.
2. Accept Ω_1^{prior} as the first observation of the posterior distribution $\Omega_1^{\text{posterior}}$.
3. For observations $i = 2 \dots N$:
   (a) Evaluate the conditional likelihood of the data D given the ith draw from the prior parameters as well as the $i-1$st draw from the posterior parameters:

$$\mathcal{L}^{\text{propose}} = \mathcal{L}(D|\Omega = \Omega_i^{\text{prior}}), \tag{17}$$

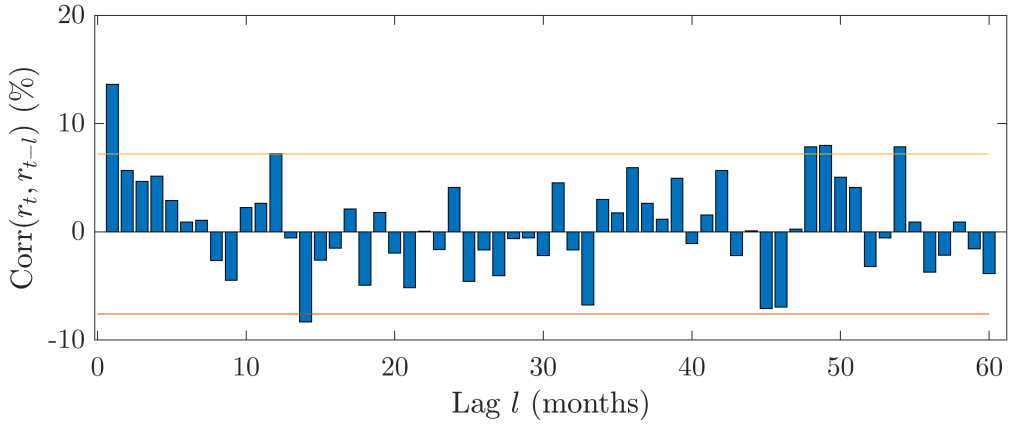
$$\mathcal{L}^{\text{previous}} = \mathcal{L}(D|\Omega = \Omega_{i-1}^{\text{posterior}}). \tag{18}$$

   (b) If $\mathcal{L}^{\text{propose}} \geq \mathcal{L}^{\text{previous}}$, accept $\Omega_i^{\text{posterior}} = \Omega_i^{\text{prior}}$.
   (c) If $\mathcal{L}^{\text{propose}} \leq \mathcal{L}^{\text{previous}}$, accept $\Omega_i^{\text{posterior}} = \Omega_i^{\text{prior}}$ with probability $\frac{\mathcal{L}^{\text{propose}}}{\mathcal{L}^{\text{previous}}}$, and otherwise retain $\Omega_i^{\text{posterior}} = \Omega_{i-1}^{\text{prior}}$.

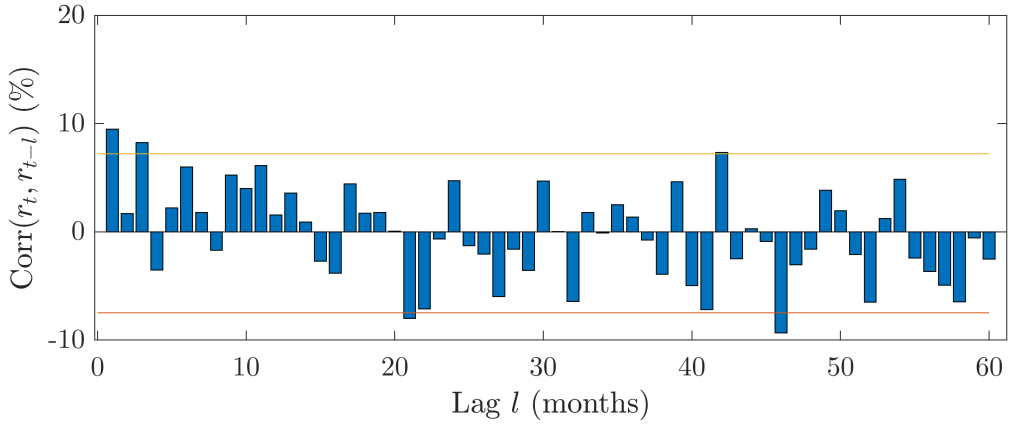
Appendix Figure 1: Monthly Autocorrelograms

This figure presents autocorrelations of monthly returns for value, investment, profitability, and size portfolios, as defined in the header of Table 1, as well the autocorrelation estimated in a pooled regression containing all four portfolios. We estimate the autocorrelation for each lag l independently. The horizontal lines represent the 95% confidence interval for autocorrelation coefficients under the zero-autocorrelation null hypothesis. Our sample consists of 678 monthly observations from Q3 1963 through Q4 2021.

