Can the representativeness heuristic explain growth-related anomalies?

Miao He*, Nishad Kapadia**, Sheri Tice**

January, 2026

Abstract

We show that the abnormal returns to four prominent growth-related anomalies—asset growth, sales growth, book-to-market, and analyst expected long-term growth—are related to the representativeness heuristic of Kahneman and Tversky (1972), as formalized in models of diagnostic investor expectations (e.g., Bordalo, Gennaioli, La Porta, and Shleifer, 2019). We construct a measure of how representative the stereotype of 'the next Google' is of high growth in the recent past. We show that this measure predicts time-variation in the premia of all four growth anomalies both in and put-of-sample. Analysts overreact to news when representativeness is high, consistent with the mechanism in the Bordalo, Gennaioli, La Porta, and Shleifer (2019) model.

* David Eccles School of Business, University of Utah, Salt Lake City, UT 84112

**A.B. Freeman School of Business, Tulane University, New Orleans, LA 70018.

Miao He: miao.he@eccles.utah.edu; Nishad Kapadia: nkapadi@tulane.edu; Sheri Tice: stice@tulane.edu

Growth related anomalies are an important feature of the cross-section of returns. Firms with rapid past or expected future growth, defined by a variety of different measures—asset growth, sales growth, analyst expected long-term growth, and book-to-market—have low average returns. Book-to-market has long been the basis of the value factor in the Fama and French (1993) model and asset growth is now a factor in both popular characteristic-based factor models (Fama and French, 2015, and Hou, Xue, Zhang, 2015). One rationale for the existence of a premium for asset growth is that it reflects investment and firms that invest a lot, all else equal, should have low discount rates. Yet, as Cooper, Gulen and Ion (2024) point out, asset growth is not a particularly good measure of investment and better measures do not share its effects on returns. Moreover, the other growth measures – sales growth, book-to-market, and analyst expected growth are even less directly linked to investment. This raises a natural question: if not investment, what explains why high-growth stocks have low returns?

The economic similarity between these measures suggests that these anomalies share a common underlying explanation. In this paper, we propose and test a single unifying hypothesis: these four growth-related anomalies arise from the representativeness heuristic, originally introduced by Kahneman and Tversky (1972) and later formalized in models of diagnostic expectations in a series of papers, including Gennaioli and Shleifer (2010), Bordalo, Coffman, Gennaioli, and Shleifer (2016), and Bordalo, Gennaioli, and Shleifer (2018). Individuals who rely on the representativeness heuristic use stereotypes when making decisions, overweighting traits that are more prevalent in a specific group relative to a reference group. Bordalo, Coffman, Gennaioli and Shleifer (2016) illustrate the representativeness heuristic using the stereotype that 'Florida residents are elderly.' They show that the trait of being elderly is representative of Florida because the relative frequency, P(t|Florida) /P(t|US), is maximized if the trait *t* is age >65. They argue that this gives rise to the "elderly Floridian" stereotype, even though in an objective sense, the frequency of Florida residents with age greater than 65 is smaller than those in the 20-44 group.

The stereotype that we investigate is that firms with high growth are the "next Google". As Bordalo, Gennaioli, La Porta, and Shleifer (BGLS1; 2019) point out, there are very few firms with exceptional performance; however, these firms are concentrated in the group of firms with high analyst expected growth. Hence, they argue that the "Google-type" is representative of high expected analyst growth firms, leading to overvaluation and low expected returns for this group of firms. We hypothesize and test whether this is true for all four growth measures. We construct an empirical measure of the degree of representativeness of the Google-type for high growth at each point in time, and test whether this single variable predicts time-variation in expected returns of all four growth-related anomalies.

To motivate our empirical measure of representativeness in the context of high growth firms, consider the theoretical definition of representativeness in Gennaioli and Shleifer (2010):

$$R(t, G) = \frac{h(T = t|G)}{h(T = t| - G)} \tag{1}$$

where h is the true distribution of traits. This variable thus captures whether a trait 't' ("Google-type") is overrepresented in the high growth group ('G') relative to the reference group ('-G'). In models of diagnostic expectations, investor beliefs reflect a 'kernel of truth'—investors use the true probability distribution distorted by a factor proportional to R, in forming their expectations.

We make three key decisions in constructing an empirical measure of R that are guided by implications of the model in BGLS1. First, we define realized Google-type firms as those with returns greater than a threshold of θ (100% in our main specification) over a year.[1] Focusing on the right tail is consistent with the implication of the BGLS1 model that the most representative firm types for high expected growth firms are in the right tail of firm performance. We could alternatively use accounting variables rather than returns to identify realized Google-type firms. We choose returns because they are

[1] This is similar to the definition of Jackpots in Conrad, Kapadia, and Xing (2014).

timely: returns in a given year may reflect expectations of several years of future earnings growth, thereby mitigating potential issues related to firms dropping out of the sample. Also, returns are likely to matter more to investors than accounting variables because they directly affect an investor's wealth.

Second, our empirical measure of the degree of representativeness of the Google-type for high growth firms at a given point in time is motivated by equation (1) above. We run univariate monthly cross-sectional regressions of $J_{i,t}\left(r_{i,t-11,t} >= \theta\right)$, a Jackpot indicator that is 1 if firm *i* has returns greater than θ over the past 12 months, on each growth measure standardized to have zero mean and unit standard deviation in each cross-section. In this regression, the coefficient on the growth measure (γ_t) reflects how representative high growth is for 'Google-type' firms at time *t*. γ_t can be written as:

$$\gamma_t = P(J_t = 1 | \, Grth = 1\sigma) - P(J_t = 1), \qquad (2)$$

where *Grth* is the growth-related variable in excess of its mean. It is easy to see the correspondence between γ in equation 2 and *R* in equation 1. Both measure whether the trait of being a Google type firm is overrepresented in the high growth group relative to a reference group. As in BGLS1, the reference group in equation 2 is all firms without conditioning on information about the trait. The sole distinction between equations 1 and 2 is that equation 1 is expressed as a ratio, whereas equation 2 is expressed as a difference.

Third, our estimate of representativeness (Rep_t) is a 36-month moving average of γ_t. The moving average smooths out the impact of estimation error in monthly estimates. Our choice of a 36-month lag length is derived from BGLS1; they report that belief distortions typically last three years in a calibration of their model.

We perform two sets of tests of the joint hypothesis that the low returns of high growth stocks are related to the representativeness heuristic and that our empirical measure is a reasonable proxy for representativeness. We test whether *Rep* predicts the differential between the returns of high and low

growth firms for each of the four growth measures. We also test if the mechanism underlying this predictive ability is an over-reaction to news, as in the BGLS1 model.

To perform our first set of tests, we run cross-sectional regressions as described above to generate the time-series of γ for each of the four growth measures. These regressions confirm that the 'kernel of truth' property is valid in the data. In particular, the time-series average of γ_t is significantly different from zero for each growth measure, confirming that Google-type firms are indeed more frequent among high growth firms. We also find that for each growth measure, *Rep* predicts returns for the corresponding low minus high growth portfolio. In particular, Rep^{AG} constructed from asset growth predicts the returns of *CMA*, the investment factor from Fama and French (2015), Rep^{MB} from market-to-book predicts *HML*, the value factor from Fama and French (1993), and Rep^{SG} and Rep^{LTG}, variables constructed from sales growth and analyst expected long-term growth, predict their corresponding portfolios (*SG* and *LTG*) constructed in a similar manner as the Fama-French factors.

The four different representativeness measures are correlated with each other, with an average correlation of 61.4%. All four measures typically rise during expansions, peaking at the onset of a recession. This common time-series variation suggests that they may be noisy estimates of the same underlying economic quantity: the representativeness of the Google-type for fast growing firms. Hence, we take a simple average of the four growth-related variables as our measure of the common underlying latent variable. This average, *Rep,* predicts the returns of each of the four growth-related factors. Thus, a single variable predicts expected returns of all four growth-related portfolios. The economic magnitude of the return predictability that we find is substantial. For example, the returns to *CMA* are 0.02% per month (statistically insignificant) when *Rep* is low (below its time-series median) and 0.4% per month when *Rep* is high (above median).

