

Huefner, Bernd; Rueenaufner, Marcel; Boesch, Martin

Article — Published Version

Value investing via Bayesian inference

Review of Financial Economics

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Huefner, Bernd; Rueenaufner, Marcel; Boesch, Martin (2023) : Value investing via Bayesian inference, Review of Financial Economics, ISSN 1873-5924, Wiley, Hoboken, NJ, Vol. 41, Iss. 4, pp. 465-492, <https://doi.org/10.1002/rfe.1185>

This Version is available at:

<https://hdl.handle.net/10419/288178>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Value investing via Bayesian inference

Bernd Huefner¹ | Marcel Rueenaufner¹  | Martin Boesch²

¹Friedrich Schiller University Jena, Jena, Germany

²Ernst-Abbe Hochschule Jena, Jena, Germany

Correspondence

Marcel Rueenaufner, Friedrich-Schiller University Jena, Carl-Zeiß-Str. 3, 07743 Jena, Germany.

Email: marcel.rueenaufner@uni-jena.de

Abstract

Classic value investing à la Graham & Dodd (Security analysis: The classic, McGrawHill, New York, 1934) focuses on selecting stocks that seem cheap relative to their intrinsic value and fundamental quality. We use Bayesian inference to account for a large amount of uncertainty within intrinsic value estimation. We find that an undervalued-minus-overvalued factor that invests in cheap quality stocks and sells expensive junk stocks selected via Bayesian inference yields high risk-adjusted returns and Sharpe ratios for equal-weighted portfolios. We also find that using value-weighted portfolios introduces size-based dilutions and shifts the focus away from actual quality characteristics like profitability, payout, safety, and past growth. Our findings suggest that while the relative benefit of accounting for uncertainty via Bayesian inference is not large over shorter holding periods, it pays off for investment horizons longer than a month.

KEYWORDS

Bayesian inference, financial statement analysis, residual income valuation, value investing

JEL CLASSIFICATION

M41, G11, G12, G14, G17, G32

1 | INTRODUCTION

Assessing the valuation accuracy and valuation bias of different equity valuation models in research is (almost) as old as the models themselves. Examples of contributions assessing the accuracy and bias of several discounted payoff valuation models are Penman and Sougiannis (1998), Ahmed et al. (2000), Francis et al. (2000), Frankel and Lee (1998), Courteau et al. (2001), Barth et al. (2005), Choi et al. (2006), Dittmann and Maug (2008), Henschke (2009), Jorgensen et al. (2011), Chang et al. (2012), Ho et al. (2017), Anesten et al. (2019) and Gao et al. (2019). Despite the large amount of research, the field has been dealing with several pervasive problems.

Most studies in the field focus on the replication of current stock prices, implying a semi-strong level of market efficiency in accordance with Fama (1970). If market prices are efficient, they form a useful benchmark for intrinsic value estimates. If they are not (e.g., due to a noticeable influence of low-information trading as shown by Dontoh et al. (2007)), then little is gained by replicating them because the benchmark is biased. It is also unclear whether a model that replicates current prices accurately and without bias has a higher predictive utility for subsequent stock returns.¹ In line with Graham and Dodd (1934) and Piotroski and So (2012), our empirical analysis anchors on the convergence assumption

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Review of Financial Economics* published by Wiley Periodicals LLC on behalf of University of New Orleans.

of the market price toward intrinsic value over time, which leads to (expectedly) predictable return patterns. Following Lee (2015) and Asness et al. (2019), firms that justify high (low) intrinsic values relative to their price should yield higher (lower) returns. In other words, cheap quality stocks are expected to yield higher returns than expensive junk stocks, and this expectation is supported by empirical research. Most studies in the field, however, take B/P as the one “true” indicator for cheapness (e.g., Piotroski (2000) or Piotroski and So (2012)). This is surprising since the true intrinsic value (and thus the cheapness of a stock) is unobservable. Consequentially, there is substantial uncertainty about the empirical validity of proxies for cheapness. We contribute to the field by comparing the ability of different accounting-based intrinsic value estimates to predict future excess stock returns.

Another issue is that existing estimates for intrinsic values are almost exclusively point estimates, again neglecting the large amount of uncertainty involved in valuation. As a potential solution, Bird and Gerlach (2003), Ying et al. (2005) and Higgins and Nandram (2006) apply Bayesian inference to stock price forecasting. Bird and Gerlach (2003) perform Bayesian model averaging using 5-year historical data on fundamentals to predict the probability of value stocks to outperform the market. Their methodology extends on Piotroski (2000) by incorporating uncertainty about the true model for quality, but they rely on book-to-price (B/P) as the sole proxy for cheapness. Higgins and Nandram (2006) rely on quarterly historical time-series data in a finite forecast horizon variant of the Ohlson (1995) model to predict the next quarter's stock price. They show that the predictions for future prices based on the autoregressive Bayesian model are more accurate than a comparable OLS-based approach. They do not, however, include a measure of a firm's inherent quality. In a similar vein, Ying et al. (2005) use a hierarchical Bayesian approach anchoring on the finite forecast Ohlson (1995) model. In line with Higgins and Nandram (2006), they find that the Bayesian model improves the accuracy of return prediction compared to other maximum-likelihood-based estimates. They also, however, do not consider quality. This empirical evidence indicates that Bayesian inference can be a fruitful approach to measure cheapness while accounting for the large amount of uncertainty involved in the process.²

Our goal is, therefore, to provide investors and researchers with a tool that:

- Accounts for a large amount of uncertainty in the measurement of cheapness.
- Includes a suitable measure of quality.
- Is empirically valid.

Our analysis primarily anchors on Frankel and Lee (1998), Gao et al. (2019), Rueenauffer (2022), and Huefner et al. (2021). As alternatives to B/P, Frankel and Lee (1998), Gao et al. (2019), and Huefner et al. (2021) provide several point-estimate measures for cheapness that form benchmarks for the Bayesian workflow proposed by Rueenauffer (2022). Huefner et al. (2021) derive a parsimonious quality index based on valuation theory similar to Piotroski (2000)'s FScore and prove to be empirically valid. This quality index complements the cheapness measures in our analysis. In the analysis, we find that portfolios sorted using the Bayesian cheapness measure (operationalized as the posterior predictive probability of intrinsic value to exceed price) yield higher Sharpe ratios (higher excess returns at lower risk) in annual return windows compared to the other point estimates. They further exhibit higher shares of positive returns. These results apply to portfolios sorted on cheapness alone and conditional sorts combined with quality. For monthly returns, portfolios sorted on B/P and the Gao et al. (2019) hybrid model yield comparable Sharpe ratios and shares of positive returns. It, therefore, seems that accounting for uncertainty mainly pays off for longer holding periods.

Further, we discuss whether the application of value-weighted makes sense in the context of value-based investment strategies (as paradox as it sounds). We find that value-weighted portfolios are (logically) dominated by fewer large stocks and assign lower weights to quality characteristics that value investors look for (profitability, payout, safety, past growth). Furthermore, value-weighted portfolios generally mimic the market, going against the underlying paradigm of Graham and Dodd (1934) that price and value might differ.

The remainder of the paper is structured as follows: First, we summarize and combine existing literature on fundamental quality, cheapness, and equity valuation errors in Section 2. Section 3 then provides an outline of the Bayesian workflow proposed by Rueenauffer (2022). In Section 4, we provide the operationalizations for the different measures for cheapness and VScore. Section 5 critically discusses equal-weighted and value-weighted approaches to portfolio formation. Section 6 then includes the development of the main hypotheses and Section 6 describes the data acquisition and includes descriptive statistics. Section 8 includes the main empirical analysis and limitations. In Section 9, we perform additional robustness tests and sensitivity analyses before we provide concluding remarks in Section 10.

2 | COMBINING RESEARCH ON VALUATION ACCURACY, VALUATION BIAS, AND FUNDAMENTAL QUALITY

2.1 | Predicting stock returns using measures of fundamental quality and cheapness

In capital-market-based accounting and finance research, a large diversity of investment strategies exist that try to predict future stock returns by shifting the return distribution to the right. A common approach anchors on the book-to-price ratio (B/P) due to the historically documented B/P premium on future returns (also labeled as the “value” premium).³ The B/P premium varies over time and has declined since the financial crisis in 2007, as Arnott et al. (2021) show. A major drawback of such a simple (heuristic) approach is that it neglects many other factors that have predictive power for subsequent stock returns. This can lead investors into the so-called value trap, as value stocks may be cheap for a fundamentally justified reason.⁴ That is why other contributions refine the B/P strategy by including additional metrics of fundamental performance or different proxies for intrinsic value—like Frankel and Lee (1998), Piotroski (2000), Bird and Gerlach (2003), Mohanram (2005), Piotroski and So (2012), Penman and Reggiani (2013, 2018), Lee (2015), Penman et al. (2018), Asness et al. (2019), Li and Mohanram (2019), Walkshäusl (2020) and Huefner et al. (2021). These studies provide evidence for the predictive ability of fundamental quality for future stock returns on the US market before and after the financial crisis. They focus on the incongruence of market expectations to fundamental data by creating portfolios based on the relation of quality and cheapness, where portfolios including cheap quality firms (expensive junk firms) exhibit higher (lower) Sharpe ratios and factor model alphas.

2.2 | Approaches to measuring valuation accuracy, valuation bias, and cheapness

A main takeaway from the studies in the field of value investing is that capturing “value” requires more than one dimension (cheapness and quality), that is, B/P is not a sufficient indicator of value. Despite a large history of research on the relation of B/P and returns however, it remains unclear under which conditions a stock should be seen as cheap: cheapness portfolios are constructed using quantile-based approaches, for example, in Piotroski (2000) or Asness et al. (2019), or using absolute boundaries, like in Huefner et al. (2021). As a starting point for the model development, we use pricing error research. Pricing error terms in the existing literature are commonly expressed as signed pricing errors (PE)—reflecting valuation bias—or absolute pricing errors (APE)—reflecting valuation accuracy; see Henschke (2009) and Anesten et al. (2019) and references therein for more details:

$$PE = \frac{V_0 - P_0}{P_0} \text{ or } APE = \left| \frac{V_0 - P_0}{P_0} \right|, \quad (1)$$

where V_0 is the investor’s current estimate of intrinsic value and P_0 is the current observable market price. These error measures can also be transformed to assign different weights to outliers, e.g. to squared or logarithmic errors as shown in Dittmann and Maug (2008). Most valuation accuracy studies focus on the replication of current prices so that ideally $V_0 = P_0$ holds. But even if a model or strategy can replicate current prices perfectly—so that both error measures are zero—that does not necessarily lead to a good predictive ability of the model for subsequent stock returns. When measuring intrinsic value, it is required to assume that $V_0 = P_1$, where P_1 is the future market price at date $t=1$. While the exact pace of convergence (the timespan between $t=0$ and $t=1$) is uncertain, it is reasonable to assume that the market distinguishes between firms with poor and strong fundamental performance well enough to eventually make “prices gravitate towards fundamentals”, as Penman (2011) states. Formally, this relation can be expressed through the total (or cum-dividend) stock return TSR_1 ⁵:

$$TSR_1 = \frac{D_1}{P_0} + \frac{V_0}{P_0} - 1 = DY_1 + R_1. \quad (2)$$

The dividend yield (DY_1) is the same for all models (only V_0 varies across the estimates, not D_1), so the following theoretical comparison focuses on the forward buy-and-hold-return R_1 as implied by the value-to-price ratio (V_0/P_0) minus 1. This way, V_0 becomes an indicator for expected returns and thus a stock’s inherent cheapness. With B/P as the indicator for cheapness, V_0 equals B_0 , which in turn is a RIM valuation with no future expected residual income. Then, the forward buy-and-hold return can be expressed in a first simple modeling variant of the intrinsic value as follows:

$$R_1^B = \frac{V_0^B}{P_0} - 1, \text{ with } V_0^B = B_0. \quad (3)$$

This is a main (often implicit) premise of common value investing strategies where firms with a high B/P ratio are expected to yield higher returns than low B/P firms (as implied by Fama and French (1992)'s high-minus-low (HML) B/P factor). This might be a rather conservative estimate, as Zhang (2000) and Skogsvik and Juettner-Nauroth (2013) theoretically show that accounting conservatism induces positive future residual income. The result of conservative accounting is that value estimates obtained through the RIM are negatively biased (i.e. they have a tendency for negative pricing errors and expected returns, so that $R_1^B < 0$).⁶ Among others, Lev and Gu (2014) argue that market and valuation dynamics have changed significantly over the course of the 21st century, but accounting numbers do not reflect those changes. One aspect of change is the increased importance of intangible investments for value creation. Since a large share of intangible investment is expensed immediately, book values are increasingly understated compared to market prices. This led to a recent development of conservatism corrections that anchor on a retroactive recognition and depreciation of R&D and SG&A for example, as in Penman and Zhang (2002), Peters and Taylor (2017), or Iqbal et al. (2022). As a result, using book value as a proxy for cheapness is a conservative measure. Huefner et al. (2021) extend B/P by incorporating short-term residual income (STE_0) into this equation, so that $V^{1,0}$ equals a RIM with no residual income after the precise forecast horizon, leading to the following slightly upgraded modeling variant of intrinsic value:

$$R_1^{STE} = \frac{V_0^{STE}}{P_0} - 1, \text{ with } V_0^{STE} = B_0 + \frac{(\text{ROE}_1 - r)B_0}{(1+r)}. \quad (4)$$

Frankel and Lee (1998) calculate V/P ratios by inserting analyst forecasts, a perpetual growth rate of zero, and CAPM-based cost of capital into a standard RIM formula. They show that their V/P ratios are positively associated with future stock returns. Based on their analysis, portfolios formed on V/P yield higher returns compared to B/P portfolios over longer-time horizons. In that case, V_0 equals a RIM with a zero-growth steady state assumption after the precise forecast horizon, leading to the third and further extended modeling variant of intrinsic value:

$$R_{FL} = \frac{V_0^{FL}}{P_0} - 1, \text{ with } V_0^{FL} = B_0 + \frac{(\text{ROE}_1 - r)B_0}{(1+r)} + \frac{(\text{ROE}_2 - r)B_1}{(1+r)r}. \quad (5)$$

Even though a zero-growth RIM seems conservative, the long-run earnings growth rate implied by that assumption can be large as it depends on the starting ROE in the numerator of the terminal value. Another stream of research uses market prices and accounting variables to determine the degree of conservatism, like Basu (1997) or Khan and Watts (2009). These approaches involve valuation multiples, such as E/P and B/P. While they are assumed to not be the sole proxy for conservatism—Khan and Watts (2009) also employ Size and Leverage—they represent the difference in earnings recognized in the accounting and future earnings implied by market expectations. That is one reason why the hybrid RIM derived by Gao et al. (2019) reduces the negative valuation bias in RIM-based approaches, as it incorporates the contemporaneous “normal” P/B (ϕ_b), leading to our fourth modeling variant of intrinsic value:

$$R_1^{FL} = \frac{V_0^H}{P_0} - 1, \text{ with } V_0^H = \frac{D_1}{(1+r)} + \frac{\phi_b}{(1+r)} \left(B_1 + \frac{(\text{ROE}_2 - r)B_1 - \left(1 - \frac{1}{\phi_b}\right)D_2}{r-g} \right). \quad (6)$$

Circularity problems concerning market (in)efficiency aside, reducing valuation bias is one of the advantages of multiple-based valuation—it captures market-wide developments in the time series and considers that firms and peers are often priced in similar ways. In turn, we use the hybrid model as the fourth proxy for cheapness.