Panel A: Value

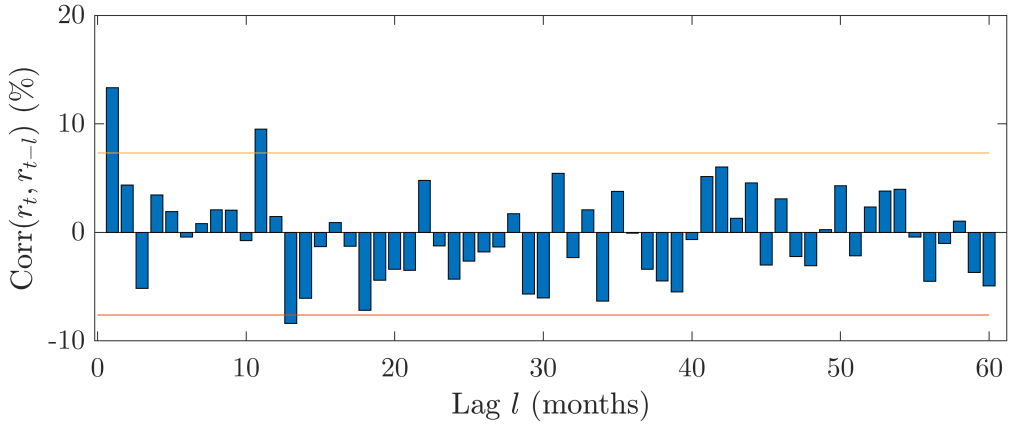


Panel B: Investment

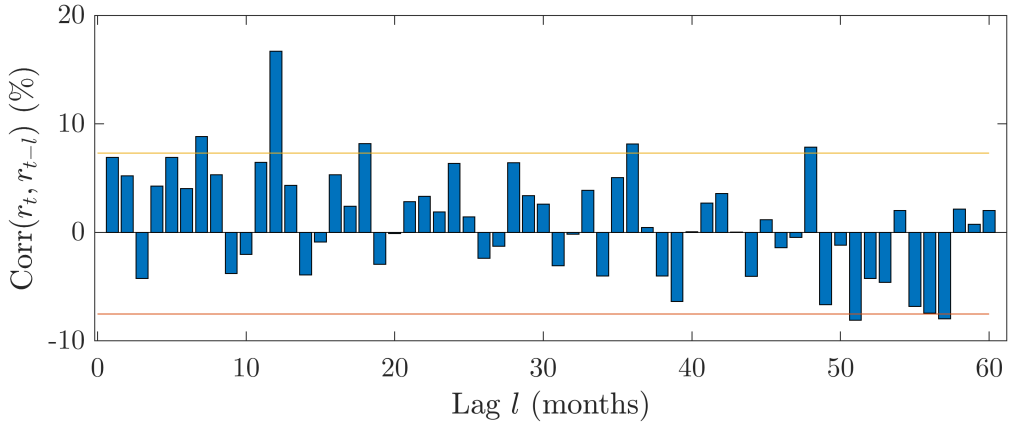


Appendix Figure 1: Monthly Autocorrelograms (continued)