We then test whether the overpricing is corrected steadily or in a crash. Models of diagnostic expectations such as Gennaioli, Shleifer, and Vishny (2015) predict boom and bust cycles in prices, as investors ignore bad news that is not representative, until a string a bad news results in a rapid change in

underlying beliefs, resulting in a crash. We define a crash as a one-year return in the bottom 10th percentile of the distribution of that portfolio's returns and a boom in the top 10th percentile. Because the growth-related anomaly portfolios are short the high growth portfolios, they experience booms when high growth portfolios crash. As hypothesized, we find that *Rep* predicts booms for all four low minus high growth portfolios because it predicts crashes for the high growth leg of the portfolios.

We next test whether *Rep* predicts the four portfolio returns out-of-sample. We find statistically and economically significant out-of-sample predictive ability, with monthly out-of-sample R^2s ranging from 1.21% for *LTG* to 2.76% for *CMA*.

Our paper examines whether there is a link between BGLS1 and Bordalo et al. (2024) (BGLS2). BGLS1 develop a model of analyst learning based on the representativeness heuristic and show that analyst forecasts of expected long-term growth and associated stock returns are consistent with the model's implications. BGLS2 document that the value-weighted analyst long-term growth expectations of S&P 500 firms forecast market returns and provide "promising" evidence that these expectations also predict time-variation in Fama–French factor premia. Their hypothesis is that cash flows expectations rather than discount rate changes predict returns; however, they do not tie their predictability results to specific behavioral model for expectations formation. For example, it is by no means clear that the common variation in value-weighted analyst expectations of the 500 largest US stocks is related to the representativeness heuristic. Intuitively, we would expect the "*next* Google" to lie in small rather than large growth stocks.

Rather than starting from analysts' expectations, we construct a direct measure of representativeness, *Rep,* which captures how representative recent high-growth firms are of high returns *relative* to all stocks. Because *Rep* relies solely on realized returns, it allows us to test whether the common (systematic) variation in analyst long-term growth forecasts identified by *BGLS2* is related to the representativeness bias. Empirically, *Rep* and LTG^{BGLS} are contemporaneously correlated (≈ 40 %), consistent with a shared behavioral foundation. Yet they contain distinct predictive content: we find that

Rep continues to forecast one-month and one-year growth portfolio returns after controlling for LTG^{BGLS}, while LTG^{BGLS} retains power for five-year returns after controlling for *Rep*. These results suggest that although these variables share substantial common variation, the predictive ability in BGLS2 is different from our return-based measure of representativeness.

Finally, we test whether the mechanism through which *Rep* predicts returns is consistent with its motivation, using the Coibion–Gorodnichenko (2015) regression framework adopted by *BGLS1* to test for over-extrapolation. We regress subsequent forecast errors in long-term earnings growth on prior revisions in analysts' growth expectations and find a negative coefficient consistent with over-reaction to news, as reported in BGLS1. Crucially, when we introduce an interaction between *Rep* and revisions, we find that it is significantly negative, showing that when representativeness is high, analysts' over-reaction intensifies. Thus, these results connect past realized return outcomes (*Rep*) to the belief-formation process modeled in *BGLS1*.

The return predictability results survive a barrage of robustness tests. First, we show that the predictive ability of *Rep* is robust to controlling for potentially confounding variables. One important such variable is sentiment because Stambaugh, Yu, and Yuan (2012) show that sentiment predicts the returns to a set of anomalies including asset growth and Cooper, Gulen, and Ion (2024) argue that changes in equity financing costs as measured by sentiment are responsible for the asset growth anomaly. It is likely that representativeness is correlated with sentiment because sentiment is typically high in market booms when speculative high growth stocks are doing well. We find that representativeness is indeed correlated with sentiment and beats sentiment in a 'horse race' predicting returns of the growth anomalies—representativeness remains significant, and sentiment is insignificant in predicting returns of three (CMA, HML, and SG) of the four anomalies. We also show that our predictability results are robust to controlling for other standard business cycle predictors such as the term and default spreads, price-earnings ratio, and consumption-wealth ratio. The predictability is also not due to momentum in factor returns documented in Ehsani and Linnainmaa (2022).

Our paper is related to research that analyzes the representativeness bias. Gennaioli and Shleifer (2010) and Bordalo, Coffman, Gennaioli, and Shleifer (2016) formalize Kahneman and Tversky's (1972) representativeness heuristic into models of diagnostic expectations. Our paper builds on BGLS, who develop, calibrate, and test a model of diagnostic expectations to understand the low returns of stocks with high analyst forecasts of long-term earnings growth. We complement their results in two related directions: different anomalies and different tests. First, besides long-term analyst forecast growth, we also examine asset growth, book-to-market, and sales growth. Asset growth is particularly important because Fama and French (2015) and Hou, Xue, Zhang (2015) show that it explains several other anomaly premia. Second, we construct an empirical measure of the degree of representativeness at a given point in time, which we show predicts future returns of all four growth-related anomalies. Thus, our results are primarily in the time-series of anomaly premia, whereas theirs are in the cross-section. Our paper is also related to Cassella and Gulen (2018), who show that extrapolative expectations can explain the predictability of returns of price-scaled variables.

Fama and French (2015) and Hou, Xue, and Zhang (2015) show that the asset growth factor explains the returns to the value factor. Yet, it is not clear why that is the case, because prior research provides very different explanations for these two anomalies—ranging from distress risk (Fama and French, 1992), displacement risk (Gârleanu, Leonid Kogan and Stavros Panageas, 2012), to duration risk (Lettau and Wachter, 2007) for value, and optimal investment according to *q* theory (Li and Zhang, 2010) to equity financing-related shocks as measured by sentiment for asset growth (Cooper, Gulen, and Ion, 2024). We provide evidence that a single explanation—the representativeness heuristic—suffices to explain all four anomalies.

Our second contribution is to the literature on predicting returns to factors. Although there is a long literature predicting the market risk premium, there are relatively fewer papers that examine the predictability of other factor returns. While some research focuses on predicting specific factor returns, such as Cohen, Polk, and Vuolteenaho (2003) on value and Daniel and Moskowitz (2016) on momentum,

other research focuses on predicting a set of anomalies using a single variable, such as Stambaugh, Yu, and Yuan (2012) who predict anomalies with sentiment. Finally, Haddad, Kozak, and Santosh (2020) predict the principal components of a wide set of anomaly returns using valuation ratios. We test models of diagnostic expectations, by examining whether a variable based on them predicts time-variation in returns to four prominent growth-related portfolios both in and put-of-sample.

Finally, our measure of representativeness is related to the measure of Jackpots in Conrad, Kapadia, and Xing (2014) that is also based on returns greater than a threshold. In practice, the measures are very different: the Conrad et al measure is firm-specific and combines several characteristics while ours is time-specific and is the same for all firms at a given point in time. Also, ours is consistent with the theory in BGLS and measures the realized probability of jackpots for high asset growth stocks *relative* to that of all stocks, while Conrad et al predict the absolute level of jackpot probability for each stock. Finally, Conrad et al rely on a theory based on preferences, while we rely on one based on how investors form expectations. Of course, these two approaches need not be mutually exclusive. While it is difficult to distinguish between preferences and expectations in practice, our results on analyst surprises are more consistent with biased expectations rather than a preference for skewness with unbiased expectations.

II. Data and Methodology

A. Data

Firms in our sample are in the intersection of CRSP and COMPUSTAT. We require that firms are NYSE, Amex, and NASDAQ listed and have ordinary common shares (share code 10 and 11). We obtain daily and monthly stock prices, returns, and shares outstanding from CRSP daily and monthly files. We adjust stock returns with CRSP delisting returns where applicable.

We obtain accounting information from COMPUSTAT quarterly and annual files. We only include domestic firms with standardized and consolidated data from COMPUSTAT. We use

COMPUSTAT data to construct firm characteristics. The characteristics underlying our growth variables are obtained from Compustat annual files. When merging Compustat information with returns, we ensure that the fiscal year end is at least six months before the month of returns, to ensure that the accounting information is known to markets before returns are measured.

We use data from IBES for some of our tests. Following Bordalo, Gennaioli, La Porta, and Shleifer (2019), we take mean analyst EPS long-term growth (LTG) estimates made at December of each year from the IBES Unadjusted Summary Statistics file to measure analyst expected growth in earnings. Actual EPSs are adjusted for stock splits using CRSP daily data on cumulative adjustment factors (CFACSHR). We obtain Fama-French factor returns from Prof. French's website and investor sentiment data from Prof. Wurgler's website.