3 | OUTLINE OF THE BAYESIAN VALUATION WORKFLOW

As proposed by Rueenaufner (2022), we model the time series of stock prices as $P_t = P_{t-1} + \tau_t$, with τ_t being an i.i.d. random variable that represents price innovations and is defined on the interval $(-P_0, \infty)$. If one further assumes that value

leads price ($P_1 = V_0$), a firm's value-to-book ratio (VB) follows the same distribution as τ_t / B , shifted by the firm's price-to-book (PB) ratio to the right and is thus strictly positive. Anchoring on this, we assume that VB follows a Γ (gamma)-distribution with unknown shape and scale parameters α and β , so that $VB \sim \Gamma(\alpha, \beta)$. Then, following Bayes' theorem, the posterior distribution for α and β given the data (VB) is proportional to prior times likelihood:

$$p(\alpha, \beta | VB) \propto p(VB | \alpha, \beta)p(\alpha, \beta). \quad (7)$$

$p(\alpha, \beta)$ is the joint prior distribution for α and β . Assuming independence between α and β , this equals $p(\alpha)p(\beta)$. Since both α and β are strictly positive, prior distributions for them should be defined on the same parameter space (e.g., via χ^2 -priors).

This joint prior is multiplied by $p(VB | \alpha, \beta)$, the likelihood of observing the data given α and β . The likelihood is the backbone of Bayesian inference and has to resemble the data-generating process closely. In turn, it should be defined on the same parameter space as the data. Note that while we ground on Rueenaufner (2022) and use RIM-based value estimates as the input data, it is possible to use any other set of point estimates as long as they fulfill the present value of expected dividends condition (PVED). Ceteris paribus, the more point estimates are generated, the larger the weight of the likelihood in the determination of the posterior. The RIM formula for the value-to-book ratio employed by Rueenaufner (2022) looks as follows:

$$\frac{V_0^r}{B_0} = 1 + \frac{(ROE_1 - r)}{(1 + r)} + \frac{B_1}{B_0} \frac{(ROE_2 - r)}{(1 + r)(r - g)} = \frac{P_0}{B_0} + \frac{\tau_1}{B_0}. \quad (8)$$

The data generation is then done by making different assumptions on the behavior of r , $ROE_{1,2}$, and g . In order to keep V_0^r positive, the point estimates have to fulfill going-concern and transversality. If closed-form inference is not suitable for the decision problem at hand, the posterior $p(\alpha, \beta | VB)$ can only be approximated. The common approach to do so is Markov-Chain-Monte-Carlo simulations, which use iterative algorithms to sample from the posterior distribution directly. Since closed forms limit the prior distributions for α and β to a particular family, we rely on MCMC in this paper.

As the basis for inference, the posterior samples obtained via MCMC can be used to generate samples from the posterior predictive distribution $p(VB^* | VB)$.⁷ The posterior is of limited use on its own, as it only includes uncertainty about the parameter values for α and β and ignores uncertainty about the data generation. The samples from the posterior predictive are then used to approximate the integral on a previously defined interval. This procedure allows investors to make probabilistic statements about the predicted values for VB, conditional on the observed data, likelihood, and the prior. The Bayesian measure for cheapness is V_0^{Bayes} , the posterior predictive probability of VB to exceed the current PB ratio so that $V_0^{\text{Bayes}} = p(VB^* | VB) \geq PB$.

4 | OPERATIONALIZATION OF INTRINSIC VALUE ESTIMATES AND VSCORE

For the operationalization of the point estimates of intrinsic value, we generally follow Frankel and Lee (1998), Gao et al. (2019), and Huefner et al. (2021). All variables are estimated at the portfolio formation date, using only information that is available at the time. B_0 is the most recent book value of equity per share, while $E_{1,2}$ and $D_{1,2}$ are forecasts for earnings per share and dividends per share for the next two fiscal years. Future book values for V_0^{FL} are predicted using clean-surplus accounting. Whenever required, the cost of equity capital r is estimated using the CAPM, so that $\hat{r} = rf + \hat{\beta}_i \text{MRP} + \varepsilon_i$. Here, rf is the 10-year U.S. treasury bond yield at portfolio formation, and the MRP is set to 5% for all firms. $\hat{\beta}_i$ is estimated using 60 monthly returns with the S&P500 as the reference index. In case $\hat{\beta}_i$ is negative, we set it to 1. For the hybrid model V_0^H , ϕ_b is measured as the median P/B ratio of the market at portfolio formation. The future expected payout ratios k_1 are calculated as dividends per share relative to earnings per share for the respective forecast period. Analogously, the forecasts for $ROE_{1,2}$ are calculated as earnings per share relative to the beginning book value of equity per share for the respective forecast period, with book values predicted via clean-surplus accounting. The long-run growth rate in the hybrid model is set to 0.9, as proposed by Gao et al. (2019). For V_0^{Bayes} , we employ the following priors for α and β :

$$\alpha \sim \chi^2(4)$$

$$\beta \sim \chi^2(3)$$

These priors ensure that the mode of the prior predictive distribution for VB is 1 (thus anchoring the prior predictive on book value). They are also weakly informative, as they are defined on the entire positive range with large variances (8 and 6, respectively). For the data input for the likelihood, we follow Rueenaufner (2022) and specify point estimates for VB for each firm at each portfolio formation date, which depend on varying assumptions on $ROE_{1,2}$, r , and g . The first estimate is always equal to 1 ($ROE_1 = ROE_2 = r$, no future residual income). For the second estimate, we assume that $ROE_2 = r$ (no long-run residual income). Then, we include three different scenarios for g : $g = 0.5r$, $g = 0$, and $g = -0.5r$. For each of these variations of payoffs, we use three different scenarios for r ($MRP = 0.03$, $MRP = 0.05$, and $MRP = 0.06$), resulting in 13 total estimates.

For the sampling process, we rely on Hamiltonian MCMC via STAN.⁸ We implement the Γ -model for each observation with a total number of 1200 iterations and 200 burn-in samples in a single chain. We sample from the posterior predictive distribution $p(VB^* | VB)$ and then count the number of samples from the posterior predictive that exceed the current PB-ratio to approximate V_0^{Bayes} . In case any value estimate (including the non-Bayesian estimates) is negative, we set it to one cent.⁹

These specifications are the same for all firms in the sample; portfolio- or industry-based approaches could be the subject of future research. Finally, we operationalize VScore the same way as Huefner et al. (2021), where all eight signals are measured at the portfolio formation date and only involve financial statement information that is available.¹⁰

5 | ABNORMAL RETURN PATTERNS AND PORTFOLIO FORMATION TECHNIQUES

After replicating several of the previously abnormal return patterns, several studies (e.g., Fama (1998) and Kothari and Warner (1997)) showed that abnormal return patterns are often fragile and disappear whenever value-weighted portfolios are applied instead of equal-weighted portfolios.¹¹ In turn, more recent papers (e.g., Asness et al. (2019)) use value-weights since they are (a) more reflective of the market environment and (b) have less liquidity problems caused by smaller stocks. However, as Chiang (2002) argues, that argument can also be turned around: value-weighted portfolios assign larger weights to stocks that are already favored on the market, with momentum becoming a driving factor of portfolio returns (past winners dominate value-weights). As of March 15, 2023, for example, seven stocks are responsible for around 50% of the NASDAQ100's total market capitalization, as shown in Table 1.

Equal-weighted portfolios “give all stocks the same chance” on each portfolio formation date, which is more in line with Graham and Dodd (1934)'s mantra of buying cheap quality stocks, regardless of their size and market capitalization. In consequence, we decided to place emphasis on equal-weighted portfolios in our analyses, but we show additional comparative statistics including value-weighted portfolios as part of our sensitivity analyses. It is important to acknowledge that illiquidity can indeed cause problems if trading strategies cannot be implemented consistently. We, therefore, control for liquidity issues in a separate section later on.

6 | HYPOTHESIS DEVELOPMENT AND RESEARCH METHODOLOGY BASED ON CHEAPNESS AND QUALITY

In our analysis, undervalued stocks are characterized as cheap relative to their quality (high or middle VScore, high cheapness), while overvalued stocks are expensive relative to their quality (low or middle VScore, low cheapness). This incongruence of market expectations with fundamentals leads to predictable return patterns, as Piotroski and So (2012) and Huefner et al. (2021) show. In expectation, portfolios that are classified as undervalued yield higher returns than portfolios that are classified as overvalued—this assumption grounds on Piotroski and So (2012). A different way of looking at this dynamic is to regard both quality and cheapness as predictors of returns. Asness et al. (2019) construct a quality factor by subtracting the returns of low-quality portfolios from the returns of high-quality portfolios (quality minus junk, QMJ). They show that QMJ yields excess returns (via factor model alphas) in various specifications of pooled factor regressions. In a similar vein, the difference in returns between high- and low-cheapness portfolios is an explanation for the documented empirical relevance of the CME (cheap minus

expensive) factor. Combining these two factors then leads to Figure 1, the errors-in-expectations matrix as described by Piotroski and So (2012).

Our main analysis tests for return differences across equal-weighted portfolios sorted on quality (VScore) and cheapness ($V_0/P_0 - 1$), where V_0 is proxied by the five different intrinsic value estimates $V_0^B, V_0^{STE}, V_0^{FL}, V_0^H$, and V_0^{Bayes} . First, we test for return differences in QMJ portfolios to assess the validity of VScore as a measure for QMJ through a comparison of Sharpe Ratios (SRs). For all comparisons of Sharpe Ratios, we employ both 1-month and 1-year ahead returns because Frankel and Lee (1998), Piotroski and So (2012), and Huefner et al. (2021) show that the errors-in-expectations are often corrected over horizons longer than 1 month. The QMJ portfolios are created using VScore. Akin to Huefner et al. (2021), firms with a VScore higher than six are labeled as quality, while firms with a Score lower than two are labeled as junk. We further run pooled calendar-time regressions with monthly QMJ returns using six factors: MKT, SMB, one of the five different CME factors, RMW, CMA, and UMD.¹² If QMJ yields excess returns on its own regardless of the measurement of CME (and the other factors), the intercept (alpha) of each regression will be positive and statistically significant.

Then, we focus on the different CME factors in isolation by looking at differences in Sharpe ratios for CME portfolios without a consideration of QMJ. Portfolios for $V_0^B, V_0^{STE}, V_0^{FL}$, and V_0^H are created using the 20th and 80th percentiles as cutoff points (the lowest 20% are expensive, highest 20% are cheap), while portfolios for $V_0^{Bayes} \leq 0.2$ are created using $V_0^{Bayes} \leq 0.2$ and $V_0^{Bayes} \geq 0.8$, respectively. This leads to balanced CME portfolios for $V_0^B, V_0^{STE}, V_0^{FL}$, and V_0^H , while the portfolios of V_0^{Bayes} do not have to be balanced. To test absolute boundaries for the point estimate measures, we also create portfolios using values of 0.2 and 0.8, respectively. The results are robust to this specification. Then, we run another set of pooled factor regressions for the different CME factors to test whether they yield excess returns and assess their relations to commonly applied risk factors.

TABLE 1 Market capitalization of top seven companies in the NASDAQ100.