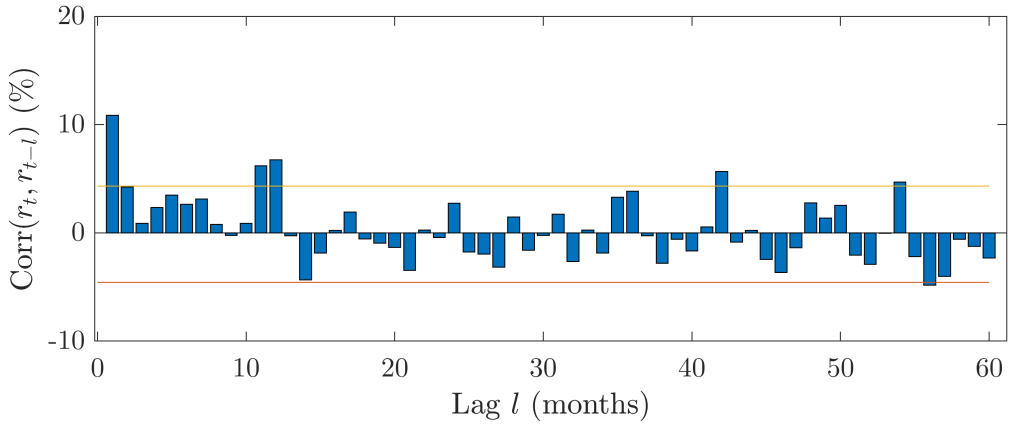
Panel C: Profitability



Panel D: Size

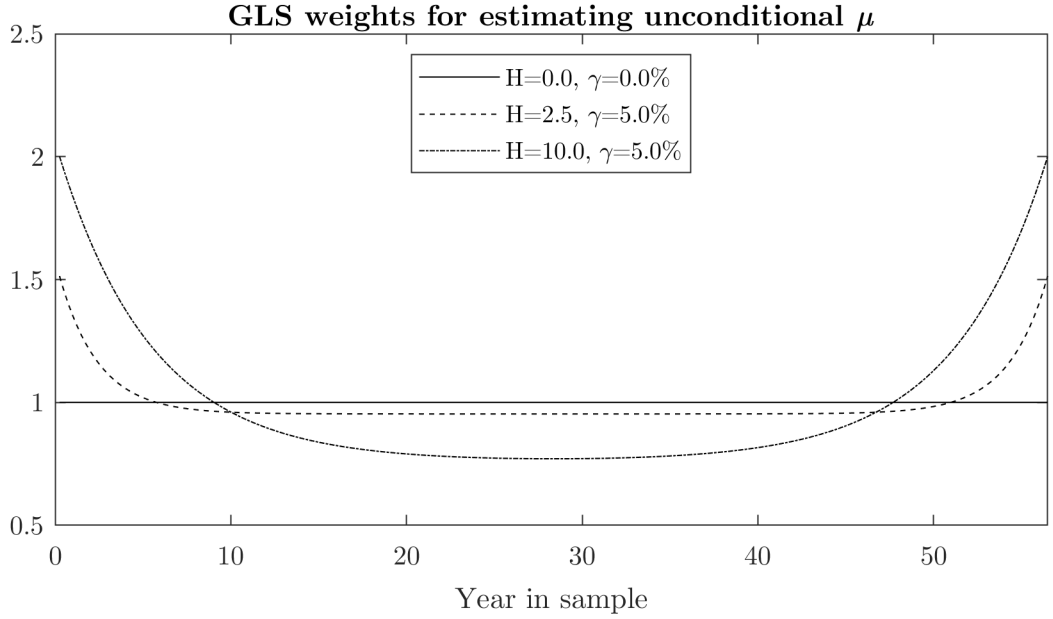


Panel E: Pooled



Appendix Figure 2: GLS Influence Functions and Return Autocorrelations

This figure presents the GLS weights that apply in the estimation of the unconditional expected return μ under different assumptions about the autocorrelation structure of returns. The parameter H is the half-life of shocks to expected returns in years, and γ is the first-order autocorrelation of quarterly returns.



Appendix Table 1: Return Autocorrelations for Different Parameterizations

This table shows the first-order autocorrelation of quarterly returns in percentage terms, $\gamma = corr(r_t, r_{t-1})$, implied by various combinations of H, the half-life of shocks to expected returns, and $\sigma_\delta / \sigma_\epsilon$, the ratio of the volatilities of expected and unexpected return shocks. We assume $\rho = -1$ throughout to provide an upper bound on the magnitude negative autocorrelations.

H	$\frac{\sigma_\delta}{\sigma_\epsilon}$						
	0.001	0.01	0.1	0.25	0.5	0.75	1
0.1	-0.0	-0.2	-1.8	-4.4	-8.6	-12.7	-16.6
0.25	-0.0	-0.5	-4.8	-11.2	-19.2	-23.7	-25.0
0.5	-0.1	-0.7	-6.3	-12.5	-14.1	-8.5	0.0
1	-0.1	-0.8	-6.2	-7.2	5.4	21.7	34.8
2.5	-0.1	-0.9	-2.9	11.2	41.1	59.2	69.1
5	-0.1	-0.8	3.3	32.1	64.3	77.5	83.7
10	-0.1	-0.7	14.1	54.0	80.1	88.1	91.6
25	-0.1	-0.3	35.7	76.6	91.4	95.1	96.6

Appendix Table 2: Downward Bias in Newey and West (1987) Standard Errors

This table uses small-sample simulations to demonstrate how the downward bias in Newey and West (1987) standard errors varies depending on the true data generating process, number of lags used, and realized sample autocorrelation. Panel A shows how the bias varies as a function of H, the half life of shocks to conditional expected