Table I presents summary statistics and a correlation matrix for selected variables used in this paper. The sample period is from 1963 to 2023. See Appendix A for variable definitions.

B. Methodology to estimate representativeness

We consider four commonly used measures of growth. While we do not know for certain what measures investors use, the four measures we consider are simple, have been used extensively in prior research, and together, cover a lot of ground. Each variable measures a different aspect of growth: balance sheet growth (asset growth; AG), income statement growth (sales growth; SG), investor expected growth (market-to-book; MB), and analyst expected EPS growth over the next five years (LTG).

To estimate our measure of representativeness, we run univariate cross-sectional regressions in the style of Fama and MacBeth (1973) every month of the Jackpot indicator, *J,* on each growth measure.

$$J_t\left(r_{t-11,t} >= 100\%\right) = \mathrm{a} + \gamma_t \mathrm{G}_{i,t-12} \qquad (3)$$

where G is a measure of growth (*AG*, *SG*, *MB*, or *LTG*) for firm i, known as of time t - 12 that is standardized to have a zero mean and unit standard deviation in every cross-section, and $J_{i,t}(r_{t-11,t} >= 100\%)$ is the Jackpot indicator for a Google-type firm—a return greater than 100% over the past 12 months.[2] γ_t is an estimate of representativeness for month t. Note γ_t estimates the probability of Jackpots for high growth firms (those with G = 1 standard deviation above the mean) relative to firms with typical growth (G = mean growth). This is similar to the definition of representativeness in BGLS except that it is a difference rather than a ratio.[3]

We take a moving average of γ_t over the last 36 months to generate our measure of representativeness, Rep_t^G, of the Google-type for growth measure G. Note that Rep_t^G only uses information that is known by investors as of the end of month t and we use it to predict returns of month t+1.

III. Predicting growth portfolio returns with representativeness

A. Estimating representativeness for each growth variable

Table II presents results of regressions that estimate representativeness by running Fama-MacBeth regressions of the Jackpot indicator on each growth measure as specified in equation (2) above. Panel A reports time-series averages of monthly coefficients. Note that because the indicator is based on 12 months of returns, the dependent variable overlaps with 11 prior cross-sections. We use Newey-West standard errors to correct for autocorrelation in errors in Panel A. As an alternative, in Panel B, we use

[2] The choice of a threshold of 100% and the horizon of 1 year are ad-hoc, but they seem reasonable and are the first benchmarks we tried. We find similar results with other benchmarks.

[3] In an earlier draft of this paper, we use sorts on growth instead of regressions and find that the ratio of realized jackpots for firms in the top tercile of asset growth relative to that for all firms also predicts *CMA* returns. We use regressions in the current draft because they are less susceptible to noise that occurs with quintile or decile sorts on growth given the low realized probability of Jackpots in small samples. In addition, the use of a regression on growth removes subjectivity in defining the cut-off for high growth.

annual regressions with Jackpots defined using July to June returns with no overlap. Since both panels provide similar inferences, we focus our discussion on Panel A. The first column shows that the unconditional probability of Jackpots (the intercept in the regression) is 4.5%. The next four specifications show that each of the four growth measures, asset growth, book-to-market, sales growth, and long-term growth expected by analysts, predict jackpots. This confirms the 'kernel of truth' property of models of extrapolative expectations: jackpots are indeed more frequent among high-growth firms relative to the average firm. In terms of magnitude, the table shows that a 1 standard deviation increase in the growth measure increases the unconditional probability of jackpots relative to the intercept by roughly 9.6% for asset growth, 8.8% for book-to-market, 14.8% for sales growth, and by 34.3% for analyst expected growth.

These regressions generate our measure of representativeness for each growth measure. We take a 36-month moving average of the monthly coefficients from these regressions to generate versions of *Rep* for each growth measure. Figure 1 plots these four versions of *Rep* over time. All four series appear correlated, typically peaking at the end of expansions. Panel B of Table III examines the time-series correlation between the four versions of *Rep*. All four versions are highly correlated with each other, with an average pairwise correlation of 61.4%.

B. Representativeness and expected returns

In this section we examine whether representativeness constructed from each growth measure predicts returns of the portfolio that is long low-growth and short high-growth stocks based on that growth measure. For asset growth and book-to-market, we use the corresponding Fama-French factors, *CMA* and *HML*. For sales growth and analyst predicted long-term growth, we construct long-short portfolios in a similar manner to the Fama-French factors. In particular, we perform independent sorts on size and the growth measure. We use the 30th and 70th NYSE breakpoints of the growth measure and the

median NYSE size to form 6 equal-weighted portfolios.[4] The returns to the long-short growth portfolio are the low growth minus high growth returns averaged across the two size groups.

We next use a regression to test whether representativeness constructed from each growth measure predicts returns of the portfolio that is long low-growth and short high-growth stocks based on that growth measure. The results are shown in Panel A of Table III. Each representativeness measure predicts its corresponding return portfolio. Consistent with our motivation, when representativeness of the Google type for high growth stocks increases, high growth stocks are overvalued, leading to low returns in the future. The R^2 are small in magnitude but are respectable (between 1.8% and 2.9%) when seen in relation to prior work on predicting returns at the monthly frequency. As Campbell and Thompson (2008) point out in the context of market return predictability, small R^2s translate into large utility gains for investors using these predictor variables to determine their trading strategy.

C. A single measure of representatives of the 'Google-type' for high growth

A natural interpretation of the high correlation between the four representativeness variables is that they are noisy measures of the same underlying latent variable: how representative is high growth for exceptional future returns? To estimate this underlying variable, we take a simple average of the four representativeness measures to obtain *Rep*.

Table IV examines whether the single variable *Rep* predicts all four growth-related portfolio returns. Panel A shows that this is indeed the case: there is a common thread running through all four growth-portfolio returns that is predictable based on the single variable. The magnitude of predictability is economically meaningful. For *CMA* for example, a 1 standard deviation increase in *Rep* is associated with a 0.34% increase in average returns; this is large compared to the unconditional average returns of

[4] We follow BGLS in using equal weights rather than value weights, since it is the anomalous returns in their paper that we wish to explain. Premia arising from value-weighted portfolios are smaller than equal-weighted premia and statistically insignificant.

CMA of 0.27% per month. That is, the premium to CMA is essentially zero, in times in which *Rep* is one standard deviation below its mean. Similar results obtain for the other portfolios: a one standard deviation shock results in changes in mean returns (0.42% for *HML*, 0.24% for *SG*, and 0.66% for *LTG*) comparable to the unconditional means of the portfolios. Panel B provides a simple metric to gauge the economic significance of the predictability by regressing growth-portfolio returns on a dummy that is one if the corresponding version of *Rep* is above its time-series median. For all hedge portfolios, average returns are essentially zero when Rep is below its time series median – there is no growth anomaly when *Rep* is low. The growth anomaly only exists when *Rep* is above its time series median, with returns between 0.3% to 0.5% per month for the 4 different growth portfolios (although only two of the four dummy variables are statistically significant).

A possible explanation for these results is a conditional CAPM, in which market betas vary with *Rep*. Panel C examines this explanation by including an interaction between the *Rep* dummy and market returns. The interaction is significant across the 4 growth anomalies, with betas becoming more negative in high *Rep* periods. However, this does not make the returns in high *Rep* times less anomalous; the *Rep* dummy is now significant in all specifications, and of equal or larger magnitude than before.

Figures 2 and 3 provide a graphical representation of these results by plotting average 2-year cumulative returns and average 2-year cumulative CAPM alphas, respectively, for the asset growth, market-to-book, sales growth, and long- term growth hedge portfolios following periods of high and low representativeness. High (low) representativeness is defined as a period when the 36-month moving average of average representativeness is above (below) its time-series median. Figure 2 shows that hedge portfolio returns are substantially higher following high representativeness periods when compared to low representativeness periods. The difference between the portfolio returns in high and low representativeness periods typically grows over the first year, after that both sets of returns grow at about the same pace. In Figure 3, CAPM alphas for *j* months ahead are computed from regressions on the

market excess return and are then cumulated to plot on the figure. Cumulative alphas behave in a similar manner as raw returns.