Company (NASDAQ 100)	Weight in %
Apple Inc.	12.426
Microsoft Corp.	12.069
Alphabet Inc	7.230
Amazon.com Inc.	6.053
NVIDIA Corp.	4.793
Tesla Inc.	3.582
Meta Platforms Inc.	3.425
Rest	50.422

		CME		
		Low V_0/P_0-1 (Expensive)	Middle V_0/P_0-1	High V_0/P_0-1 (Cheap)
QMJ	Low VScore (Junk)	$E[V_0/P_0] > E[VScore]$ Likely Overvalued	$E[V_0/P_0] > E[VScore]$ Potentially Overvalued	$E[V_0/P_0] = E[VScore]$ Aligned
	Middle VScore	$E[V_0/P_0] > E[VScore]$ Potentially Overvalued	$E[V_0/P_0] = E[VScore]$ Aligned	$E[V_0/P_0] < E[VScore]$ Potentially Undervalued
	High VScore (Quality)	$E[V_0/P_0] < E[VScore]$ Aligned	$E[V_0/P_0] < E[VScore]$ Potentially Undervalued	$E[V_0/P_0] < E[VScore]$ Likely Undervalued

FIGURE 1 Errors-in-expectations matrix with CME and QMJ factors. This figure shows the classification of firms into overvalued, aligned and undervalued portfolios based on the firm's fundamental quality implied by VScore and cheapness implied by the expected stock return. Here, V_0 is the estimate of the firm's intrinsic value at portfolio formation and P_0 is the current market price. VScore is the sum of eight signals of fundamental quality as shown in Huefner et al. (2021).

Afterward, we examine returns for portfolios sorted on quality and cheapness simultaneously. Out of the five estimates, V_0^B has the lowest data requirements, but comes at the cost of not including any explication of future expectations. V_0^B and V_0^{STE} are very similar since the incremental change from added STE is small. V_0^{FL} is more volatile because it depends on viable point estimates for ROE_1 and ROE_2 to yield feasible results. Therefore, we expect a higher share of misclassified firms for V_0^{FL} due to one-cent estimates whenever ROE_1 or ROE_2 are negative. Excluding negative estimates for V_0^{FL} reduces the sample size, leading to lower practical utility. How the misclassification of firms for V_0^{FL} changes the return patterns is hard to predict ex-ante. This misclassification distorts the results if quantile-based portfolios are used and may weaken or nullify the expected abnormal returns earned by CME. Overall, we expect the hybrid model V_0^H to increase intrinsic value estimates compared to the others since historical P/B ratios are larger than 1 (it has a median of ~ 2 in our sample). This lowers valuation bias, but comes at the cost of lower return predictability because including the contemporaneous market sentiment may create a larger deviation from firm fundamentals (a dilution of quality). Finally, V_0^{Bayes} mimics the behavior of V_0^B , but does not require market-wide data to function due to its probabilistic nature (it stays between 0 and 1). Furthermore, it excludes firms with a large uncertainty concerning the intrinsic value estimate compared to V_0^B . We, therefore, expect it to provide additional assurance for quality when paired with VScore. Empirically, we expect V_0^{Bayes} to increase Sharpe ratios compared to V_0^B by avoiding the value trap more reliably. Finally, we examine the time-series behavior of the share of quality firms that are classified as cheap by V_0^{Bayes} over time to assess the consistency of the strategy. Expectedly, market sentiment plays an important role here since the share of cheap firms should be larger in times when overall market expectations are low and vice versa. Similarly, the share of quality firms should be smaller in times of economic downturn.

7 | DATA ACQUISITION AND DESCRIPTIVE STATISTICS

7.1 | Data acquisition

Our sample includes all firms listed on the US Stock Market (Nasdaq and NYSE) and historical prices of the S&P500 price index from June 2002 to July 2022 (monthly and annual dataset range from June 2002 until July 2021–230 Months), taken from Refinitiv Eikon. Historical data is obtained from Refinitiv DataStream and analyst forecasts stem from I/B/E/S. Portfolios are formed at the beginning of each month (calendar time approach). Factor model data is obtained from the Kenneth French Database. QMJ and the five different CME factors are calculated for each portfolio formation date. Firms need to have data for VScore and CME factors available. We set dividends and investment income to zero when they are unavailable to increase coverage. Firms with a per-share value lower than 3\$ are excluded because they are subject to microstructure effects and have a highly right-skewed distribution, as Ball et al. (1995). Our results are robust to alternative low-dollar thresholds (the higher the threshold, the more extreme returns are excluded). This mainly affects junk stocks, because the majority of low-dollar stock is concentrated there. Total stock returns need to be available for both 1 month ahead and 1 year ahead of portfolio formation. Unless they are available on DataStream, delist returns are set to -30% following Shumway (1997). After applying all filters, 425,967 monthly observations remain. Due to the large sample size, test statistics (even for subsamples) may exhibit natively high significance levels, which is why we focus on both statistical and economic significance, at the cost of (potentially) more subjective inferences.¹³

7.2 | Descriptive statistics and basic correlations

As the baseline of the empirical analysis, Table 2 includes descriptive statistics for the input variables. Panel A shows return statistics, Panel B shows statistics for the eight VScore variables and Panel C shows statistics for the expected returns implied by the five different intrinsic value estimates. It is visible in Panel A that the average annual (monthly) total stock return in the sample is 9.03% (0.25%), indicating that the sample covers a bull market period. That is further reinforced by the median and share of positive returns since the median is positive and the share of positive returns is greater than 50%. Panel B includes a few large outliers for many input variables for VScore, making the means and standard deviations unreliable. The percentile statistics and shares of binaries set to 1 are more robust to outliers, which is why we concentrate on them. Payout ratios exhibit (as expected) a large share of zeroes, since they are set to zero if they are unavailable. They are also among the biggest gatekeepers for quality, with only 51.52% and 45.19% cases of positive binaries. Aside from

that, the VScore statistics indicate that many EROEs are only positive because firms use the leverage effect, not because of actual excess operating profitability (high positive binary $EROE_0$ and FLG_0 , but lower $EORBA_0$ and $IRBA_0$). Finally, Panel C reveals that both returns implied by V_0^B and V_0^{STE} have a negative bias (consistent with prior research), since only a small fraction of these returns is positive and both medians are negative. Akin to Panel B, the means and standard deviations for the expected returns are unreliable (aside from V_0^B and V_0^{Bayes}). Part of the negative bias in V_0^B and V_0^{STE} is eliminated in V_0^{FL} and V_0^H since they increase both the share of positives and the medians. This, however, comes at the cost of very high standard deviations caused by large positive outliers, reducing the overall accuracy. The last intrinsic value estimate V_0^{Bayes} also includes a negative bias, since the average posterior probability for a positive return is only 32.30% and the median is only 26.80%.

As a means to test for statistical relationships between the input variables, Table 3 shows basic correlations. Note that significance levels for either panel might be affected by the large sample size, which causes them to be statistically significant, even for values that appear economically negligible. Panel A includes Pearson correlations between the eight binaries for VScore and Panel B includes Spearman correlations between the actual returns and the expected returns implied by the different intrinsic value estimates.

Large correlations in Panel A (0.634 and 0.743) exist between historical and expected EROEs and payout ratios, in line with prior research (e.g. Huefner et al. (2021)). It thus seems that analyst forecasts tend to project prior profitability and payout behavior into the future. Another relatively strong positive correlation exists between the two EROEs and

TABLE 2 Descriptive statistics for actual returns, VScore, and expected returns.

Panel A: Returns—descriptive statistics								
Returns	Mean (μ)	1st Perc.	20th Perc.	Median	80th Perc.	99th Perc.	SD (σ)	% > 0
TSR ($t=1/12$)	0.25%	-30.39%	-6.34%	0.59%	7.30%	24.76%	10.05%	58.65%
TSR ($t=1$)	9.03%	-74.21%	-19.62%	6.37%	33.81%	136.50%	39.50%	53.59%
Panel B: VScore—descriptive statistics								
VScore	Mean (μ)	1st Perc.	20th Perc.	Median	80th Perc.	99th Perc.	SD (σ)	Binary > 0
$EROE_0$	4.76%	-62.97%	-2.86%	3.66%	12.94%	151.29%	8.41	68.04%
$EORBA_0$	2.61%	-28.56%	-5.00%	0.20%	6.58%	38.45%	1.51	51.26%
$IRBA_0$	0.19%	0.00%	0.00%	0.00%	0.24%	2.85%	0.06	51.49%
FLG_0	19.36%	-5.21%	0.74%	5.87%	14.54%	182.40%	3.10	86.03%
k_0	52.66%	0.00%	0.00%	18.26%	55.94%	525.00%	3.36	51.52%
$EROE_1$	112.97%	-16.41%	-1.55%	4.28%	12.78%	139.89%	100.27	73.48%
k_1	29.26%	0.00%	0.00%	5.54%	45.16%	288.24%	2.19	45.19%
STRIG	20.11%	-200.00%	-29.11%	9.12%	86.71%	200.00%	0.96	60.20%
Panel C: Intrinsic value-based expected returns—descriptive statistics								
Expected return	Mean (μ)	1st Perc.	20th Perc.	Median	80th Perc.	99th Perc.	SD (σ)	% > 0
R_1^B	-44.26%	-96.07%	-74.09%	-51.92%	-20.63%	78.57%	58.87%	9.34%
R_1^{STE}	80.48%	-92.81%	-71.55%	-49.92%	-19.83%	73.72%	12,516.90%	9.69%
R_1^{FL}	2284.66%	-96.08%	-52.70%	-21.76%	23.04%	224.72%	232,145%	33.73%
R_1^H	2207.77%	-79.81%	-23.07%	19.58%	75.71%	256.90%	216,424%	64.36%
V_0^{Bayes}	32.30%	0.00%	5.80%	26.80%	58.20%	92.20%	26.40%	/

Note: This table includes descriptive statistics for the input data of the empirical analysis of 425,967 monthly firm observations, starting in June 2022 and ending in July 2021 (230 months). Panel A includes descriptive statistics for monthly ($t=1/12$) and annual ($t=1$) returns after portfolio formation at the beginning of each month (calendar time returns). TSR is the total stock return, including dividend payments if they fall into the return window. Panel B includes descriptive statistics for the eight VScore inputs. The detailed descriptions of the input variables for VScore are shown in Huefner et al. (2021). $EROE_0$ is the excess ROE over the cost of equity capital for the last fiscal year, $EORBA_0$ is the excess operating return over the cost of equity capital for the last fiscal year, $IRBA_0$ is the investment return for the last fiscal year, FLG_0 is the financial leverage gain over the cost of debt for the last fiscal year, k_0 is the payout ratio for the last fiscal year, $EROE_1$ is the excess ROE over the cost of equity capital for the next fiscal year, k_1 is the payout ratio for the next fiscal year and STRIG is the growth rate in residual income from the last to the next fiscal year. R_1^B , R_1^{STE} , R_1^{FL} , R_1^H are the expected returns implied by the intrinsic value estimates described in Section 2 and V_0^{Bayes} is the posterior predictive probability of intrinsic value to exceed price.

operating profitability, with values of 0.450 and 0.489 for $EROE_0$ and $EROE_1$, respectively. Aside from those correlations, most Pearson correlations are moderately positive, moderately negative, or close to zero. This indicates that overall, the eight variables add incremental value to VScore. In Panel B, the Spearman correlations between actual and expected returns are rather weak, while several different expected return measures exhibit strong positive correlations with each other. Especially for V_0^B and V_0^{STE} , the correlation is almost perfectly positive (0.990). This indicates a low incremental benefit of adding short-term expectations to the anchoring book value. There is also a strong positive correlation between V_0^{Bayes} and the expected returns implied by the other estimates. Since V_0^{Bayes} includes the other estimates aside from V_0^H as input, that is not surprising.

8 | EMPIRICAL RESULTS

8.1 | Comparison of returns for Quality-Minus-Junk portfolios (QMJ)

The first endeavor of the empirical analysis is to test whether quality can predict returns on its own, as shown by Piotroski (2000) and Asness et al. (2019). We calculated return statistics (excess returns, standard deviations, and Sharpe Ratios) for the nine different portfolios formed on the different values for VScore. Panel A of Table 4 shows these statistics for both monthly and annual returns after the portfolio formation date, while Panel B shows t -statistics and F -statistics for QMJ portfolios. It is visible that there is a larger share of observations that are classified as quality compared to junk. As is common in these event studies over longer horizons, this could be the result of a

TABLE 3 Basic correlations between VScore binaries and returns (actual and expected).