returns (in years), and γ, the first-order autocorrelation of returns. Each of the 50,000 simulations has 234 quarters, zero unconditional expected return μ, and return volatility equal to the volatility of the value portfolio. For each H and γ pair, Panel A presents Newey and West (1987) standard errors for μ with $2H$ years of lags (NW SE) relative to the standard deviation across simulations of estimates for μ (Sim. SE), and the fraction of simulated samples in which the t-statistic for $\mu = 0$ is above 1.96. Panel B presents the same summary statistics when the Newey and West (1987) standard errors are calculated using a variety of different lags. Panel C shows how the bias statistics when using $2H$ years of lags vary across simulated samples sorted by the in-sample estimate of b_{20}, the coefficient in a regression of r_t on $\frac{1}{20}\sum_{l=1}^{20} r_{t-l}$.

Panel A: Bias in Newey and West (1987) standard errors with $2H$ years of lags as a function of H and γ

		γ				
H		0%	1%	2.5%	5%	10%
0.5	NW SE/Sim. SE	0.98	0.97	0.95	0.91	0.87
	$\lvert$NW t-stat$\rvert > 1.96$	5.8%	6.0%	6.8%	7.9%	9.4%
1	NW SE/Sim. SE	0.98	0.94	0.91	0.86	0.81
	$\lvert$NW t-stat$\rvert > 1.96$	6.0%	7.1%	8.2%	9.7%	12.1%
2.5	NW SE/Sim. SE	0.94	0.88	0.82	0.76	0.70
	$\lvert$NW t-stat$\rvert > 1.96$	7.7%	9.8%	12.1%	15.0%	18.5%
5	NW SE/Sim. SE	0.89	0.79	0.72	0.66	0.61
	$\lvert$NW t-stat$\rvert > 1.96$	10.8%	14.8%	18.7%	22.5%	25.8%
10	NW SE/Sim. SE	0.79	0.65	0.57	0.52	0.48
	$\lvert$NW t-stat$\rvert > 1.96$	16.3%	24.3%	30.1%	34.6%	38.1%

Panel B: Varying number of lags

			Number of lags (years)				
H	γ		1	3	5	10	20
5	0%	NW SE/Sim. SE	0.99	0.96	0.94	0.89	0.79
		$\lvert$NW t-stat$\rvert > 1.96$	5.7%	6.8%	7.9%	10.8%	16.3%
5	5%	NW SE/Sim. SE	0.56	0.60	0.63	0.66	0.64
		$\lvert$NW t-stat$\rvert > 1.96$	28%	25%	23%	23%	26%