Table V examines whether prices of aggressive growth firms crash or correct slowly following periods of high representativeness. The model in Gennaioli, Shleifer, and Vishny (2015) predicts that prices crash when investors with the representativeness bias finally revise their expectations following a string of bad news. We define a crash (boom) as a return in the bottom (top) 10^{th} of the distribution of annual returns of the portfolio. We expect that *Rep* predicts crashes for the high growth portfolio and not the low growth portfolio; we also do not expect that *Rep* predicts booms for either the low or the high growth portfolio. Thus, when looking at booms and crashes for each of the four growth variables, we expect four out of sixteen coefficients on *Rep* to be significant in regressions of booms and crashes of each high and low growth portfolio on *Rep*.

Table V shows a pattern consistent with our expectations. *Rep* significantly predicts crashes for all four high growth portfolios; does not predict booms for any of the four low growth portfolio; and also does not predict booms for any of the high or low growth portfolios. The magnitude of predictability is economically meaningful: a 1 standard deviation increase in representativeness results in a 5.6% increase in the probability of a crash for the high asset growth portfolio, 6.7% for the high book-to-market, 4.7% for the high sales growth, and 8.9% for the high expected analyst growth portfolio. These numbers are large relative to the unconditional probability of a crash, which is 10% by definition. Because the long-short growth portfolios are short the high growth portfolio, a greater probability of crashes for high growth portfolios translates into a greater probability of booms for the long-short portfolio. *Rep* significantly predicts booms in all four growth-related long-short portfolios.

D. Out-of-sample Tests

In this section we test whether the predictability we find persists out-of-sample. Goyal and Welch (2008) find that many variables that predict market returns in-sample do not do so out-of-sample. They

argue that it is important to examine out-of-sample R^2s for two reasons: to understand whether investors could have used these variables to time the market in real-time, and to diagnose the stability of the in-sample tests. By showing that most of the predictability for market returns stems from the oil shocks of 1973-75, they argue that the in-sample predictability is unstable. Both these reasons motivate us to examine whether the predictive ability of *Rep* for growth-related portfolios is present out-of-sample.

In our out-of-sample tests, we compare the mean squared prediction error from an OLS model that uses *Rep* to predict each portfolio's return (MSE_{Rep}) to that of a naïve model that uses the historical mean (MSE_N). We estimate both models over an estimation sample, use the parameters estimated over this sample to forecast the next month's return, and compute squared prediction errors. Each initial estimation sample is 20 years. We then roll forward this procedure, by expanding the estimation period by 1 month. For *CMA*, *HML*, and *SG*, the initial estimation period is from July 1963 to June 1983, and the forecast evaluation period is July 1983 to December 2023. The *LTG* return series only begins in July 1982, hence the initial estimation sample is July 1982 to June 2002, and the forecast evaluation period is July 2003 to December 2023.

Following Goyal and Welch (2008), Figure 4 presents the cumulative sum of the squared prediction error differences between the naïve model and *Rep* model. As Goyal and Welch (2008) point out, if the predictive ability of a variable is robust, we expect the cumulative sum of squared errors to steadily increase over the sample period and to end up positive at the end of the sample. The figure shows increasing and positive squared error differentials for all four portfolios. Table VI presents out-of-sample R^2s (defined as $1- MSE_{Rep} / MSE_N$), which are economically and statistically significant. We assess statistical significance using the forecast encompassing test of Clark and McCracken (2001), which they show has better small sample properties than other tests. We assess economic significance using the metric proposed by Campbell and Thompson (2008) – the proportional increase in expected returns for a mean-variance investor that observes the forecasting variable, approximately equals the ratio of the R^2 to the squared Sharpe ratio. The smallest out-of-sample R^2 we find is 1.21% for LTG (has the shortest

evaluation period) that translates to a proportional improvement of 51%. For *CMA*, the R^2 over the full sample is 2.76%, which translates to a proportional improvement of 160%.

CMA, *HML*, and *SG* show a clear spike in the squared error differential around the crash following the internet bubble in Figure 4. This is intuitive because we would expect that the representativeness bias would be especially prominent during the period when internet stocks were delivering really high returns, leading to high prices and low subsequent returns. However, to have confidence in the model's ability to endure in the future, its predictive power should not come solely from any one period. Hence, Table VI also reports out-of-sample R^2s excluding the internet period (1997-2003). Although R^2s drop, they remain economically and statistically significant. Table VII provides additional evidence on whether the internet period is solely responsible for the predictive ability of Rep. We split the sample into two equal parts: 1963-1993 and 1994-2023. Table VII shows that coefficients on *Rep* are significant in both subsamples for all four variables. While coefficients are higher in magnitude in the second subsample for asset growth and book-to market, they are of similar magnitude in both subsamples for *LTG* and *SG*.

IV. Confirming the mechanism

BGLS use regressions suggested by Coibion and Gorodnichenko (2015) to test whether analysts over-extrapolate news in a manner consistent with their model of representativeness. They run a regression of analyst forecast errors in long-term earnings growth on revisions in their expectations of long-term growth:

$$(EPS_{t+n}/EPS_t)^{\frac{1}{n}} - LTG_t = \alpha + \beta\,(LTG_t - LTG_{t-\delta})$$

where n=3, 4, or 5 and δ=1, 2, or 3. These regressions test over-reaction. Presumably, analysts revise their forecasts because of the arrival of news. If analysts over-extrapolate as in the BGLS model, they will revise their earnings expectations too far in the direction of the news, leading to forecast errors in the opposite direction.

To confirm that our measure of representativeness results in overreaction as in BGLS, we add an interaction of *Rep* to the forecast revisions. We expect that the tendency to over-extrapolate will increase with *Rep*, and hence we expect a negative and significant interaction term. Table VIII confirms that this is indeed the case. Following BGLS, we examine forecast errors over the 3-year (Panel A), 4-year (Panel B), and 5-year (Panel C) horizons and forecast revisions over the past 1, 2, and 3 years. As in BGLS, we find that analysts over-extrapolate: forecast revisions generally predict future forecast errors as all coefficients on forecast revisions are significant at the four- and five-year horizons. Most important, the degree of extrapolation increases with *Rep*; seven of nine interaction terms are significant. BGLS report that the absolute magnitude of the coefficients on the forecast revision increase as the horizon increases (as we move from left to right in the table). We also find that the absolute magnitude of the coefficients on the interaction term also increases as the horizon increases.

V. Alternative explanations

A. Sentiment, Analyst Optimism, and Representativeness

In this section, we consider two related explanations for the effects of *Rep* on returns: sentiment and analyst optimism. Stambaugh, Yu, and Yuan (2012; SYY) show that stocks in the short legs of anomaly variables are over-valued in times of high sentiment due to short-sale constraints, leading to low returns. This explanation is related to ours because SYY consider asset growth in the set of anomalies they evaluate. Moreover, Cooper, Gulen, and Ion (2024) also argue that changes in the costs of equity financing as measured by sentiment also predict returns to asset growth. We expect that sentiment and representativeness are correlated because it seems plausible that speculative, high growth firms do well in times of high sentiment. Also, BGLSII hypothesize that cash flow expectations rather than discount rate changes explain returns and show that the value-weighted analyst long-term growth expectations predict market returns negatively. They also show that these expectations predict Fama-French factor returns.

It is plausible that representativeness is the underlying explanation for the predictive ability of both sentiment and analyst optimism. The origins and ultimate causes of sentiment are nebulous: Baker and Wurgler (2006) define sentiment as "a propensity to speculate". But where does this propensity come from? A possibility is that this propensity arises because of the representativeness heuristic: investors over-value growth stocks in times in which a greater fraction of growth stocks has really high returns, resulting in a "propensity to speculate". Thus, representativeness may provide a microfoundation for sentiment and explain why high sentiment results in the over-valuation of high growth stocks. Similarly, BGLSI argue that analyst long-term growth expectations reflect the representativeness bias; thus, it is possible that the average analyst forecast reflects this bias. However, it is by no means clear that this is the case, because the BGLSII measure is the value-weighted average of the largest stocks—it is unlikely that these conform to the "next Google" stereotype, since they already contain current "Googles".