Panel A: Pearson correlations for VScore binaries								
ρ (Pearson)	$EROE_0$	$EORBA_0$	$IRBA_0$	FLG_0	k_0	$EROE_1$	k_1	STRIG
$EROE_0$								
$EORBA_0$	0.450***							
$IRBA_0$	-0.021***	0.124***						
FLG_0	0.061***	0.006***	-0.121***					
k_0	0.283***	0.153***	-0.109***	0.178***				
$EROE_1$	0.634***	0.489***	-0.013***	0.090***	0.227***			
k_1	0.216***	0.150***	-0.092***	0.196***	0.743***	0.234***		
STRIG	-0.161***	0.057***	0.046***	0.044***	-0.072***	0.143***	0.016***	
Panel B: Spearman correlations for returns (actual and expected)								
ρ (Spearman)	TSR ($t=1/12$)	TSR ($t=1$)	R_1^B	R_1^{STE}	R_1^{FL}	R_1^H	V_0^{Bayes}	
TSR ($t=1/12$)								
TSR ($t=1$)	0.265***							
R_1^B	0.028***	0.077***						
R_1^{STE}	0.030***	0.082***	0.990***					
R_1^{FL}	0.029***	0.094***	0.331***	0.417***				
R_1^H	0.015***	0.033***	0.725***	0.785***	0.762***			
V_0^{Bayes}	0.034***	0.110***	0.497***	0.576***	0.965***	0.831***		

Note: This table includes correlations for the input data of the empirical analysis of 425,967 monthly firm observations, starting in June 2022 and ending in July 2021 (230 months). Panel A includes Pearson correlations for the for the eight VScore binaries. The detailed descriptions of the input variables for VScore are shown in Huefner et al. (2021). $EROE_0$ is the excess ROE over the cost of equity capital for the last fiscal year, $EORBA_0$ is the excess operating return over the cost of equity capital for the last fiscal year, $IRBA_0$ is the investment return for the last fiscal year, FLG_0 is the financial leverage gain over the cost of debt for the last fiscal year, k_0 is the payout ratio for the last fiscal year, $EROE_1$ is the excess ROE over the cost of equity capital for the next fiscal year, k_1 is the payout ratio for the next fiscal year and STRIG is the growth rate in residual income from the last to the next fiscal year. Panel B includes Spearman rank correlations for monthly ($t=1/12$) and annual ($t=1$) returns after portfolio formation at the beginning of each month (calendar time returns). TSR is the total stock return, including dividend payments if they fall into the return window. R_1^B , R_1^{STE} , R_1^{FL} , R_1^H , and V_0^{Bayes} are the expected returns implied by the intrinsic value estimates described in Sections 2 and 3 and P_0 is the current market price. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively.

survivorship bias on the market (successful quality firms survive longer and thus form a larger share of the sample). Looking at the excess returns and standard deviations for the nine portfolios in Panel A, it is clearly visible that returns increase with VScore, while the standard deviation behaves the opposite way. It is, therefore, not surprising that the test-statistics in Panel B are both economically and statistically significant, indicating that QMJ predicts Sharpe Ratios on its own.¹⁴

In order to further validate VScore as a quality measure, we also calculate pooled calendar time regressions with the five different cheap-minus-expensive (CME) measures described in Section 2. V_0^B is analogous to the classic HML factor outlined by Fama and French (1992). Table 5 reports the regression coefficients, t -statistics, and adjusted R^2 s for pooled regressions of QMJ on the other factors. Here, all alphas are positive, range between 1.10% and 1.23% and remain statistically significant. Aside from V_0^H , the hybrid model, and V_0^{FL} , the terminal value RIM, the CME factors yield a negative sign and have small significance levels on the respective coefficients. For the hybrid model, the factor loading on CME is close to zero and statistically insignificant, while the coefficient on V_0^{FL} is positive and statistically significant. The latter could be an outcome of the misclassification on CME due to the inconsistencies in the terminal value. The coefficient on RMW is strongly positive and significant; since profitability is a factor of quality as well (gross profit relative to total assets should correlate positively with ROE), the positive sign is in line with the expectation (we perform additional analyses involving RMW later). For SMB and MKT, the signs are consistently negative for all CME factors. This implies that quality stocks are less exposed to those risk factors. This reinforces the impression that quality stocks seem safer than junk stocks overall, consistent with Asness et al. (2019). Finally, CMA and UMD are inconsistent in terms of sign and/or significance, indicating that they have rather weak explanatory power for QMJ returns. The adjusted R^2 always ranges between 0.52 and 0.55, which implies that a relatively large share of the variance in QMJ returns is explained by the six respective factors. In summary, it seems that QMJ based on VScore predicts Sharpe ratios for both monthly and annual return windows following portfolio formation. The main inference remains the same after controlling for commonly applied risk factors, including several different operationalizations of CME.

TABLE 4 Return statistics for portfolios formed on quality and junk (QMJ).

Panel A: Quality returns			Monthly TSR- r_f ($t = 1/12$)			Annual TSR- r_f ($t = 1$)		
Classification	VScore	n	μ	σ	SR	μ	σ	SR
Junk	0	1732	-1.66%	0.14	-11.79%	8.11%	51.22%	15.84%
	1	14,975	-0.38%	0.12	-3.19%	8.78%	50.40%	17.41%
Middle	2	37,639	-0.12%	0.12	-1.08%	9.17%	47.20%	19.44%
	3	53,789	-0.02%	0.11	-0.18%	7.77%	44.54%	17.44%
	4	67,133	0.08%	0.11	0.73%	8.45%	42.39%	19.93%
	5	82,045	0.22%	0.10	2.15%	8.10%	39.76%	20.38%
	6	77,342	0.46%	0.09	5.07%	10.11%	35.43%	28.54%
Quality	7	57,831	0.62%	0.08	7.46%	10.10%	31.17%	32.39%
	8	33,472	0.69%	0.08	8.84%	10.18%	28.70%	35.48%

Classification	Monthly TSR- r_f ($t = 1/12$)			Annual TSR- r_f ($t = 1$)		
	μ	σ	SR	μ	σ	SR
Junk	-0.40%	12.11%	14.83%	8.71%	50.49%	-3.29%
Quality	0.65%	8.18%	10.13%	10.13%	30.29%	7.98%
QMJ	1.05%	-3.93%	22.64%	1.42%	-20.20%	11.27%
Stat	11.81***	0.44***		4.26***	0.35***	

Note: This table shows average excess total stock returns (μ), standard deviations (σ), and Sharpe Ratios (SR) for returns of equal-weighted portfolios classified based on their quality for a total amount of 425,967 monthly observations of US-Firms from June 2002 to July 2021. For Panel A, firms are classified based on VScore each month and for Panel B, firms are classified as Quality if they achieve a VScore greater than 5 and as Junk when their VScore is lower than 2. TSR- r_f is the excess total stock return for the subsequent month, including dividend payments if they occur in the respective window, less the risk-free rate. SR is calculated as excess returns divided by the respective standard deviation. For QMJ, t -statistics and F -statistics are shown below the excess return and standard deviation, with ***, **, and * indicating significance levels of 0.01, 0.05 and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

TABLE 5 Asset pricing tests for QMJ factors.

QMJ	V_0^B	V_0^{STE}	V_0^{FL}	V_0^H	V_0^{Bayes}
Alpha (in %)	1.23	1.21	1.10	1.15	1.15
	9.07***	8.88***	8.02***	8.19***	8.21***
MKT	-0.10	-0.10	-0.07	-0.10	-0.10
	-2.94***	-2.88***	-2.1**	-2.86***	-2.82***
SMB	-0.38	-0.40	-0.47	-0.47	-0.47
	-6.27***	-6.61***	-8.16***	-7.8***	-8.12***
RMW	0.55	0.56	0.44	0.52	0.52
	7.53***	7.48***	5.51***	6.73***	6.87***
CMA	0.09	0.07	-0.08	-0.03	-0.03
	1.02	0.73	-0.89	-0.37	-0.36
UMD	-0.05	-0.05	-0.04	-0.05	-0.05
	-1.89*	-1.92*	-1.28	-1.68*	-1.68*
CME	-0.22	-0.18	0.19	0.00	0.00
	-3.93***	-3.24***	3.02***	0.08	-0.01
Adj. R^2	0.55	0.54	0.54	0.52	0.52

Note: This table shows six-factor alphas for different operationalizations of equal-weighted CME portfolios. The regression is a pooled time-series regression of QMJ on six commonly applied risk factors. Alpha is the intercept and can be interpreted as the average excess return (in basis points) of the QMJ portfolios. MKT, SMB, RMW, and UMD are taken from Kenneth French's website. The descriptions of the five factors on the website are as follows: MKT is the excess return of the market portfolio, including all NYSE, AMEX, and NASDAQ firms. SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios and UMD is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. The CME factors are constructed using V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes} . The t statistics are shown below the coefficients, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

8.2 | Comparison of returns for CME portfolios

Based on the errors-in-expectations matrix, CME complements QMJ as the second explanatory factor for return patterns, as Piotroski and So (2012) and Huefner et al. (2021) show. Our next focal point in the analysis is, therefore, the comparison of different operationalizations of CME in isolation, that is, which CME factor appears as the best complement for QMJ. In order to test the respective validity, we first calculate return statistics for each CME factor. Table 6 shows those statistics, with Panel A including monthly returns and Panel B including annual returns following portfolio formation. Overall, monthly and annual returns for all CME factors aside from V_0^{FL} yield consistent patterns. Cheap stocks yield the highest returns, but also inherit higher risk; this risk is only consistent for annual returns (for monthly returns, it is largely statistically and/or economically insignificant). Cheap stocks include a larger share of positive returns, particularly over the annual horizon. Hence, positive CME returns manifest over short and long horizons, but cheap portfolios have heavier tails with larger gains and losses on individual stocks over the subsequent year. One could see the higher returns as a compensation for risk (over longer horizons), but as Huefner et al. (2021) show, there are fundamental differences across the three cheap portfolios that can be explained by QMJ. Interestingly, V_0^{Bayes} yields by far the largest Sharpe Ratios and the largest share of positive returns for annual returns, while the difference is not economically significant for monthly returns (it does still have the highest return difference, with 0.51%). It, therefore, seems that the utility of the Bayesian estimate manifests over horizons longer than 1 month after portfolio formation.

The large annual returns for cheap firms identified by V_0^{Bayes} exhibit a jump in standard deviation, indicating that the higher returns are not free lunch. It is, therefore, of interest whether combining V_0^{Bayes} with a quality measure might counter that affect, if at least partially. The unbalanced portfolios for V_0^{Bayes} also lead to a much smaller number of cheap firms. Due to the probabilistic nature of the Bayesian CME factor, a larger number of stocks is excluded, so that the cheap portfolios for V_0^{Bayes} are smaller than the portfolios formed on the other CME factors (25,330 vs. 85,292 observations).

To control for risk factors, we again run pooled calendar-time regressions, this time with one of the five CME factors as the variable to explain each regression. Consistent with Table 6, all factors in Table 7 aside from V_0^{FL} yield positive and

significant alphas, with lower alphas for V_0^H and V_0^{Bayes} compared to V_0^B and V_0^{STE} . The adjusted R^2 is lower for V_0^H and V_0^{Bayes} , indicating that the other risk factors explain less of the variance of these two CME factors. The loading on RMW is once again consistently positive and significant; high expected returns implied by CME go along with high expected returns implied by RMW. Additionally, it seems that momentum (UMD) does not affect CME factors. QMJ is highly significant for V_0^B , V_0^{STE} , and V_0^{FL} , but has a lower significance level for V_0^{Bayes} . Negative signs on QMJ indicate that cheaper stocks have lower quality overall for V_0^B , V_0^{STE} , and V_0^{Bayes} . This is consistent with the finding of Asness et al. (2019) that quality usually comes at a price. The main difference between V_0^{Bayes} and V_0^B lies in SMB: for V_0^{Bayes} there is no significant effect of size on the returns, while the coefficient for V_0^B is significantly positive. In turn, it seems that V_0^B generally identifies larger firms, while there is no size difference between stocks that are identified as cheap and expensive for V_0^{Bayes} . Overall, the lower R^2 shows a weaker exposure of V_0^H and V_0^{Bayes} to the different risk factors.

TABLE 6 Return statistics for portfolios formed different CME factors.

Factor	Class	n	Panel A: Monthly TSR-rf ($t = 1/12$)				Panel B: Annual TSR-rf ($t = 1$)			
			μ	σ	Pos (%)	SR (%)	μ	σ	Pos (%)	SR (%)
V_0^B	Expensive	85,292	0.10%	0.10	53.39	0.95	8.34%	0.39	58.33	21.23
	Middle	255,383	0.22%	0.10	53.37	2.23	8.15%	0.38	58.07	21.44
	Cheap	85,292	0.50%	0.10	54.48	4.86	12.34%	0.44	60.79	28.24
	CME		0.41%	0.00	1.10	3.91	4.00%	0.04	2.46	7.01
	Stat		8.21***	1.04			19.90***	1.24		
V_0^{STE}	Expensive	85,292	0.08%	0.10	53.28	0.82	8.35%	0.40	58.07	20.93
	Middle	255,383	0.22%	0.10	53.38	2.27	8.17%	0.38	58.11	21.55
	Cheap	85,292	0.51%	0.10	54.58	4.92	12.30%	0.43	60.96	28.31
	CME		0.42%	0.00	1.30	4.09	3.95%	0.04	2.89	7.38
	Stat		8.49***	1.01			19.58***	1.19		
V_0^{FL}	Expensive	85,292	-0.02%	0.11	51.95	-0.21	8.49%	0.45	54.94	19.08
	Middle	255,383	0.35%	0.10	53.98	3.58	9.30%	0.38	59.49	24.68
	Cheap	85,292	0.24%	0.10	54.10	2.48	8.76%	0.39	59.95	22.21
	CME		0.26%	-0.02	2.16	2.68	0.28%	-0.05	5.01	3.13
	Stat		5.18***	0.75***			1.351*	0.786***		
V_0^H	Expensive	85,409	-0.02%	0.10	52.49	-0.23	8.10%	0.41	57.11	19.76
	Middle	255,769	0.29%	0.10	53.70	2.93	8.58%	0.38	58.75	22.86
	Cheap	85,409	0.43%	0.10	54.41	4.15	11.32%	0.43	59.99	26.05
	CME		0.45%	0.00	1.92	4.38	3.22%	0.02	2.88	6.29
	Stat		8.96***	0.96***			15.718***	1.12		
V_0^{Bayes}	Expensive	177,691	0.02%	0.10	51.85	0.15	5.35%	0.39	53.63	13.61
	Middle	222,946	0.41%	0.10	54.68	4.17	10.65%	0.38	61.58	27.97
	Cheap	25,330	0.53%	0.12	56.36	4.50	20.60%	0.49	68.37	42.22
	CME		0.51%	0.02	4.51	4.34	15.25%	0.09	14.74	28.61
	Stat		6.62***	1.38			47.58***	1.54		

Note: This table shows average excess total stock returns (μ), standard deviations (σ), the share of positive returns (Pos), and Sharpe Ratios (SR) for equal-weighted portfolios that are classified as low cheapness, middle cheapness, or high cheapness for a total amount of 425,967 observations of US-Firms from June 2002 to July 2021 (230 months). Each month, firms are classified based on cheapness via different proxies for intrinsic value (V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes}). The cheapness portfolios for overvalued (undervalued) consist of the 20% lowest (highest) observations for intrinsic value estimates at portfolio formation. For V_0^{Bayes} , portfolios are constructed based on the posterior predictive probability. Panel A shows characteristics of monthly returns following the portfolio formation date, while Panel B Shows the same statistics for annual returns. TSR is the total stock return for the subsequent month, including dividend payments if they fall into the respective window. Pos is the share of positive TSRs in the portfolio. Sharpe Ratios are calculated as excess returns divided by the respective standard deviation. For the CME returns, t -statistics and F -statistics are shown below the excess return and standard deviation, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. t -test are one-sided tests for High > Low and F -Tests are one-sided for High < Low. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

TABLE 7 Asset pricing tests for CME factors.