Panel C: Simulated samples sorted by in-sample autocorrelation

					$\hat{b}_{20}$ quintile				
H	γ	b_{20}		All	1	2	3	4	5
5	0%	0.00	$\hat{b}_{20}$	-0.20	-0.75	-0.35	-0.15	0.02	0.24
			NW SE/Sim. SE	0.89	0.65	0.77	0.87	0.97	1.17
			$\lvert$NW t-stat$\rvert > 1.96$	10.8%	20.3%	14.2%	10.2%	6.6%	2.9%
5	5%	0.41	$\hat{b}_{20}$	0.18	-0.32	0.05	0.22	0.37	0.55
			NW SE/Sim. SE	0.66	0.44	0.55	0.64	0.74	0.92
			$\lvert$NW t-stat$\rvert > 1.96$	22.5%	39.8%	28.0%	21.4%	15.3%	8.2%

Appendix Table 3: GLS with Time-Varying Expected Returns

This table presents estimates of unconditional expected returns of value, investment, profitability, and size portfolios, as defined in the header of Table 1, under a variety of assumptions about the magnitude and persistence of variations in conditional expected returns. The model-implied autocorrelation structure of returns are summarized by H, the half-life of shocks to expected returns in years, and γ, the first-order autocorrelation of quarterly returns. We estimate unconditional expected returns μ using both OLS and GLS and calculate t-statistics using the model-implied correlation matrix of the regression error terms. Our sample consists of 234 quarterly observations from Q3 1963 through Q4 2021.

Panel A: Value

	H (years)	2.5			5		10	
	γ (%)	2.5	5	10	2.5	5	2.5	5
OLS	$\hat{\mu}$ (%)	3.80	3.80	3.80	3.80	3.80	3.80	3.80
	Model t-stat	(1.47)	(1.24)	(0.99)	(1.27)	(1.02)	(1.07)	(0.83)
GLS	$\hat{\mu}$ (%)	3.68	3.62	3.57	3.50	3.36	3.28	3.03
	Model t-stat	(1.43)	(1.19)	(0.93)	(1.18)	(0.91)	(0.94)	(0.67)

Panel B: Investment

	H (years)	2.5			5		10	
	γ (%)	2.5	5	10	2.5	5	2.5	5
OLS	$\hat{\mu}$ (%)	4.75	4.75	4.75	4.75	4.75	4.75	4.75
	Model t-stat	(2.62)	(2.21)	(1.75)	(2.26)	(1.81)	(1.91)	(1.47)
GLS	$\hat{\mu}$ (%)	4.65	4.62	4.63	4.46	4.35	4.23	4.00
	Model t-stat	(2.57)	(2.15)	(1.72)	(2.14)	(1.68)	(1.72)	(1.26)

Panel C: Profitability

	H (years)	2.5			5		10	
	γ (%)	2.5	5	10	2.5	5	2.5	5
OLS	$\hat{\mu}$ (%)	4.91	4.91	4.91	4.91	4.91	4.91	4.91
	Model t-stat	(2.67)	(2.25)	(1.79)	(2.30)	(1.84)	(1.94)	(1.50)
GLS	$\hat{\mu}$ (%)	5.04	5.16	5.36	5.07	5.23	5.04	5.23
	Model t-stat	(2.74)	(2.37)	(1.96)	(2.39)	(1.98)	(2.02)	(1.62)

Panel D: Size

	H (years)	2.5			5		10	
	γ (%)	2.5	5	10	2.5	5	2.5	5
OLS	$\hat{\mu}$ (%)	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	Model t-stat	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.03)
GLS	$\hat{\mu}$ (%)	0.16	0.13	0.08	-0.04	0.16	0.12	0.23
	Model t-stat	(0.05)	(0.03)	-(0.01)	(0.05)	(0.03)	(0.06)	(0.04)

Appendix Table 4: Inference and Posteriors for 174 Portfolios

This table presents summary statistics and estimation results for the four portfolios that we focus on in our main analysis (the first four rows) and 170 long-short equity portfolios analyzed by Chen and Zimmermann (2021) (the remaining rows). The first two columns report $E(r_t)$, the in-sample average of annualized quarterly returns, and SR, the in-sample annualized Sharpe ratio. The next two columns report the p-values for the hypothesis that the unconditional expected return μ equals zero using OLS regressions that assume i.i.d. returns and maximum likelihood estimations (MLE) that allow for time-varying expected returns. The next four columns report the means of Bayesian posterior distributions for $P(\mu > 0)$, the probability that $\mu > 0$; μ_{sr}, the unconditional Sharpe ratio; $\mu_{sr,2022}$, the conditional Sharpe ratio for Q1 of 2022; and $\hat{\sigma}_{sr}$, the volatility of the Sharpe ratio conditional on past returns. The last column reports ρ(size), which denotes the correlation of each portfolio's returns with the size portfolio's return.