We first examine correlations between the three sets of measures in Table IX, Panel A. We include the Baker and Wurgler (2006) sentiment measure (SENT), along with the value-weighted analyst expected long-term growth of S&P500 (LTG^{VW}) from BGLS2. In particular, BGLS use two components of LTG^{VW} separately in their regressions: $L1TG^{VW}$ and $D1.LTG^{VW}$, where L1 denotes the 1-year (12 month) lag and $D1.LTG^{VW}$ is LTG^{VW}-$L1.LTG^{VW}$. It's clear that $L1.LTG^{VW}$, *Rep*, and *Sent* are all correlated with each other, with pairwise correlations of 53% ($L1.LTG^{VW}$, *Rep*), 60% (*Sent*, *Rep*), and 41% (*Sent*, $L1.LTG^{VW}$). $D1.LTG^{VW}$ does not display meaningful correlations with *Sent* or *Rep*.

To ensure that our results are not solely driven by the correlation between representativeness, LTG^{VW}, and sentiment, Panels B and C run a horse race between these variables. Panel B considers sentiment and *Rep*. Representativeness clearly wins the horse race. For specifications with both sentiment and representativeness, sentiment is only marginally significant for one portfolio (LTG), and representativeness retains significance in predicting three of the four portfolio returns.

Panel C examines LTG^{VW}, using the specifications, sample period, and data from BGLS2. [5] We test whether Rep, $L1.LTG^{VW}$, and $D1.LTG^{VW}$ predict 1-month, 1-year, 3-year, and 5-year ahead returns for the four growth portfolios. *Rep* continues to predict returns to all four portfolios both 1 month and 1 year ahead. The two LTG^{VW} variables predict returns at longer horizons—3 to 5 years ahead. Because longer horizons have potential statistical concerns because returns are overlapping, we use Newey-West standard errors, with lag length equal to the horizon. Thus, it appears that although the $L1LTG^{VW}$ is correlated with *Rep*, both predict returns at different horizons. Hodrick (1992) shows that taking moving averages of the independent variable to predict short horizon returns is equivalent to and has better statistical properties than predicting long horizon returns (that are moving averages of short-horizon log returns) with the original independent variable. It is possible that because *Rep* is a 3-year moving average, it is implicitly following the Hodrick (1992) prescription and predicting short-horizon returns. Hence, it is possible that a 3-year moving average of $L1LTG^{VW}$ will also predict short-horizon returns. However, in unreported tests we find no evidence for this.

B. Momentum and other business cycle predictors

A possible explanation for our results is that the representativeness variable reflects past portfolio returns, which Ehsani and Linnainmaa (2022) show are related to future returns for many long-short factor portfolios. The correlation between representativeness and past returns seems possible because if stocks with high growth have high average returns in the past year, it is plausible that they will also have a greater fraction of stocks with returns greater than 100%, leading to a higher value of representativeness. Note that the direction of predictability we find is opposite to that from factor momentum—high representativeness (which we expect to be correlated with high portfolio returns)

[5] We download BGLS2 data from the replication files uploaded by BGLS2 to the Harvard Dataverse. They provide a longer time-series of LTG than used in paper. We use the sample period from 1982-2015 used in BGLS2, for ease of comparison. Using the full available sample (ending in 2022) results in similar results for Rep, but weaker results for the LTG variables.

predicts lower (rather than higher) returns for high growth firms. Nevertheless, we control for prior 12- and 36-month returns of each growth portfolio in Panel A of Table X. We find that these controls do not affect the significance of representativeness in predicting any of the four growth-portfolio returns.

Panel B examines whether the predictive ability of *Rep* for portfolio returns survives controls for the standard business cycle variables that have be shown by prior research to predict market returns. We control for the cyclically adjusted price-earnings ratio (CAPE), the consumption wealth ratio (CAY), the spread between BAA and AAA bonds (DFS), and the term spread (TMS). We find that controlling for these variables does not affect the significance of *Rep.*

VI. Conclusion

We show that the low returns of stocks with high growth are consistent with the representativeness heuristic of Kahneman and Tversky (1972). We find evidence consistent with a key implication of models of diagnostic expectations that formalize this heuristic: over-extrapolation of outcomes that become more likely in a particular group relative to a reference group in light of recent news. We develop an empirical measure that reflects the definition of representativeness in the context of high growth firms for four different measures of growth and average across these four measures to obtain a simple, theory-driven measure of representativeness of the 'Google-type' for high growth. We show that time-series variation in this measure predicts the returns of the four growth-related portfolios including the Fama-French asset growth and value factors, (CMA and HML) and portfolios designed to track sales growth and analyst long-term growth in a manner similar to the Fama-French factors. We also show that our measure of representativeness is a significant variable in modulating the extent of over-reaction to news by analysts, in a manner consistent with the model in BGLS.

One critique of behavioral finance is that there is a new behavioral bias to explain every new empirical finding. In fact, the opposite is true for this paper. We show that a theory based on the representativeness heuristic that is derived from old experimental evidence (Kahneman and Tversky,

1972) explains a relatively recently discovered empirical fact—the low returns of firms with high asset growth.

REFERENCES

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645-1680.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, 2016, Stereotypes, *Quarterly Journal of Economics* 131, 1753-1794.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2018, Diagnostic expectations and credit cycles, *Journal of Finance* 73, 199-227.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic expectations and stock returns, *Journal of Finance* 74, 2839-2874.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. Belief overreaction and stock market puzzles, *Journal of Political Economy* 132, no. 5 (2024): 1450-1484.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns put-of-sample: Can anything beat the historical average? *The Review of Financial Studies* 21, 1509-1531.

Cassella, Stefano, and Huseyin Gulen, 2018, Extrapolation bias and the predictability of stock returns by price-scaled variables, *The Review of Financial Studies* 31, 4345-4397.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2003. The value spread, *Journal of Finance* 58, 609-641.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644-2678.

Conrad, Jennifer, Nishad Kapadia, and Yuhang Xing, 2014, Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics* 151, 455-475.

Cooper, Michael J., Huseyin Gulen, and Mihai Ion, 2024, The use of asset growth in empirical asset pricing models, *Journal of Financial Economics* 122, 103746.

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221-247.

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

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

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

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

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

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

Gennaioli, Nicola, and Andrei Shleifer, 2010, What comes to mind, *Quarterly Journal of Economics* 125, 1399-1433.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Neglected risks: The psychology of financial crises, *American Economic Review* 105, 310-314.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455-1508.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh, 2020, Factor timing, *The Review of Financial Studies* 33, 1980-2018.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650-705.

Kahneman, Daniel, and Amos Tversky, 1972, Subjective probability: A judgment of representativeness, *Cognitive psychology* 3, 430-454.

Lettau, Martin, and Jessica A. Wachter, 2007, Why is long-horizon equity less risky? A duration-based explanation of the value premium, *Journal of Finance* 62, 55-92.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815-849.

Li, Dongmei, and Lu Zhang, 2010, Does q-theory with investment frictions explain anomalies in the cross section of returns? *Journal of Financial Economics* 98, 297-314.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288-302.