CME factor	V_0^B	V_0^{STE}	V_0^{FL}	V_0^H	V_0^{Bayes}
Alpha (in %)	0.68	0.65	0.04	0.38	0.38
	3.87***	3.5***	0.27	2.05**	2.05**
MKT	-0.02	-0.01	-0.12	-0.03	-0.11
	-0.64	-0.37	-3.17***	-0.63	-2.71***
SMB	0.27	0.28	0.06	0.31	0.03
	3.7***	3.57***	0.92	3.97***	0.34
RMW	0.29	0.33	0.35	0.31	0.17
	3.08***	3.37***	4.07***	3.18***	1.8*
CMA	0.56	0.55	0.26	0.36	0.24
	5.65***	5.34***	2.76***	3.41***	2.32**
UMD	-0.03	-0.04	-0.06	-0.04	-0.07
	-0.98	-1.27	-1.83*	-1.27	-1.95*
QMJ	-0.29	-0.25	0.21	0.01	0.00
	-3.93***	-3.24***	3.02***	0.08	-0.01
Adj. R^2	0.32	0.28	0.30	0.15	0.10

Note: This table shows six-factor alphas for different operationalizations of equal-weighted high-minus-low cheapness portfolios (CME). The regression is a time-series regression of CME on six commonly applied risk factors. Alpha is the intercept and can be interpreted as the average excess return of the respective CME portfolios. MKT, SMB, RMW, and UMD are taken from Kenneth French's website. The descriptions of the five factors on the website are as follows: MKT is the excess return of the market portfolio, including all NYSE, AMEX, and NASDAQ firms. SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios and UMD is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. The CME factors are constructed using V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes} . The t statistics are shown below the coefficients, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

8.3 | Comparison of returns for portfolios formed on QMJ and CME

Since both QMJ and CME-based portfolios yield positive alphas, the next step in the analysis is the combination of both strategies to exploit the errors-in-expectations hypothesis further. In order to do so, we form portfolios on quality and cheapness simultaneously, in the form of undervalued (UV) and overvalued (OV) portfolios. The result is another factor, Undervalued-minus-Overvalued (UMO). Table 8 includes return statistics for portfolios formed on UMO for the five different intrinsic value estimates, with Panel A of Table 8 showing monthly statistics and Panel B showing annual statistics. For both monthly and annual return windows, the results are consistent with the errors-in-expectations hypothesis, as UMO is positive for all five cheapness measures. Except for V_0^{FL} (with a UMO of 0.81%), all return differences on UMO are economically significant. Interestingly, V_0^{Bayes} yields the highest UMO return and the largest difference in the share of positive returns for both monthly and annual returns. As expected, the standard deviation decreases consistently from overvalued to undervalued for all five CME factors. It is also important to acknowledge that the SR for UV portfolios formed on V_0^{Bayes} in Table 8 is lower than the SR for cheap portfolios formed on V_0^{Bayes} in Table 6, but only for annual returns. V_0^{Bayes} as a standalone CME factor yields high returns, but picks up on risk for both monthly and annual returns. Adding the quality dimension leads to a reduction of both risk and return for the longer window because cheap junk firms are removed, while more quality firms are added. For monthly returns, combining cheapness and quality increases returns and reduces risk for all five measures, with V_0^B showing the largest difference returns and Sharpe ratios (but not by far compared to V_0^{STE} and V_0^{Bayes}).

It, therefore, seems that the Bayesian estimate starts outperforming the others only for holding periods longer than 1 month. Table 8 includes the UMO factors, but does not grant insight into the full errors-in-expectations pattern as shown in Figure 1. To provide full insight, Table 9 includes return statistics for the nine V_0^{Bayes} -portfolios (Panel A again with monthly returns, Panel B with annual returns). For the monthly returns in Panel A, the contrarian portfolios (top left and bottom right) follow the pattern implied by errors-in-expectations. Cheap quality stocks have the highest returns at low (but not the lowest) risk and the highest share of positive returns. In Panel A, the cheap junk stocks stand out,

TABLE 8 Return statistics for portfolios formed on QMJ and CME factors.

Factor	Class	n	Panel A: Monthly returns ($t = 1/12$)				Panel B: Annual returns ($t = 1$)			
			μ	σ	Pos	SR	μ	σ	Pos	SR
V_0^B	OV	64,924	-0.19%	0.11	51.76	-1.69	7.76%	0.44	54.85	17.22
	A	224,637	0.17%	0.10	53.09	1.69	7.97%	0.39	57.25	20.51
	UV	136,406	0.60%	0.09	55.31	6.37	11.39%	0.38	62.72	29.83
	UMO		0.79%	-0.02	3.55	8.06	3.63%	-0.06	7.87	12.60
	Stat		15.52***	0.71***			17.90***	0.74***		
V_0^{STE}	OV	67,306	-0.17%	0.11	51.74	-1.55	7.80%	0.45	54.80	17.44
	A	220,699	0.17%	0.10	53.11	1.72	8.03%	0.39	57.33	20.76
	UV	137,962	0.59%	0.09	55.29	6.27	11.24%	0.38	62.71	29.58
	UMO		0.76%	-0.02	3.54	7.83	3.43%	-0.07	7.91	12.14
	Stat		15.24***	0.70***			17.13***	0.72***		
V_0^{FL}	OV	85,490	-0.07%	0.11	51.74	-0.62	8.44%	0.45	54.52	18.57
	A	194,834	0.27%	0.10	53.38	2.63	9.13%	0.39	58.14	23.15
	UV	145,643	0.42%	0.09	54.99	4.68	9.25%	0.36	61.82	25.95
	UMO		0.49%	-0.02	3.25	5.31	0.81%	-0.10	7.30	7.38
	Stat		10.87***	0.63***			4.46***	0.61***		
V_0^H	OV	77,068	-0.16%	0.11	51.59	-1.47	7.85%	0.44	55.14	17.84
	A	203,506	0.21%	0.10	53.22	2.06	8.40%	0.39	57.61	21.61
	UV	145,393	0.53%	0.09	55.19	5.72	10.55%	0.38	62.02	27.91
	UMO		0.70%	-0.02	3.60	7.20	2.70%	-0.06	6.88	10.07
	Stat		14.91***	0.71***			14.47***	0.74***		
V_0^{Bayes}	OV	155,016	-0.06%	0.10	51.46	-0.60	5.46%	0.41	52.79	13.19
	A	185,843	0.33%	0.10	53.98	3.32	9.82%	0.39	59.70	25.28
	UV	85,108	0.65%	0.09	56.67	6.99	13.82%	0.37	67.13	37.64
	UMO		0.72%	-0.01	5.21	7.59	8.36%	-0.05	14.34	24.46
	Stat		17.22***	0.80***			50.99***	0.79***		

Note: This table shows average excess total stock returns (μ), standard deviations of returns (σ), the share of positive returns (Pos) and Sharpe Ratios (SR) for equal-weighted portfolios that are classified as overvalued (OV), aligned (A) or undervalued (UV) for a total amount of 425,967 observations of US-Firms from June 2002 to July 2021 (230 months). Each month, firms are classified based on cheapness via different proxies for intrinsic value (V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes}) and then further classified based on quality through VScore. The cheapness portfolios for overvalued (undervalued) firms for V_0^B , V_0^{STE} , V_0^{FL} , and V_0^H consist of the 20% lowest (highest) observations for intrinsic value estimates at portfolio formation. For V_0^{Bayes} , portfolios are constructed based on the posterior predictive probability, so that overvalued (undervalued) portfolios include all firms with a posterior predictive probability of intrinsic value to exceed price lower than 20% (larger than 80%). Panel A shows characteristics of monthly returns following the portfolio formation date, while Panel B shows the same statistics for annual returns. TSR is the total stock return for the subsequent timeframe, including dividend payments if they fall into the respective window. Rf is the 1 month and 10-year US-treasury bond yield at the time. Pos is the share of positive excess returns (TSR-rf) in the portfolio. SR is calculated as excess return divided by the respective standard deviation. For the CME returns, t -statistics and F -statistics for Stat $\neq 0$ are shown below the excess return and standard deviation, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

because they have negative returns (-0.53%) and inherit the highest standard deviation (0.14). These negative returns reverse for the annual return window in Panel B. There, cheap junk stocks, middle cheapness junk stocks, and cheap middle-quality stocks yield very high excess returns (19.78%, 19.18%, and 21.15%, respectively). These returns come at a cost, as they also have fairly high standard deviations (0.53, 0.57, and 0.51, respectively). Ball et al. (1995) argue that low-dollar value stocks (penny stocks) create issues in event studies (like ours) because they have very large positive returns over longer horizons with little impact on the market due to low market capitalizations. This leads to a large sensitivity of returns of portfolios including these stocks to alternative specifications (caused by large impacts of brokerage/transaction costs and liquidity problems). Indeed, setting different low-dollar thresholds changes the excess returns from these two portfolios noticeably, while other portfolio statistics remain robust. A noticeable portion of the large positive excess returns for these portfolios comes from investing in low-dollar stocks. Considering their high standard deviation,

TABLE 9 Return statistics for portfolios formed on quality and cheapness.

Panel A: Monthly returns ($t = 1/12$)						Panel B: Annual returns ($t = 1$)					
	Expensive	Middle	Cheap	Total	CME	Expensive	Middle	Cheap	Total	CME	
μ											
Junk	-0.58%	0.21%	-0.53%	-0.40%	0.05%	5.23%	19.18%	19.78%	8.71%	14.55%	
Middle	-0.02%	0.31%	0.46%	0.17%	0.48%	5.09%	10.32%	21.15%	8.73%	16.06%	
Quality	0.50%	0.70%	0.91%	0.65%	0.41%	6.77%	10.98%	18.23%	10.13%	11.46%	
Total	0.02%	0.41%	0.53%			5.35%	10.65%	20.60%			
QMJ	1.08%***	0.49%***	1.44%***			1.54%***	-8.20%	-1.55%			
σ											
Junk	0.12	0.13	0.14	0.12	0.03	0.48	0.57	0.53	0.50	0.05	
Middle	0.10	0.10	0.12	0.10	0.02	0.40	0.40	0.51	0.41	0.11	
Quality	0.08	0.08	0.10	0.08	0.02	0.29	0.30	0.38	0.30	0.09	
Total	0.10	0.10	0.12			0.39	0.38	0.49			
QMJ	-0.04***	-0.05***	-0.04***			-0.18***	-0.27***	-0.14***			
Pos											
Junk	49.07%	53.69%	49.81%	51.85%	0.74%	50.03%	60.62%	61.89%	53.63%	11.86%	
Middle	51.61%	53.92%	55.90%	54.68%	4.29%	52.82%	59.72%	67.79%	61.58%	14.96%	
Quality	54.37%	56.77%	58.72%	56.36%	4.34%	59.56%	66.59%	71.27%	68.37%	11.71%	
Total	50.16%	53.04%	56.17%			52.68%	57.23%	64.78%			
QMJ	5.30%	3.08%	8.91%			9.53%	5.96%	9.38%			
SR											
Junk	-4.98%	1.59%	-3.67%	-3.29%	1.32%	10.95%	33.65%	37.65%	17.25%	26.70%	
Middle	-0.23%	2.96%	3.78%	1.66%	4.00%	12.68%	25.67%	41.63%	21.23%	28.95%	
Quality	6.36%	8.59%	8.95%	7.98%	2.59%	23.13%	36.80%	47.45%	33.44%	24.32%	
Total	0.15%	4.17%	4.50%			13.61%	27.97%	42.22%			
QMJ	11.35%	7.00%	12.62%			12.18%	3.15%	9.80%			

TABLE 9 (Continued)

		Panel C: Number of observations					Panel D: Share of portfolio in the sample				
		Expensive	Middle	Cheap	Sum	CME	Expensive	Middle	Cheap	Sum	CME
<i>N</i>											
	% of <i>N</i>										
Junk		12,549	3893	265	16,707	-12,284	2.95%	0.91%	0.06%	3.92%	-2.88%
Middle		138,574	159,010	20,373	317,957	-118,201	32.53%	37.33%	4.78%	74.64%	-27.75%
Quality		26,568	60,043	4692	91,303	-21,876	6.24%	14.10%	1.10%	21.43%	-5.14%
Sum		177,691	222,946	25,330	425,967		41.71%	52.34%	5.95%	100.00%	
QMJ		14,019	56,150	4427			3.29%	13.18%	1.04%		

Note: This table shows average excess total stock returns (μ), standard deviations of returns (σ), the share of positive returns (Pos), Sharpe Ratios (SR), number of observations (*N*) as well as the share of the respective portfolio in the entire sample (% of *N*) for equal-weighted portfolios that are classified as junk, middle, and quality on the y-axis and classified as expensive, middle and cheap on the x-axis for a total amount of 425,967 observations of US-Firms from June 2002 to July 2021 (230 months). Each month, firms are classified based on cheapness V_0^{Baves} and then further classified based on quality through VScore. For V_0^{Baves} , portfolios are constructed based on the posterior predictive probability $p(VB^*|VB)$, so that overvalued (undervalued) portfolios include all firms with a posterior predictive probability lower than 20% (larger than 80%). Panel A shows characteristics of monthly returns following the portfolio formation date, while Panel B Shows the same statistics for annual returns. TSR is the total stock return for the subsequent timeframe, including dividend payments if they fall into the respective window. Rf is the 1-month and 10-year US-treasury bond yield at the time. Pos is the share of positive excess returns (TSR-rf) in the portfolio. SR is calculated as excess return divided by the respective standard deviation. For the CME and QMJ statistics, ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1 for the *t*-statistic or *F*-statistic, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%. Panel C and D show the number of stocks and share of the respective portfolio in the sample, respectively.

the low-dollar stocks entail a high risk of being cheap for a justifiable reason. The excess return, standard deviation, and share of positive returns for the cheap quality portfolio remain robust to alternative dollar value thresholds. Aside from this deviation, the results in Panel B are consistent with the returns implied by errors-in-expectations using V_0^{Bayes} .