			$\mu = 0$ p-values		Bayesian Posterior Means				
Portfolio	$E(r_t)$	SR	OLS	MLE	$P(\mu > 0)$	μ_{sr}	$\mu_{sr,2022}$	$\hat{\sigma}_{sr}$	ρ(size)
Value (FF)	3.80	0.25	5.4%	10.4%	87.8%	0.20	0.10	0.14	0.40
Investment (FF)	4.75	0.45	0.1%	1.2%	95.6%	0.36	0.37	0.17	0.04
Profitability (FF)	4.91	0.45	0.0%	1.4%	98.0%	0.39	0.60	0.12	-0.52
Size (FF)	0.16	0.01	93.5%	97.5%	50.9%	0.00	-0.30	0.38	1.00
AM	1.34	0.08	54.6%	61.4%	65.6%	0.06	-0.01	0.12	0.18
AOP	3.88	0.34	2.0%	3.7%	93.3%	0.28	0.39	0.12	-0.26
AbnormalAccruals	4.92	0.53	0.0%	0.0%	99.0%	0.45	0.53	0.08	0.04
Accruals	4.59	0.47	0.0%	0.0%	99.5%	0.42	0.43	0.05	-0.19
Activism1	-0.59	-0.10	68.7%	69.7%	38.3%	-0.07	-0.13	0.19	0.20
AdExp	5.79	0.38	0.7%	4.4%	89.0%	0.27	0.13	0.22	0.11
AgeIPO	13.77	0.59	0.0%	0.3%	97.5%	0.46	0.50	0.12	-0.30
AnalystRevision	5.40	0.57	0.0%	1.2%	97.9%	0.47	0.57	0.13	-0.14
AnalystValue	5.39	0.32	3.1%	9.1%	90.9%	0.25	0.28	0.15	-0.11
AnnouncementReturn	8.37	1.05	0.0%	0.1%	99.3%	0.82	0.99	0.18	-0.07
AssetGrowth	4.38	0.40	0.2%	5.5%	91.5%	0.29	0.24	0.21	0.17
BM	2.30	0.11	41.7%	48.6%	71.1%	0.08	0.00	0.12	0.36
BMdec	3.78	0.27	3.9%	10.9%	82.3%	0.18	-0.07	0.25	0.33
BPEBM	1.75	0.25	5.5%	3.6%	92.9%	0.23	0.32	0.10	0.22
Beta	-7.13	-0.36	0.6%	3.7%	6.5%	-0.29	-0.49	0.19	0.60
BetaFP	-7.70	-0.53	0.0%	1.2%	3.5%	-0.41	-0.41	0.16	0.31
BetaLiquidityPS	1.56	0.16	24.9%	18.8%	74.5%	0.11	-0.03	0.15	0.02
BetaTailRisk	-0.14	-0.01	93.3%	93.0%	47.7%	-0.01	-0.22	0.24	0.42
BetaVIX	8.38	0.72	0.0%	2.3%	91.3%	0.41	0.53	0.34	-0.29
BidAskSpread	-10.93	-0.46	0.0%	6.2%	4.9%	-0.36	-0.51	0.19	0.68
BookLeverage	-0.48	-0.03	80.3%	82.6%	42.7%	-0.02	-0.03	0.08	0.11
BrandInvest	-0.86	-0.06	70.2%	32.5%	38.9%	-0.04	-0.07	0.09	0.41
CBOperProf	6.87	0.74	0.0%	0.4%	99.4%	0.62	0.93	0.15	-0.52
CF	6.35	0.44	0.1%	0.5%	97.6%	0.37	0.38	0.10	0.09
Cash	2.84	0.20	16.2%	3.5%	86.3%	0.17	0.15	0.08	0.21
CashProd	1.56	0.12	34.4%	43.4%	75.6%	0.10	0.04	0.10	0.08
Cfp	6.85	0.50	0.0%	0.2%	98.7%	0.42	0.45	0.07	-0.53
ChAssetTurnover	2.06	0.26	4.6%	4.0%	91.4%	0.22	0.16	0.10	0.03
ChEQ	5.32	0.50	0.0%	0.1%	99.0%	0.43	0.46	0.08	0.06
ChInv	5.78	0.70	0.0%	0.2%	98.7%	0.56	0.59	0.17	-0.09
ChInvIA	3.30	0.41	0.2%	0.0%	99.2%	0.38	0.58	0.13	-0.09
ChNNCOA	3.01	0.42	0.1%	0.4%	98.0%	0.36	0.34	0.08	-0.02
ChNWC	2.89	0.39	0.3%	1.4%	95.5%	0.32	0.29	0.14	-0.06
ChTax	3.68	0.38	0.4%	8.1%	95.6%	0.32	0.24	0.14	-0.02
ChangeInRecommendation	2.15	0.28	13.8%	24.9%	80.2%	0.20	0.39	0.22	-0.06
CompEquIss	3.85	0.30	2.3%	0.3%	95.8%	0.26	0.30	0.07	-0.09
CompositeDebtIssuance	1.86	0.25	5.9%	10.4%	84.4%	0.18	0.03	0.18	0.10
CoskewACX	0.67	0.06	65.5%	57.7%	65.7%	0.05	0.04	0.09	0.20
Coskewness	1.71	0.18	18.0%	8.0%	82.8%	0.15	0.13	0.11	-0.02
CustomerMomentum	9.93	0.39	1.3%	5.4%	94.0%	0.31	0.31	0.12	0.09
DNoa	6.26	0.72	0.0%	0.2%	96.6%	0.51	0.45	0.28	-0.01
DelBreadth	1.90	0.19	21.6%	2.1%	83.4%	0.16	0.20	0.08	0.06
DelCOA	2.70	0.27	4.0%	7.9%	91.4%	0.22	0.20	0.11	-0.13
DelCOL	-0.91	-0.08	53.2%	23.1%	29.2%	-0.08	-0.10	0.06	-0.02
DelDRC	4.51	0.40	7.5%	3.3%	84.6%	0.25	0.30	0.16	-0.15