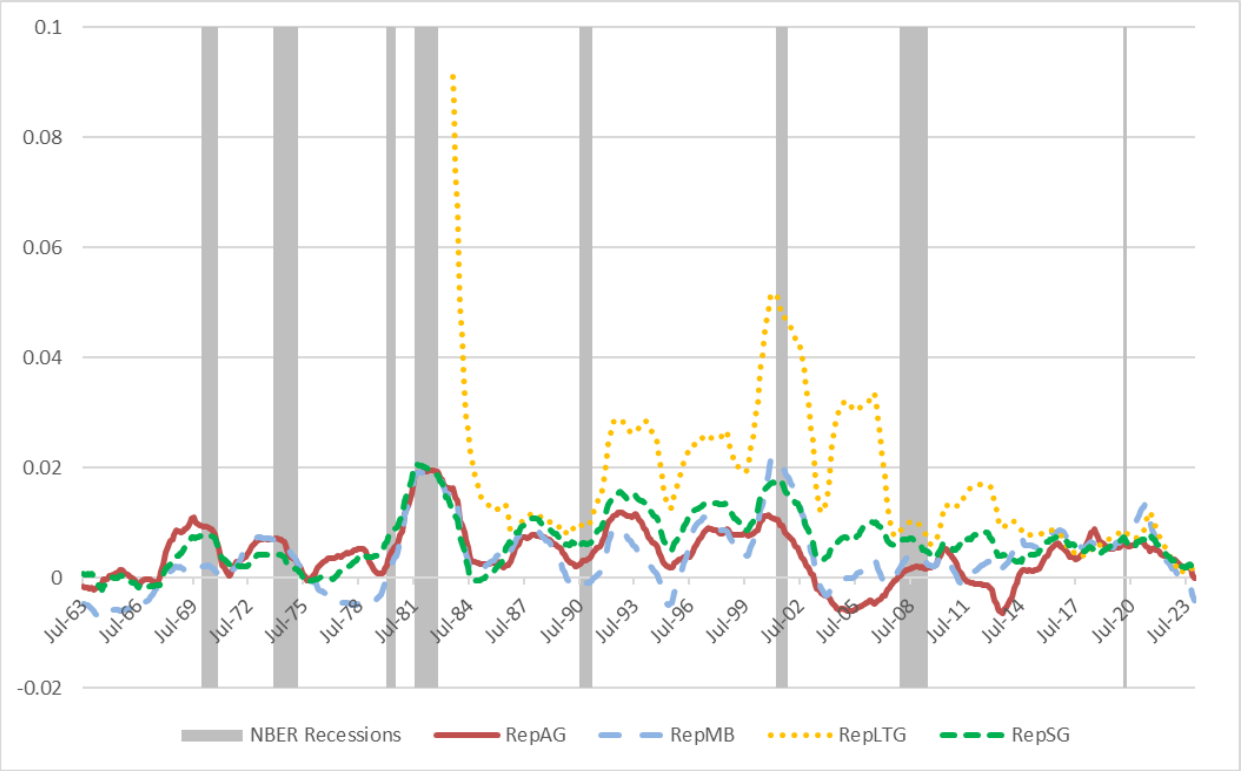


Figure 1. The time-series of representativeness based on alternative measures of growth. This figure plots the time series of representativeness based on asset growth (Rep^{AG}), market-to-book (Rep^{MB}), sales growth (Rep^{SG}), and analyst long-term growth (Rep^{LTG}). NBER recessions are shaded in gray.

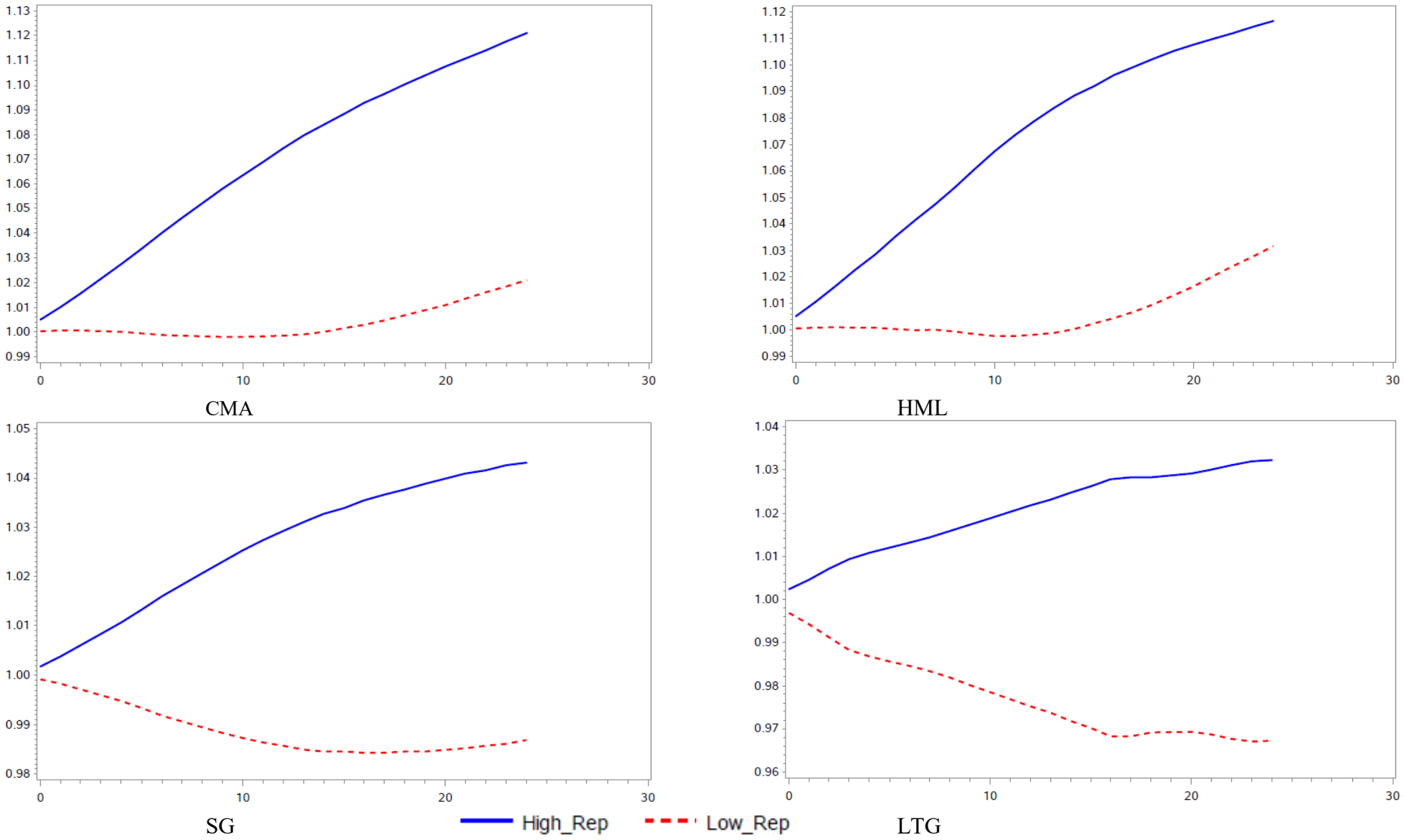


Figure 2. Cumulative raw returns and average representativeness. This figure plots 2-year cumulative average returns for asset growth, market-to-book, sales growth, and analyst long-term growth following high and low representativeness periods. A period is high (low) representativeness when the 36-month moving average of average representativeness is above (below) its time-series median. The blue solid (red dashed) line represents returns following high (low) representativeness periods.