8.4 | Time-series analysis of monthly returns on QMJ and CME

Even though the return statistics and regressions are insightful, they do not show the degree of consistency of the investment strategy over time. There may be times when market-wide dynamics cause the UMO strategy to fail (like, e.g., market bubbles or larger changes on input variables like benchmark yields or even a lack of under- or overvalued firms). In order to illustrate the consistency of the strategy, Figure 2 plots monthly differences in annual Sharpe Ratios between undervalued and overvalued portfolios. It is visible that there are only few cases where the difference is negative. The only two rather persistent windows of negative difference cover the financial crisis in 2008 & 2009 and the first 1½ years of the COVID pandemic (April 2020–August 2021). A potential explanation for the drawdowns is the weaker fundamental performance of firms during those windows; quality firms have more ground to lose compared to junk firms and thus experience larger drawdowns because the market anticipates the loss in quality.

In order to illustrate that, Figure 3 shows the share of cheap firms and the share of quality firms in the sample over time, combined with the evolution of the S&P500 price index over time. It is visible that the share of quality firms in the market fluctuates and ranges between 14.81% and 27.68%. There is a particularly low share of quality firms after the financial crisis of 2008–2009, and a noticeable dip during the COVID-crisis in 2020. It is (logically) more difficult to find quality firms in times of economic downturn. The share of cheap firms behaves the opposite way, as it exhibits peaks in times of crises, with the largest peaks reaching 22.42% in December 2008 and 33.31% in May 2020. It seems that the market anticipates the drops in quality (resulting in stronger drawdowns for quality firms), and then recovers when quality improves again. V_0^{Bayes} identifies firms based on absolute probability boundaries, so that in times of crises, more firms will be identified as undervalued because market expectations are lower. Expectations investing bets against the market—buy quality when expectations are low, sell

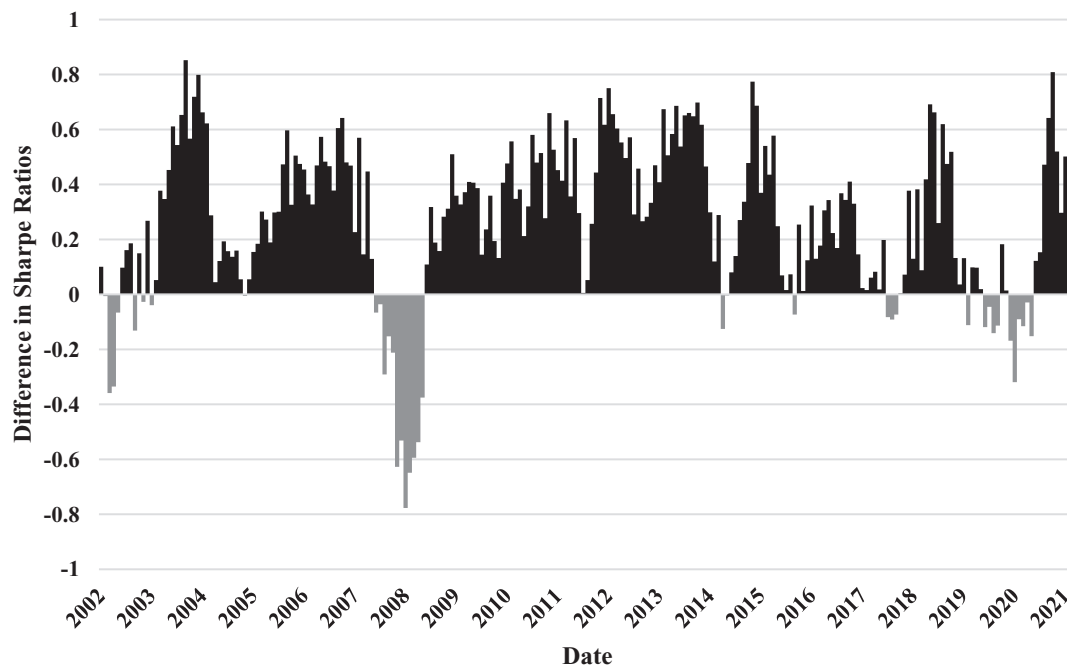


FIGURE 2 Monthly difference in annual sharpe ratios for UMO portfolios. This figure plots the (rolling) monthly difference in annual Sharpe Ratios between firms that are classified as undervalued and firms that are classified as overvalued (equal-weighted portfolios). Firms are classified as undervalued if they are cheap relative to their quality as implied by Figure 1, with the additional restriction of excluding the middle quality portfolios. The undervalued portfolio includes all firms with (a) high quality and high cheapness and (b) high quality and middle cheapness, while overvalued portfolio includes firms with (a) low quality and low cheapness and (b) low quality and middle cheapness. Sharpe Ratios are calculated as the annual excess total return (TSR- r_f) following portfolio formation, divided by the standard deviation (σ) of the excess returns in the portfolio. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30% .

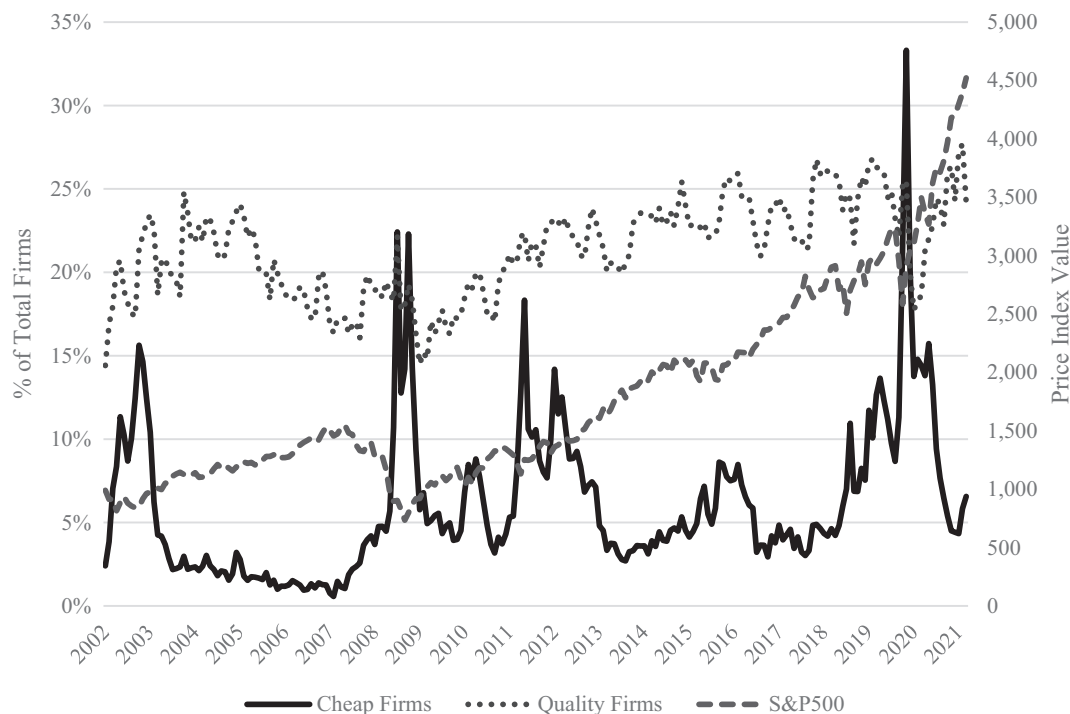


FIGURE 3 Bayesian Contrarian Investing – Time-Series of Cheap and Quality Firms. This figure plots the number of firms that are identified as cheap ($V_0^{\text{Bayes}} \geq 80\%$) and the number of quality firms (Firms with VScore greater than 6), both relative to all firms in the sample at the start of every month between June 2006 and July 2021 (230 months) on the left-hand side axis. V_0^{Bayes} is equal to the posterior predictive probability of intrinsic value to exceed price. On the right-hand side axis, it plots the index evolution of the S&P500 price index in the same timeframe.

when expectations are high. To a degree, it thus shares the idea of contrarian investing, similar to DeBondt and Thaler (1985), Chan (1988), Ball and Kothari (1989), Jones (1993), or Lakonishok et al. (1994). Overall, the strategy seems not only very consistent over time but also feasible even under extraordinary market circumstances.

9 | ROBUSTNESS TESTS AND SENSITIVITY ANALYSES

A main point of criticism in longer-run event studies like ours is the application of equal-weighted portfolios and potential liquidity issues caused by smaller stocks. Furthermore, the RMW factor (operating profitability) has a strongly positive and significant factor loading in the QMJ factor model regressions. We therefore perform three additional analyses:

- Test the differences in returns and alphas between equal-weighted and value-weighted portfolios.
- Control for the impact of liquidity on the results.
- Use quintiles of operating profitability as described by Novy-Marx (2013) as an alternative measure for quality.

Since value-weighted portfolios involve market capitalization, we had to exclude all observations for which market capitalization was not available at portfolio formation. This results in an adjusted sample of 418,686 observations. This has no significant impact on the return statistics but may lead to slight deviations for the equal-weighted portfolios compared to the previous sections.

9.1 | Equal-weighted versus value-weighted returns

Given that standard practice in factor research is to employ value-weighted portfolios, we performed the main analyses for value-weighted portfolios as well (only V_0^{Bayes} is tabulated). Table 10 shows comparison of regression results of monthly asset pricing tests for both equal-weighted and value-weighted portfolios. Similar to other anomalies, the alpha

TABLE 10 Asset pricing test comparisons for weightings.

V_0^{Bayes}		
Weighting	Equal-weights	Value-weights
Alpha (in %)	0.36	-0.06
	1.97**	-0.23
MKT	-0.11	-0.06
	-2.77***	-0.92
SMB	0.04	-0.07
	0.48	-0.59
RMW	0.17	0.36
	1.75*	2.45**
CMA	0.25	0.30
	2.43**	1.92*
UMD	-0.07	-0.05
	-2.04**	-0.91
QMJ	0.01	-0.16
	0.12	-2.01**
Adj. R^2	0.10	0.05

Note: This table shows six factor alphas for different operationalizations of equal-weighted and value-weighted high-minus-low cheapness portfolios (CME). The regression is a time-series regression of CME on six commonly applied risk factors. Alpha is the intercept and can be interpreted as the average excess return of the respective CME portfolios. MKT, SMB, RMW, and UMD are taken from Kenneth French's website. The descriptions of the five factors on the website are as follows: MKT is the excess return of the market portfolio, including all NYSE, AMEX, and NASDAQ firms. SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios and UMD is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. The CME factors are constructed using V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes} . The t statistics are shown below the coefficients, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

on CME turns negative when value-weights are employed. It, therefore, seems that the abnormal returns are driven by stocks with smaller market capitalizations, at least in the time series.

In order to examine whether the loss in significance of the alphas is also present for the nine UMO portfolios and in annual return windows, we also tabulated the annual returns for them in Table 11. Interestingly, the extreme returns on junk firms have a more severe impact when value-weighted portfolios are applied, as they exhibit extremely high returns and Sharpe ratios despite having the highest weighted standard deviations. These stocks also have an impact on the total junk stock portfolio, so that the returns on junk stocks become larger than the returns on quality stocks for value-weighted portfolios despite having a much smaller share of positive returns.

Value-weighted returns, therefore, do not necessarily provide insights into the probability distribution of future positive returns. Given that the number of stocks in these portfolios is small, this hints to a large impact of few stocks with large market capitalizations. In order to test whether this might be the case, we calculated Gini coefficient on the market capitalizations of the respective portfolio. It is visible that the coefficients are high for all portfolios, generally increasing from the bottom left to the top right. This shows that there are substantial differences in market capitalization in all portfolios, with the highest disparity for cheap junk portfolios.

9.2 | Controlling for liquidity

Given that the abnormal returns on equal-weighted portfolios are driven by stocks with smaller market capitalizations, it could be that liquidity issues might reduce the consistency and feasibility of the investment strategy. In order to control for the impact of liquidity, we ran control regressions including that traded liquidity factor of Pastor and Stambaugh (2003). The results are shown in Table 12.