Appendix Table 4: Inference and Posteriors for 174 Portfolios (continued)

			$\mu = 0$ p-values		Bayesian Posterior Means				
Portfolio	$E(r_t)$	SR	OLS	MLE	$P(\mu > 0)$	μ_{sr}	$\mu_{sr,2022}$	$\hat{\sigma}_{sr}$	ρ(size)
DelEqu	3.47	0.30	2.3%	0.6%	94.8%	0.26	0.29	0.09	0.15
DelFINL	3.91	0.63	0.0%	1.2%	97.8%	0.50	0.55	0.18	0.08
DelLTI	1.75	0.33	1.2%	2.5%	95.2%	0.28	0.44	0.12	0.04
DelNetFin	1.56	0.22	9.3%	13.8%	86.7%	0.18	0.06	0.12	-0.08
DolVol	1.25	0.09	47.9%	31.6%	66.0%	0.06	-0.10	0.18	0.56
EBM	2.61	0.22	8.7%	13.5%	87.2%	0.18	0.15	0.13	0.24
EP	3.47	0.25	5.3%	1.7%	92.0%	0.22	0.25	0.09	0.22
EarnSupBig	3.43	0.25	5.5%	0.7%	94.5%	0.22	0.22	0.06	-0.14
EarningsConsistency	3.42	0.27	4.0%	1.8%	94.5%	0.24	0.21	0.10	0.10
EarningsForecastDisparity	3.68	0.25	12.0%	0.2%	89.7%	0.20	0.20	0.08	-0.03
EarningsStreak	8.37	0.93	0.0%	0.0%	98.9%	0.71	0.82	0.14	-0.17
EarningsSurprise	2.43	0.33	1.3%	0.4%	96.5%	0.29	0.29	0.09	-0.05
EntMult	5.81	0.39	0.3%	1.5%	95.3%	0.32	0.26	0.13	0.08
EquityDuration	7.81	0.48	0.0%	1.2%	95.7%	0.37	0.36	0.18	0.01
ExclExp	2.68	0.54	0.1%	0.3%	97.2%	0.42	0.46	0.10	-0.33
FEPS	11.88	0.77	0.0%	0.0%	98.8%	0.61	0.87	0.15	-0.63
FR	1.19	0.15	35.3%	50.1%	71.2%	0.14	0.54	0.37	-0.17
Fgr5yrLag	7.67	0.43	0.7%	1.7%	96.1%	0.35	0.42	0.12	-0.31
FirmAge	-1.19	-0.14	28.4%	31.3%	22.4%	-0.12	-0.30	0.15	0.50
FirmAgeMom	18.00	0.78	0.0%	0.9%	98.6%	0.60	0.95	0.24	0.00
ForecastDispersion	6.11	0.38	1.0%	0.6%	95.5%	0.32	0.42	0.10	-0.56
Frontier	1.80	0.09	51.2%	13.9%	67.4%	0.07	-0.04	0.12	0.52
GP	4.47	0.39	0.3%	1.5%	96.1%	0.34	0.58	0.15	-0.16
GrAdExp	3.47	0.29	4.6%	2.6%	92.0%	0.24	0.28	0.11	-0.08
GrLTNOA	1.71	0.21	10.6%	0.2%	90.0%	0.18	0.19	0.07	-0.16
GrSaleToGrInv	3.02	0.37	0.5%	12.2%	89.3%	0.28	0.29	0.29	0.07
GrSaleToGrOverhead	0.63	0.07	61.8%	40.3%	66.1%	0.06	0.04	0.06	0.07
Grcapx	4.08	0.41	0.2%	1.7%	95.2%	0.33	0.28	0.16	0.16
Grcapx3y	2.93	0.29	2.8%	6.4%	91.7%	0.24	0.14	0.13	0.18
Herf	0.77	0.07	59.2%	67.5%	65.3%	0.06	0.03	0.10	0.25
HerfAsset	0.85	0.08	54.1%	64.4%	68.3%	0.07	0.07	0.11	0.31
HerfBE	0.79	0.08	56.4%	68.1%	67.7%	0.07	0.04	0.13	0.35
High52	15.21	0.76	0.0%	0.1%	99.5%	0.64	0.70	0.12	-0.44
Hire	2.22	0.22	10.0%	17.3%	84.8%	0.17	0.10	0.13	-0.06
IdioRisk	15.92	0.78	0.0%	1.6%	95.8%	0.51	0.91	0.39	-0.65
IdioVol3F	15.71	0.74	0.0%	0.6%	96.8%	0.52	0.86	0.34	-0.68
IdioVolAHT	13.94	0.56	0.0%	1.5%	95.4%	0.41	0.60	0.25	-0.71
Illiquidity	3.01	0.25	5.1%	46.9%	69.8%	0.12	-0.85	0.45	0.77
IndMom	2.88	0.19	13.8%	18.1%	82.9%	0.16	0.19	0.18	0.16
IndRetBig	20.46	1.26	0.0%	1.9%	98.0%	0.66	0.48	0.46	0.06
IntMom	11.92	0.62	0.0%	3.1%	95.5%	0.44	0.16	0.24	-0.09
IntanBM	0.31	0.02	89.5%	95.5%	52.7%	0.01	-0.17	0.15	0.41
IntanCFP	2.43	0.14	28.7%	10.4%	79.0%	0.11	0.05	0.08	0.22
IntanEP	1.92	0.12	36.2%	10.3%	77.0%	0.10	0.07	0.07	0.31
IntanSP	-1.14	-0.06	66.8%	71.0%	37.8%	-0.05	-0.13	0.15	0.56
InvGrowth	5.14	0.49	0.0%	5.1%	96.1%	0.39	0.49	0.19	-0.03
InvestPPEInv	4.91	0.49	0.0%	0.6%	94.4%	0.37	0.21	0.22	0.10
Investment	1.16	0.12	34.0%	40.8%	79.0%	0.11	0.08	0.07	0.37
LRreversal	1.68	0.09	51.3%	61.2%	67.1%	0.07	0.04	0.15	0.48
Leverage	1.83	0.10	43.5%	51.2%	72.2%	0.08	0.05	0.11	0.10
MRreversal	3.59	0.24	6.5%	18.4%	85.2%	0.19	0.10	0.20	0.20
MaxRet	11.38	0.65	0.0%	0.3%	97.3%	0.49	0.75	0.30	-0.59
MeanRankRevGrowth	3.62	0.35	0.7%	2.1%	95.2%	0.29	0.25	0.10	0.05
Mom12m	15.26	0.68	0.0%	0.2%	99.4%	0.58	0.56	0.09	-0.15
Mom12mOffSeason	8.86	0.44	0.1%	0.0%	99.3%	0.39	0.40	0.05	-0.06
Mom6m	10.45	0.55	0.0%	0.4%	98.8%	0.47	0.58	0.12	-0.08
Mom6mJunk	5.33	0.30	6.5%	7.4%	86.4%	0.22	0.23	0.15	0.14
MomOffSeason	5.28	0.30	2.0%	7.0%	91.8%	0.24	0.17	0.13	0.20
MomOffSeason06YrPlus	8.79	0.73	0.0%	1.1%	98.5%	0.58	0.69	0.17	0.07
MomOffSeason11YrPlus	3.70	0.35	0.7%	1.0%	96.0%	0.30	0.27	0.09	0.04
MomOffSeason16YrPlus	2.81	0.31	1.9%	0.6%	96.5%	0.27	0.27	0.07	-0.01
MomSeason	4.91	0.42	0.1%	16.4%	83.9%	0.26	0.32	0.48	-0.06
MomSeason06YrPlus	6.15	0.58	0.0%	3.5%	94.7%	0.41	0.38	0.33	0.00