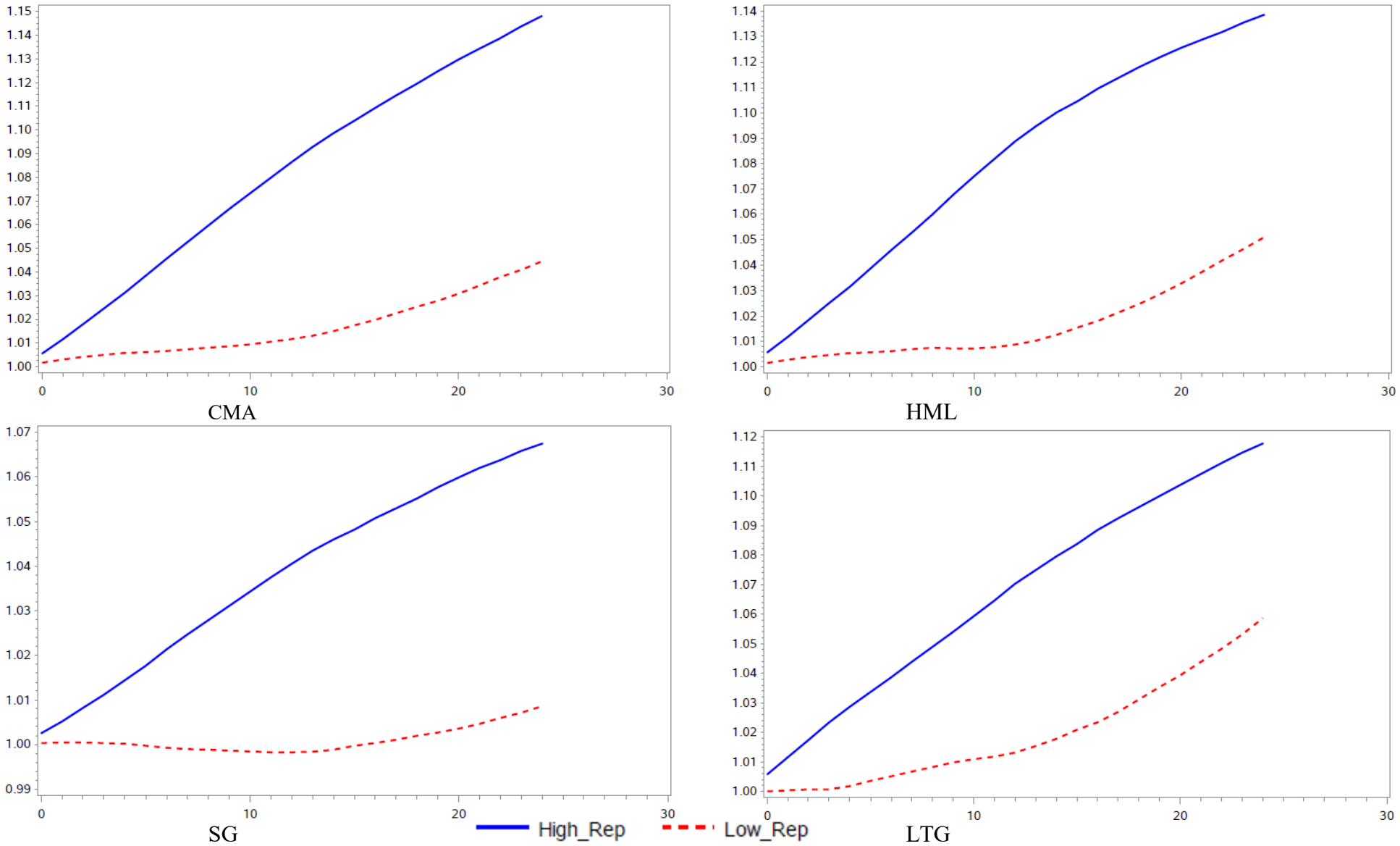
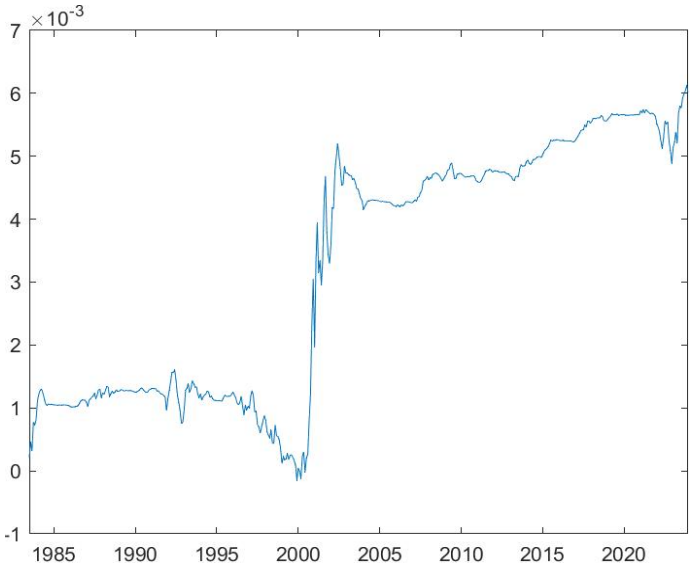
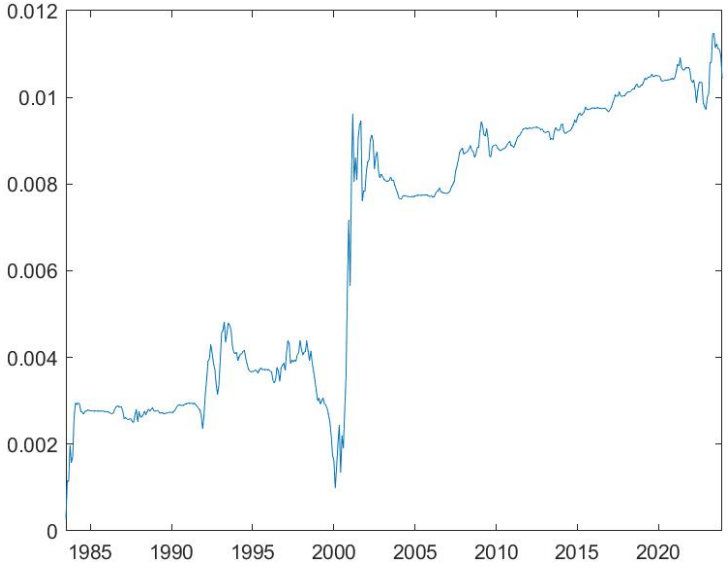


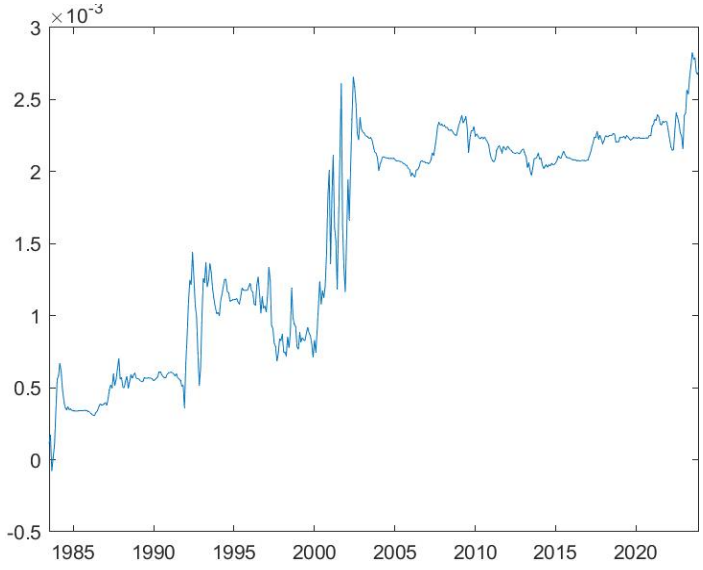
Figure 3. Cumulative abnormal returns and average representativeness. This figure plots 2-year cumulative average CAPM abnormal returns for asset growth, market-to-book, sales growth, and analyst long-term growth following high and low representativeness periods. A period is high (low) representativeness when the 36-month moving average of average representativeness is above (below) its time-series median. The blue solid (red dashed) line represents returns following high (low) representativeness periods.



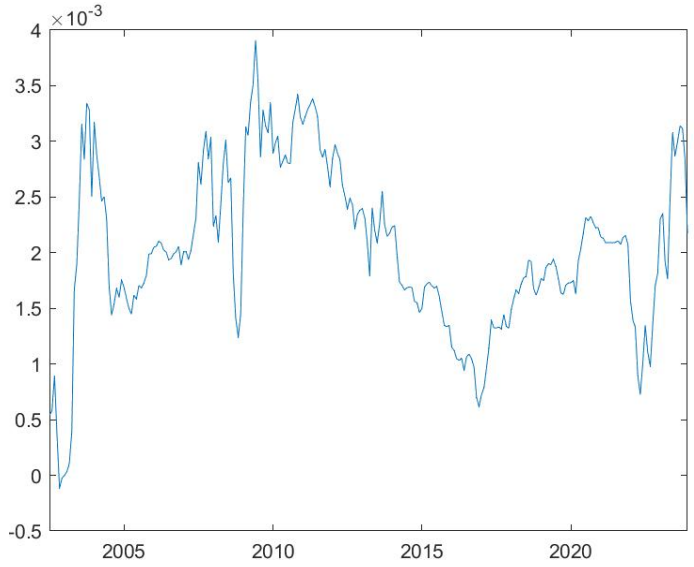
CMA



HML



SG



LTG

Figure 4. Out-of-sample Performance. This figure plots cumulative sum of squared error differences between a forecasting model that uses representativeness to forecast a given portfolio's returns and a naïve model that uses the historical average of that portfolio, for four portfolios--*CMA, HML, SG, and LTG.* A higher number indicates better performance for the model using representativeness. Errors are out-of-sample, and based on an expanding window approach with an initial estimation period of 20 years.

Table I
Summary Statistics

This table presents summary statistics and the correlation matrix for key variables used in this paper. Variable definitions are in Appendix A. All variables are winsorized at 2.5% and 97.5% levels. Panel A reports summary statistics for all variables with sufficient data. Panel B reports Spearman rank correlation in the upper triangle and Pearson correlation in the lower triangle. The sample period is from 1963 to 2023.