TABLE 11 Annual return statistics for value-weighted portfolios.

	Expensive	Middle	Cheap	Total	CME
Panel A: Annual returns ($t = 1$)—value—weighted					
TSR-rf					
Junk	9.61%	29.02%	46.12%	13.34%	36.51%***
Middle	8.25%	7.49%	19.74%	8.19%	11.49%***
Quality	11.37%	9.64%	13.57%	10.24%	2.20%***
Total	9.46%	8.76%	16.94%		
QMJ	1.76%***	-19.37%***	-32.55%***		
σ					
Junk	0.39	0.54	0.44	0.43	0.05***
Middle	0.34	0.34	0.40	0.33	0.06***
Quality	0.29	0.24	0.28	0.27	0.00
Total	0.32	0.28	0.35		
QMJ	-0.10***	-0.29***	-0.16***		
SR					
Junk	24.78%	54.10%	104.28%	31.18%	79.51%
Middle	24.60%	23.85%	49.97%	25.03%	25.37%
Quality	39.59%	39.65%	48.09%	38.44%	8.50%
Total	29.49%	31.00%	47.34%		
QMJ	14.82%	-14.45%	-56.19%		
	Expensive	Middle	Cheap	Total	CME
Panel B: Gini coefficients (market capitalization)					
Gini					
Junk	0.75	0.81	0.85	0.77	0.10
Middle	0.79	0.82	0.82	0.79	0.03
Quality	0.73	0.73	0.79	0.74	0.06
Total	0.77	0.80	0.81		
QMJ	-0.02	-0.07	-0.06		
	Expensive	Middle	Cheap	Total	CME
Panel C: Share of positive returns					
Pos					
Junk	50.35%	60.91%	61.48%	53.02%	11.13%
Middle	52.86%	59.81%	68.08%	57.31%	15.22%
Quality	59.50%	66.74%	71.10%	64.87%	11.60%
Total	53.68%	61.71%	68.57%		
QMJ	9.15%	5.83%	9.62%		
	Expensive	Middle	Cheap	Total	CME
Panel D: Number of observations per portfolio					
N					
Junk	12,113	3799	270	16,182	-11,843
Middle	135,827	156,398	19,849	312,074	-115,978

(Continues)

TABLE 11 (Continued)

	Expensive	Middle	Cheap	Total	CME
Panel D: Number of observations per portfolio					
Quality	26,123	59,674	4633	90,430	-21,490
Total	174,063	219,871	24,752	418,686	
QMJ	14,010	55,875	4363		

Note: This table shows average excess total stock returns (μ), weighted standard deviations of returns (σ), Sharpe Ratios (SR), Gini coefficients the share of positive returns (Pos), number of observations (N) as well as the share of the respective portfolio in the entire sample (% of N) for value-weighted portfolios that are classified as junk, middle, and quality on the y-axis and classified as expensive, middle and cheap on the x-axis for a total amount of 418,686 observations of US-Firms from June 2002 to July 2021 (230 months). Each month, firms are classified based on cheapness via different proxies for intrinsic value (V_0^B , V_0^{STE} , V_0^{FL} , V_0^H , and V_0^{Bayes}) and then further classified based on quality through VScore. The cheapness portfolios for overvalued (undervalued) firms for V_0^B , V_0^{STE} , V_0^{FL} , and V_0^H consist of the 20% lowest (highest) observations for intrinsic value estimates at portfolio formation. For V_0^{Bayes} , portfolios are constructed based on the posterior predictive probability $p(VB^*|VB)$, so that overvalued (undervalued) portfolios include all firms with a posterior predictive probability lower than 20% (larger than 80%). Panel A shows basic return statistics; TSR is the total stock return for the subsequent timeframe, weighted by the respective market capitalization, including dividend payments if they fall into the respective window. rf is the 1 month and 10-year US-treasury bond yield at the time. SR is calculated as excess return divided by the respective weighted standard deviation. For the CME and QMJ statistics, ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1 for the t -statistic or F -statistic, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%. Panel B includes Gini coefficients of the distribution of market capitalization within the respective portfolio. Panels C and D show the share of positive returns and number of observations in the portfolio, respectively.

TABLE 12 Control regressions for Pastor and Stambaugh (2003)'s liquidity factor.

V_0^{Bayes}	Equal-weighted		Value-weighted	
Alpha (in %)	0.36	0.36	-0.32	-0.32
	1.97**	1.97*	-1.13	-1.12
MKT	-0.11	-0.11	-0.01	0.00
	-2.77***	-2.72***	-0.16	-0.07
SMB	0.04	0.04	0.09	0.10
	0.48	0.48	0.78	0.85
RMW	0.17	0.17	0.16	0.15
	1.75*	1.74*	1.07	1.03
CMA	0.25	0.24	0.34	0.31
	2.43**	2.29**	2.16**	1.89*
UMD	-0.07	-0.07	-0.03	-0.03
	-2.04**	-2.03**	-0.57	-0.59
QMJ	0.01	0.01	0.16	0.16
	0.12	0.12	1.32	1.35
LIQ		0.00		-0.03
		-0.03		-0.5
Adj. R^2	0.08	0.07	0.03	0.02

Note: This table shows six- and seven-factor alphas for different operationalizations of equal-weighted and value-weighted high-minus-low cheapness portfolios (CME) for . The regression is a time-series regression of CME on six commonly applied risk factors. Alpha is the intercept and can be interpreted as the average excess return of the respective CME portfolios. MKT, SMB, RMW, and UMD are taken from Kenneth French's website. The descriptions of the five factors on the website are as follows: MKT is the excess return of the market portfolio, including all NYSE, AMEX, and NASDAQ firms. SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios and UMD is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. LIQ is the traded liquidity factor from portfolios sorted on historical (not predicted) liquidity betas as computed in Pastor and Stambaugh (2003), taken from Robert Stambaugh's website. The CME factors are constructed using V_0^{Bayes} . The t statistics are shown below the coefficients, with ***, **, and * indicating significance levels of 0.01, 0.05, and 0.1, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%.

Aside from a few minor effects on the other factors, alphas are unaffected by the introduction of LIQ. We do not perform additional tests for the annual return window because liquidity is not as problematic over longer investment horizons.

TABLE 13 Return statistics for portfolios formed on operating profitability and cheapness.

Panel A: Monthly returns ($t = 1/12$)							Panel B: Annual returns ($t = 1$)										
	Expensive	Middle	Cheap	Total	CME		Expensive	Middle	Cheap	Total	CME		Expensive	Middle	Cheap	Total	CME
μ						μ											
Junk	-1.01%	-0.70%	-0.46%	-0.82%	0.54%***	Junk	-0.22%	3.58%	9.89%	2.36%	10.11%***						
Middle	0.17%	0.55%	0.96%	0.43%	0.79%***	Middle	5.71%	11.17%	25.36%	9.95%	19.64%***						
Quality	0.75%	1.00%	1.20%	0.89%	0.45%***	Quality	10.34%	15.31%	29.60%	13.33%	19.26%***						
Total	0.03%	0.43%	0.54%			Total	5.40%	10.71%	20.94%								
QMJ	1.75%***	1.70%***	1.66%***			QMJ	10.56%***	11.73%***	19.71%***								
σ						σ											
Junk	0.11	0.11	0.12	0.11	0.01	Junk	0.40	0.42	0.45	0.42	0.05						
Middle	0.10	0.09	0.12	0.10	0.02	Middle	0.38	0.36	0.49	0.38	0.11						
Quality	0.10	0.10	0.13	0.10	0.03	Quality	0.41	0.39	0.58	0.41	0.17						
Total	0.10	0.10	0.12			Total	0.39	0.38	0.49								
QMJ	-0.01	-0.01	0.01			QMJ	0.01	-0.03*	0.13								
Pos						Pos											
Junk	48.37%	50.03%	51.74%	49.40%	3.38%	Junk	47.71%	52.11%	60.11%	50.76%	12.41%						
Middle	52.18%	55.32%	58.35%	54.31%	6.18%	Middle	54.28%	63.01%	72.30%	60.26%	18.02%						
Quality	54.67%	56.86%	59.47%	55.90%	4.80%	Quality	58.27%	65.66%	72.63%	62.37%	14.37%						
Total	51.87%	54.74%	56.41%			Total	53.68%	61.71%	68.57%								
QMJ	6.30%	6.83%	7.73%			QMJ	10.56%	13.54%	12.52%								
SR						SR											
Junk	-9.31%	-6.32%	-3.99%	-7.48%	5.31%	Junk	-0.55%	8.52%	21.95%	5.68%	22.50%						
Middle	1.77%	5.80%	8.21%	4.44%	6.44%	Middle	15.10%	30.92%	52.21%	26.24%	37.11%						
Quality	7.57%	10.12%	9.43%	8.91%	1.86%	Quality	25.08%	38.78%	51.10%	32.52%	26.02%						
Total	0.33%	4.35%	4.61%			Total	13.74%	28.19%	42.79%								
QMJ	16.87%	16.45%	13.42%			QMJ	25.63%	30.25%	29.16%								

(Continues)

TABLE 13 (Continued)

N	Panel C: Number of observations					Panel D: Share of Portfolio in the sample				
	Expensive	Middle	Cheap	Sum	CME	Expensive	Middle	Cheap	Sum	CME
Junk	39,610	36,591	7624	83,825	-31,986	9.46%	8.74%	1.82%	20.02%	-7.64%
Middle	95,290	140,624	15,122	251,036	-80,168	22.76%	33.59%	3.61%	59.96%	-19.15%
Quality	39,163	42,656	2006	83,825	-37,157	9.35%	10.19%	0.48%	20.02%	-8.87%
Sum	174,063	219,871	24,752	418,686		41.57%	52.51%	5.91%	100.00%	
QMJ	-447	6065	-5618			-0.11%	1.45%	-1.34%		

Note: This table shows average excess total stock returns (μ), standard deviations of returns (σ), the share of positive returns (Pos), Sharpe Ratios (SR), number of observations (N) as well as the share of the respective portfolio in the entire sample (% of N) for equal-weighted portfolios that are classified as junk, middle and quality on the y-axis and classified as expensive, middle and cheap on the x-axis for a total amount of 418,686 observations of US-Firms from June 2002 to July 2021 (230 months). Each month, firms are classified based on V_0^{Bayes} and then further classified based on quality through quantiles of operating profitability (calculated as net operating profit after tax divided by net operating assets). For V_0^{Bayes} , portfolios are constructed based on the posterior predictive probability $p(VB^*|VB)$, so that overvalued (undervalued) portfolios include all firms with a posterior predictive probability lower than 20% (larger than 80%). Panel A shows characteristics of monthly returns following the portfolio formation date, while Panel B shows the same statistics for annual returns. TSR is the total stock return for the subsequent timeframe, including dividend payments if they fall into the respective window. Rf is the 1 month and 10-year US-treasury bond yield at the time. Pos is the share of positive excess returns (TSR-rf) in the portfolio. SR is calculated as excess return divided by the respective standard deviation. For the CME and QMJ statistics, ***, ** and * indicate significance levels of 0.01, 0.05, and 0.1 for the t-statistic or F-statistic, respectively. If a firm is delisted within the return window, we follow Shumway (1997) and set it to -30%. Panel C and D show the number of stocks and share of the portfolio in the sample, respectively.

9.3 | Operating profitability as an alternative measure for quality.

Finally, the large correlation and highly significant positive coefficient of RMW indicate that operating profitability (as shown in Novy-Marx (2013)) might be a suitable substitute for VScore. Given that operating profitability is included in VScore as well, the positive correlation is not surprising. In order to test whether RMW may lead to similar inferences, we formed quantile-based portfolios on operating portfolios on each portfolio formation date, tabulated in Table 13.

The results on operating profitability show that (a) there are fewer stocks in the cheap quality portfolio than in the VScore portfolio and (b) that the high Sharpe ratios for the cheap portfolios are caused by few heavy return tails in both directions, which is not the case for VScore. The operating profitability portfolios, therefore, seem to pick up on risk, resulting in larger standard deviations. Nonetheless, the high shares of positive returns for the undervalued portfolios indicate that operating profitability is a suitable measure for quality when combined with the Bayesian cheapness measure. Using RMW in investment strategies comes at the cost of requiring market-wide data, while VScore can be calculated and evaluated for each firm separately. This is particularly useful for small-budget, less sophisticated investors.