Appendix Table 4: Inference and Posteriors for 174 Portfolios (continued)

			$\mu = 0$ p-values		Bayesian Posterior Means				
Portfolio	$E(r_t)$	SR	OLS	MLE	$P(\mu > 0)$	μ_{sr}	$\mu_{sr,2022}$	$\hat{\sigma}_{sr}$	ρ(size)
MomSeason11YrPlus	5.76	0.61	0.0%	0.1%	99.5%	0.53	0.52	0.08	-0.04
MomSeason16YrPlus	3.09	0.32	1.5%	0.3%	98.0%	0.29	0.27	0.09	-0.02
MomSeasonShort	5.76	0.46	0.0%	14.7%	91.2%	0.32	0.01	0.26	-0.06
NOA	4.23	0.41	0.2%	5.4%	93.1%	0.32	0.34	0.24	-0.06
NetDebtFinance	3.55	0.54	0.0%	0.8%	97.2%	0.43	0.50	0.14	0.00
NetDebtPrice	2.41	0.15	25.3%	2.4%	82.0%	0.13	0.11	0.07	0.17
NetEquityFinance	5.35	0.55	0.0%	0.2%	98.6%	0.46	0.52	0.10	-0.34
NetPayoutYield	7.01	0.53	0.0%	0.2%	97.4%	0.43	0.74	0.17	-0.35
NumEarnIncrease	2.68	0.47	0.0%	0.4%	98.4%	0.40	0.49	0.10	-0.06
OPLeverage	3.41	0.28	2.9%	4.2%	95.4%	0.25	0.27	0.07	0.20
OperProf	6.19	0.46	0.0%	0.9%	97.8%	0.40	0.58	0.15	-0.47
OperProfRD	7.04	0.58	0.0%	1.3%	94.4%	0.42	1.34	0.37	-0.56
OptionVolume1	3.35	0.28	15.7%	32.5%	68.7%	0.12	-0.09	0.31	0.12
OptionVolume2	2.95	0.38	5.6%	8.5%	84.7%	0.24	0.19	0.18	0.01
OrderBacklog	-1.08	-0.07	59.6%	65.3%	36.6%	-0.05	0.06	0.12	-0.09
OrderBacklogChg	-0.52	-0.04	77.8%	55.6%	42.7%	-0.03	-0.06	0.09	-0.05
OrgCap	5.12	0.55	0.0%	0.0%	99.3%	0.48	0.48	0.08	-0.02
PS	8.39	0.42	0.3%	0.2%	96.3%	0.34	0.29	0.09	-0.20
PayoutYield	4.31	0.33	1.1%	2.7%	94.5%	0.28	0.44	0.13	-0.09
PctAcc	2.34	0.26	5.5%	2.5%	94.0%	0.23	0.22	0.08	-0.25
PctTotAcc	0.83	0.10	56.0%	30.6%	67.7%	0.09	0.07	0.10	-0.14
PredictedFE	2.88	0.19	24.9%	33.1%	82.2%	0.16	0.20	0.10	-0.13
Price	-5.90	-0.21	10.7%	31.7%	24.2%	-0.15	-0.35	0.33	0.72
PriceDelayRsq	-1.76	-0.16	22.5%	4.1%	17.8%	-0.14	-0.20	0.11	0.23
PriceDelaySlope	-0.90	-0.09	49.6%	42.7%	31.7%	-0.07	-0.08	0.11	0.32
PriceDelayTstat	0.23	0.03	79.2%	27.6%	58.1%	0.03	0.02	0.09	-0.32
ProbInformedTrading	15.75	1.21	0.0%	0.7%	95.2%	0.62	0.81	0.29	-0.26
RD	1.50	0.09	53.8%	61.0%	64.9%	0.07	0.08	0.20	0.39
RDAbility	5.35	0.26	9.3%	4.8%	87.5%	0.21	0.28	0.13	-0.02
RDS	2.01	0.35	1.9%	5.4%	90.7%	0.26	0.21	0.15	0.13
RDcap	1.63	0.07	64.6%	81.6%	57.6%	0.03	-0.36	0.23	0.47
REV6	4.34	0.20	18.4%	25.2%	75.8%	0.13	-0.06	0.19	-0.19
Realestate	3.22	0.31	2.6%	4.5%	90.1%	0.24	0.19	0.15	-0.19
ResidualMomentum	5.96	0.57	0.0%	0.3%	98.9%	0.48	0.43	0.10	-0.04
RetConglomerate	6.44	0.46	0.3%	0.2%	96.1%	0.37	0.35	0.12	-0.04
ReturnSkew	-0.09	-0.01	91.2%	95.3%	50.3%	0.01	0.32	0.25	-0.20
ReturnSkew3F	-1.77	-0.33	1.1%	17.0%	9.1%	-0.25	-0.19	0.16	-0.23
RevenueSurprise	1.51	0.20	12.4%	7.3%	88.6%	0.18	0.19	0.09	-0.04
RoE	2.88	0.26	4.3%	6.4%	91.1%	0.23	0.34	0.13	-0.53
Roaq	6.65	0.40	0.5%	2.4%	93.2%	0.31	0.51	0.16	-0.50
SP	4.76	0.27	3.8%	8.9%	90.0%	0.22	0.17	0.14	0.39
STreversal	-0.18	-0.01	92.8%	92.3%	46.4%	-0.01	-0.15	0.14	0.16
Sfe	10.93	0.56	0.0%	0.2%	98.4%	0.47	0.49	0.10	-0.30
ShareIss1Y	6.51	0.83	0.0%	0.1%	99.1%	0.65	0.70	0.20	-0.19
ShareIss5Y	4.95	0.59	0.0%	0.1%	98.3%	0.48	0.47	0.17	-0.02
ShortInterest	4.40	0.41	0.4%	0.5%	97.5%	0.35	0.40	0.09	-0.26
Size	-0.05	0.00	98.6%	97.6%	50.5%	0.01	-0.15	0.31	0.82
Skew1	5.89	0.47	1.8%	3.4%	91.4%	0.33	0.41	0.17	-0.10
SmileSlope	13.06	1.34	0.0%	2.2%	95.9%	0.64	0.88	0.40	-0.12
Std_turn	11.35	0.69	0.0%	4.1%	94.3%	0.44	0.49	0.36	-0.50
Tang	0.60	0.05	71.5%	38.8%	62.4%	0.04	0.07	0.07	0.08
Tax	4.34	0.53	0.0%	0.2%	96.7%	0.43	0.78	0.22	-0.27
TotalAccruals	2.48	0.29	2.8%	6.5%	92.8%	0.25	0.34	0.13	0.10
TrendFactor	11.88	0.74	0.0%	4.8%	96.4%	0.50	0.21	0.24	0.15
VarCF	1.48	0.10	44.9%	57.0%	67.9%	0.09	0.30	0.23	-0.71
VolMkt	4.78	0.33	1.1%	0.9%	94.9%	0.28	0.34	0.12	-0.42
VolSD	4.09	0.52	0.0%	0.3%	98.6%	0.44	0.46	0.10	0.17
VolumeTrend	4.68	0.42	0.1%	4.8%	92.4%	0.32	0.49	0.26	0.21
XFIN	9.91	0.70	0.0%	1.3%	97.5%	0.53	0.91	0.23	-0.50
Zerotrade	5.04	0.51	0.0%	0.7%	96.7%	0.40	0.49	0.17	-0.26
ZerotradeAlt1	4.29	0.38	0.3%	6.8%	89.4%	0.28	0.56	0.30	-0.38
ZerotradeAlt12	5.17	0.52	0.0%	1.4%	96.3%	0.41	0.46	0.18	-0.21