10 | CONCLUDING REMARKS AND IMPLICATIONS FOR FUTURE RESEARCH

In line with prior research, we provide evidence that simple investment strategies that combine factors of QMJ with factors of CME yield high excess returns at low risk. In turn, our results indicate that the market tends to undervalue quality and overvalue junk, creating both cheap quality and expensive junk stocks. We exploit that dynamic by creating UMO portfolios, where undervalued (overvalued) portfolios include stocks that appear cheap (expensive) relative to their quality. We further contribute to the field by assessing the empirical validity of different proxies for cheapness, first in isolation and then as part of portfolios formed on QMJ. We find that a probabilistic Bayesian estimate of intrinsic value (V_0^{Bayes}) yields superior Sharpe ratios and higher probabilities for positive returns compared to the common B/P-based strategy and other point estimates in an annual return window. For monthly returns, the results for V_0^{Bayes} are comparable to the literature's standard approach, the B/P ratio. We control for common risk factors for the UMO strategy based on several multi-factor models through calendar-time regressions. None of the commonly used risk factors (like the market return, size, momentum, or the investment strategy) can explain the abnormal returns, so that alphas on the UMO portfolios remain positive and statistically significant. In our sample, UMO portfolios increase monthly and annual Sharpe ratios by 7.59% and 24.46%, respectively. This improves upon the difference in Sharpe ratios for both QMJ and CME portfolios in our sample, with respective monthly and annual values of 11.27% and 16.19% for QMJ and 5.68% and 20.91% for CME. We do (of course) not claim to employ the ideal combination of input factors, so that alternative prior specifications for V_0^{Bayes} and other binaries for VScore can be tested in future research. There is a large share of stocks that do not follow the patterns implied by errors-in-expectations (roughly one third), so that a further analysis of them (through a more thorough fundamental analysis) is still advisable to separate winners from losers. While we do not find that equal-weighted portfolios in our analysis exhibit liquidity issues that hinder consistent investment, the regression alphas for monthly returns become statistically insignificant for value-weighted portfolios. It thus seems that abnormal returns on cheap quality stocks are driven by smaller stocks with lower investment emphasis and market capitalization. On the contrary, value-weighted returns lead to few stocks dominating portfolio returns, without necessarily providing insights into the probability for positive investment outcomes. The high Gini coefficients of market capitalizations in all portfolios suggest that this might be the case. We further find that portfolios formed on operating profitability as proposed by Novy-Marx (2013) yield similar results compared to the standalone index-based measure. Since the general concept behind the two quality measures is similar, that is not surprising.

We do, however, believe that Bayesian inference can improve value investment strategies due to its (almost unavoidable) transparency about the uncertainty involved in the decision-making process and the possibility to regulate the output through the prior and likelihood. Our results reinforce that impression and, therefore, have implications on the methodology in both research and practice that is concerned with equity valuation.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

ORCID

Marcel Rueenauffer  <https://orcid.org/0000-0003-0482-4958>

ENDNOTES

- ¹ Kothari (2001) and Ang and Zhang (2014) provide overviews of problems related to longer-horizon event studies. Since valuation models have specific data requirements and longer term event studies require a specification of a “normal” benchmark return and risk, misspecifications and data problems can limit the usefulness of the results.
- ² Other studies emphasize the capability of Bayesian inference to intuitively and transparently resolve decision problems under uncertainty, like Cremers (2002), Goldstein (2006), Johnstone (2018, 2021) or Breuer and Schütt (2021).
- ³ Fama and French (1992) demonstrated that including the difference in returns between firms with high and low B/P and Size as additional explanatory variables increases the explainability of cross-sectional returns.
- ⁴ See Penman and Reggiani (2018) for a detailed analysis of the value trap and why B/P alone does not suffice to avoid it.
- ⁵ Operators for expectational values (all parameters with $t > 0$) are suppressed for simplicity of notation.
- ⁶ This negative bias is also demonstrated in empirical research, such as Francis et al. (2000) or Asness et al. (2019).
- ⁷ The posterior predictive distribution is approximated by using the posterior samples for α and β to sample from the likelihood. Due to the law of total probability, this results in the posterior predictive.
- ⁸ STAN is a C++ library for Bayesian inference with interfaces to common statistical packages (such as R or STATA) and has the advantage of requiring only basic input information to operationalize complex models.
- ⁹ Francis et al. (2000) set negative estimates to zero, but price and value must be strictly positive, so we set it to the lowest realistically feasible value.
- ¹⁰ VScore measures quality as a sum of binary variables, where the binary variables are set to 1 if a particular condition is fulfilled (e.g., the ROE for the previous fiscal year must exceed the cost of capital). Huefner et al. (2021) provide detailed descriptions of the eight binaries in their appendices.
- ¹¹ For a recent review of research on equal-weighted versus value-weighted portfolios, see Plyakha et al. (2021).
- ¹² Descriptions of the factors can be found on Kenneth French's website.
- ¹³ See Ohlson (2020) for a recent discussion on the drawbacks of frequentist statistics in the age of “big data”.
- ¹⁴ We understand economic significance in the context of equity valuation as impacting the choices of investors (as subjective as that is), in that they would e.g. prefer to invest in a quality portfolio over a junk portfolio based on our findings. For further discussions on the matter, see Ziliak and McCloskey (2007) and Engsted (2009).

REFERENCES

- Ahmed, A., Morton, T., & Schaefer, T. (2000). Accounting conservatism and the valuation of accounting numbers: Evidence on the Feltham-Ohlson (1996) model. *Journal of Accounting, Auditing and Finance*, 15(3), 271–292.
- Anesten, S., Moeller, N., & Skogsvik, K. (2019). The pricing accuracy of alternative equity valuation models: Scandinavian evidence. *Journal of International Financial Management & Accounting*, 31(1), 5–34.
- Ang, J. S., & Zhang, S. (2014). Evaluating long-horizon event study methodology. In *Handbook of financial econometrics and statistics* (1st ed., pp. 383–411). Springer.
- Arnott, R. D., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. T. (2021). Reports of value's death may be greatly exaggerated. *Financial Analysts Journal*, 77(1), 44–67.
- Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24, 34–112.
- Ball, R., & Kothari, S. P. (1989). Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics*, 25, 51–74.
- Ball, R., Kothari, S. P., & Shanken, J. (1995). Problems in measuring portfolio performance: An application to contrarian investment strategies. *Journal of Financial Economics*, 38, 79–107.
- Barth, M. E., Beaver, W. H., Hand, J. R. M., & Landsman, W. R. (2005). Accruals, accounting-based valuation models, and the prediction of equity values. *Journal of Accounting, Auditing & Finance*, 20(4), 311–345.
- Basu, S. (1997). The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics*, 24, 3–37.
- Bird, R., & Gerlach, R. (2003). The good and the bad of value investing: Applying a Bayesian approach to develop enhancement models. *SSRN Unpublished Working Paper*.
- Breuer, M., & Schütt, H. (2021). Accounting for uncertainty: An application of Bayesian methods to accruals models. *Review of Accounting Studies*, 28, 726–768.
- Chan, K. K. (1988). On the contrarian investment strategy. *Journal of Business*, 61(2), 147–163.
- Chang, W.-J., Landsman, W. R., & Richardson, G. D. (2012). Selecting an accounting-based valuation model. *Working Paper INSEAD/UNC*.
- Chiang, K. C. H. (2002). Portfolio return metric: Equal weights versus value weights. *SSRN Working Paper*.

- Choi, Y.-S., O'Hanlon, J. F., & Pope, P. F. (2006). Conservative accounting and linear information valuation models. *Contemporary Accounting Research*, 23(1), 73–101.
- Courteau, L., Kao, J. L., & Richardson, G. D. (2001). Equity valuation employing the ideal versus ad-hoc terminal value expressions. *Contemporary Accounting Research*, 18(4), 625–661.
- Cremers, K. J. M. (2002). Stock return predictability: A Bayesian model selection perspective. *The Review of Financial Studies*, 15(4), 1223–1249.
- DeBondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40(3), 793–805.
- Dittmann, I., & Maug, E. G. (2008). Biases and error measures: How to compare valuation methods. *Mannheim Finance Working Paper*.
- Dontoh, A., Radakrishnan, S., & Ronen, J. (2007). Is stock price a good measure for assessing value-relevance of earnings? An empirical test. *Review of Managerial Science*, 1(1), 3–45.
- Engsted, T. (2009). Statistical vs. economic significance in economics and econometrics: Further comments on McCloskey and Ziliak. *Journal of Economic Methodology*, 16(4), 393–408.
- Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283–306.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427–465.
- Francis, J., Olsson, P., & Oswald, D. R. (2000). Comparing the accuracy and Explainability of dividend, free cash flow, and abnormal earnings equity value estimates. *Journal of Accounting Research*, 38(1), 45–70.
- Frankel, R., & Lee, C. M. C. (1998). Accounting valuation, market expectation, and cross-sectional returns. *Journal of Accounting and Economics*, 25, 283–319.
- Gao, Z., Myers, J. N., Myers, L. A., & Wu, W.-T. (2019). Can a hybrid method improve equity valuation? An empirical evaluation of the Ohlson and Johannesson (2016) model. *The Accounting Review*, 94(6), 227–252.
- Goldstein, M. (2006). Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, 1(3), 403–420.
- Graham, B., & Dodd, D. (1934). *Security analysis: The classic 1934 edition*. McGrawHill.
- Henschke, S. (2009). *The accuracy of equity valuation models*. Gabler.
- Higgins, H. N., Nandram, B., & Lu, Q. F. (2006). Predicting stock Price by applying the residual income model and Bayesian statistics. AAA 2007 Financial Accounting and Reporting Section (FARS) Meeting Paper. <https://doi.org/10.2139/ssrn.926183>
- Ho, K.-C., Lee, S. C., Lin, C.-T., & Yu, M.-T. (2017). A comparative analysis of accounting-based valuation models. *Journal of Accounting, Auditing and Finance*, 32(4), 561–575.
- Huefner, B., Rueenaufner, M., & Boesch, M. (2021). Does the incongruence of market expectations with fundamentals explain stock return patterns? *SSRN Working Paper*.
- Iqbal, A., Rajgopal, S., Srivastava, A., & Zhao, R. (2022). Value of internally generated intangible capital. *SSRN Working Paper*.
- Johnstone, D. (2018). Accounting theory as a Bayesian discipline. *Foundations and Trends in Accounting*, 13(1–2), 1–266.
- Johnstone, D. (2021). Accounting research and the significance test crisis. *Critical Perspectives on Accounting*, 89. <https://doi.org/10.1016/j.cpa.2021.102296>
- Jones, S. L. (1993). Another look at time-varying risk and return in a long-horizon contrarian strategy. *Journal of Financial Economics*, 33(1), 119–144.
- Jorgensen, B. N., Lee, Y. G., & Yoo, Y. K. (2011). The valuation accuracy of equity value estimates inferred from conventional empirical implementations of the abnormal earnings growth model: US evidence. *Journal of Business Finance & Accounting*, 38(3), 446–471.
- Khan, M., & Watts, R. L. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of Accounting and Economics*, 2(3), 132–150.
- Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, 31, 105–231.
- Kothari, S. P., & Warner, J. B. (1997). Measuring long-horizon security price performance. *Journal of Financial Economics*, 43(3), 301–339.
- Lakonishok, J., Shleifer, A., & Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5), 1541–1578.
- Lee, C. M. C. (2015). Value investing: Bridging theory and practice. *China Accounting and Finance Review*, 16(2), 10–38.
- Lev, B., & Gu, F. (2014). *The end of accounting and the path forward for investors*. Wiley and Sons.
- Li, K. K., & Mohanram, P. (2019). Fundamental analysis: Combining the search for quality with the search for value. *Contemporary Accounting Research*, 36(3), 1–36.
- Mohanram, P. S. (2005). Separating winners from losers among low book-to-market stocks using financial statement analysis. *Review of Accounting Studies*, 10, 133–170.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1), 1–28.
- Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11(2), 661–687.
- Ohlson, J. A. (2020). Researchers' data analysis choices: An excess of false positives? *Working Paper*.
- Pastor, L., & Stambaugh, R. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3), 642–685.
- Penman, S. H. (2011). *Accounting for value*. Columbia University Press.
- Penman, S. H., & Reggiani, F. (2013). Returns to buying earnings and book value: Accounting for growth and risk. *Review of Accounting Studies*, 18, 1021–1049.
- Penman, S. H., & Reggiani, F. (2018). Fundamentals of value vs. growth investing and an explanation for the value trap. *Financial Analysts Journal*, 74(4), 103–119.
- Penman, S. H., Reggiani, F., Richardson, S. A., & Tuna, I. (2018). A framework for identifying accounting characteristics for asset pricing models, with an evaluation of book-to-price. *European Financial Management Journal*, 24, 488–520.

- Penman, S. H., & Sougiannis, T. (1998). A comparison of dividend, cash flow, and earnings approaches to equity valuation. *Contemporary Accounting Research*, 15(3), 343–383.
- Penman, S. H., & Zhang, X.-J. (2002). Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review*, 77(2), 237–264.
- Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment- q relation. *Journal of Financial Economics*, 123, 251–272.
- Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to Separate winners from losers. *Journal of Accounting Research*, 38, 1–41.
- Piotroski, J. D., & So, E. C. (2012). Identifying expectation errors in value/glamour strategies: A fundamental analysis approach. *Review of Financial Studies*, 25(9), 2841–2875.
- Plyakha, Y., Uppal, R., & Vilkov, G. (2021). Equal or value weighting? Implications for asset-pricing tests. In *Financial risk management and modeling* (pp. 295–347). Cham.
- Rueenaufner, M. (2022). Equity valuation: To Bayes or not to Bayes? *SSRN Working Paper*.
- Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance*, 51(1), 327–340.
- Skogsvik, K., & Juettner-Nauroth, B. (2013). A note on accounting conservatism in residual income and abnormal earnings growth equity valuation. *The British Accounting Review*, 45, 70–80.
- Walkshäusl, C. (2020). Piotroski's FSCORE: International evidence. *Journal of Asset Management*, 21, 106–118.
- Ying, J., Kuo, L., & Seow, G. S. (2005). Forecasting stock prices using a hierarchical Bayesian approach. *Journal of Forecasting*, 24, 39–59.
- Zhang, X.-J. (2000). Conservative accounting and equity valuation. *Journal of Accounting and Economics*, 29, 125–149.
- Ziliak, S. T., & McCloskey, D. N. (2007). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press.

How to cite this article: Huefner, B., Rueenaufner, M., & Boesch, M. (2023). Value investing via Bayesian inference. *Review of Financial Economics*, 41, 465–492. <https://doi.org/10.1002/rfe.1185>