Panel A: Summary Statistics

Variables	MEAN	MEDIAN	STD	MIN	MAX
AG	0.184	0.079	0.463	-0.577	5.184
BE	967.894	71.401	6677.823	0.000	595878.000
LTG	0.164	0.145	0.100	-0.284	0.952
MB	2.593	1.550	3.181	0.134	37.369
ME	2073.037	87.518	18780.992	0.000	2902368.097
SG	0.174	0.097	0.409	-0.712	3.361
SIZE	4.724	4.506	2.266	-3.225	14.926

Panel B: Pearson and Spearman Rank Correlations

Variables	AG	BE	LTG	MB	ME	SG	SIZE
AG	1	0.013	0.314	0.242	0.114	0.549	0.106
BE	-0.021	1	-0.424	0.001	0.913	-0.050	0.906
LTG	0.344	-0.112	1	0.365	-0.250	0.406	-0.255
MB	0.166	-0.009	0.301	1	0.370	0.226	0.350
ME	-0.009	0.639	-0.046	0.111	1	0.047	0.983
SG	0.494	-0.026	0.409	0.180	-0.012	1	0.043
SIZE	0.020	0.317	-0.212	0.221	0.278	-0.010	1

Table II
Predicting Jackpots with Measures of Growth

This table presents regressions that use different measures of growth to predict Jackpot returns. Jackpot is a dummy variable that equals 1 if cumulative returns over the next 12 months are greater than 100% and 0 otherwise. Panels A and Panel B present monthly and annual Fama-MacBeth regression results, respectively. Annual regressions start at July of each year, consistent with the date of annual portfolio rebalancing. Growth measures (AG, MB, SG, and LTG) are standardized to have mean zero and standard deviation of 1 for easy interpretation. t-statistics are based on Newey-West standard errors with 12-month lag for monthly regressions in Panel A and no Newey-West adjustments are performed for annual regressions. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance levels respectively.

Panel A: Predicting Jackpot with Measures of Growth at Monthly Frequency					
Intercept	0.046***	0.044***	0.044***	0.044***	0.042***
	(10.19)	(9.81)	(9.84)	(9.73)	(7.76)
AG		0.004***			
		(4.27)			
MB			0.004***		
			(3.16)		
SG				0.007***	
				(6.90)	
LTG					0.016***
					(5.48)
Panel B: Predicting Jackpot with Measures of Growth at Annual Frequency					
Intercept	0.047***	0.045***	0.045***	0.045***	0.043***
	(7.04)	(6.82)	(6.77)	(6.75)	(5.20)
AG		0.004***			
		(3.24)			
MB			0.005***		
			(3.04)		
SG				0.006***	
				(5.21)	
LTG					0.015***
					(4.48)

Table III
Representativeness Constructed from Alternative Measures of Growth

This table presents regressions that use measures of representativeness based on four different measures of growth (AG,MB,SG,LTG) to predict returns for portfolios constructed from sorts on that measure of growth. Rep^{AG} is the 36-month moving averge of coefficients from monthly Fama MacBeth regression of Jackpot returns on AG. Rep^{MB}, Rep^{SG}, and Rep^{LTG} are constructed in a similar manner. Panel A presents regression results using Rep^{AG}, Rep^{MB}, Rep^{SG}, or Rep^{LTG} to predict hedged portfolio returns of CMA, HML, Sales Growth, and Analyst Expected Growth, respectively. Panel B presents the correlation matrix, mean, and standard deviations of different measures of representativeness. Spearman rank correlation is reported in the upper triangle and Pearson correlation is reported in the lower triangle. t-statistics are based on Newey-West standard errors with 2 month lag. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level respectively.

Panel A: Representativeness predicting growth-related portfolio returns

	CMA		HML		Sales Growth		Analyst Growth
	Estimate		Estimate		Estimate		Estimate
Intercept	0.000	Intercept	0.000	Intercept	-0.002	Intercept	-0.009***
	(0.28)		(0.38)		(-1.54)		(-3.09)
Rep^{AG}	0.568***	Rep^{MB}	0.631***	Rep^{SG}	0.393**	Rep^{LTG}	0.516***
	(3.94)		(2.90)		(2.23)		(3.14)
Adj R^2	0.017	Adj R^2	0.015	Adj R^2	0.007	Adj R^2	0.029

Panel B: Correlation, Mean, and Standard Deviation of representativeness

	Rep^{AG}	Rep^{MB}	Rep^{SG}	Rep^{LTG}	Rep	Mean	Stdev
Rep^{AG}	1	0.679	0.583	0.213	0.648	0.004	0.005
Rep^{MB}	0.695	1	0.661	0.173	0.796	0.004	0.006
Rep^{SG}	0.653	0.776	1	0.678	0.910	0.007	0.005
Rep^{LTG}	0.290	0.492	0.675	1	0.748	0.018	0.013
Rep	0.661	0.859	0.903	0.894	1	0.007	0.006

Table IV
Average Representativeness Predicts Growth Portfolio Returns

This table presents regressions that use average representativeness to predict growth portfolio returns. *Rep* is the 36-month moving average of average representativeness, which is obtained by averaging coefficients from univariate Fama MacBeth regressions of jackpot returns on the four growth variables (AG,MB, SG, and LTG). *HighRep* equals 1 if *Rep* is above its sample median and 0 otherwise. Panel A (B) presents regression results using the continuous (dummy) measure of average representativeness to predict the hedged portfolio returns of CMA, HML, Sales Growth, and Analyst Expected Growth respectively. Panel C presents regression results with interactions of representativeness dummy and market return. t-statistics are based on Newey-West standard errors with 2 month lag. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level respectively.

Panel A: Representativeness predicts growth-related portfolio returns

	CMA	HML	Sales Growth	Analyst Growth
Intercept	-0.001	-0.002	-0.002*	-0.009***
	(-0.98)	(-1.10)	(-1.87)	(-2.85)
Rep	0.568***	0.697***	0.397***	1.095***
	(3.77)	(3.11)	(2.84)	(2.63)
Adj R^2	0.026	0.018	0.011	0.024

Panel B: Representativeness dummy predicts growth-related portfolio returns

	CMA	HML	Sales Growth	Analyst Growth
Intercept	0.000	0.001	-0.001	-0.003
	(0.43)	(0.43)	(-0.69)	(-1.36)
HighRep	0.004***	0.005*	0.003	0.005
	(2.69)	(1.84)	(1.54)	(1.62)
Adj R^2	0.010	0.004	0.002	0.003

Panel C: Interactions with market returns

	CMA	HML	Sales Growth	Analyst Growth
Intercept	0.001	0.001	-0.000	-0.001
	(1.16)	(0.42)	(-0.10)	(-0.56)
HighRep	0.005***	0.006**	0.003*	0.008***
	(2.91)	(2.27)	(1.79)	(2.63)
MKT	-0.117***	-0.002	-0.099***	-0.257***
	(-4.07)	(-0.04)	(-3.03)	(-4.65)
HighRep X MKT	-0.093**	-0.253***	-0.118**	-0.288***
	(-2.10)	(-3.16)	(-2.55)	(-3.51)
Adj R^2	0.146	0.079	0.131	0.305

Table V

Predicting Booms and Crashes of One Year Cumulative Portfolio Returns with Representativeness

This table presents regressions that predict booms and crashes in one-year cumulative growth portfolio returns with representativeness. Rep is the 36-month moving average of average representativeness, which is obtained by averaging coefficients from univariate Fama MacBeth regressions of jackpot returns on the four growth variables (AG,MB, SG, and LTG). A boom (crash) is a dummy variable that is 1 if the cumulative future 12-month return is in the top (bottom) 10th percentile of that portfolio's full sample return distribution. t-statistics are based on Newey-West standard errors with 12 month lag. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level respectively.

Panel A: Predicting Booms and Crashes in 1-year Asset Growth Portfolio Returns						
	Aggressive		Conservative		CMA	
	Boom	Crash	Boom	Crash	Boom	Crash
Intercept	0.105***	0.034	0.085***	0.081**	-0.012	0.150***
	(3.07)	(1.01)	(2.59)	(2.07)	(-0.38)	(3.84)
Rep	-0.802	9.228**	2.039	2.528	15.739***	-7.204**
	(-0.21)	(2.12)	(0.47)	(0.69)	(3.21)	(-2.53)
Adj R^2	-0.001	0.033	0.000	0.001	0.099	0.020
Panel B: Predicting Booms and Crashes in 1-year Book to Market Portfolio Returns						
	Low BM		High BM		HML	
	Boom	Crash	Boom	Crash	Boom	Crash
Intercept	0.098***	0.020	0.084**	0.092**	0.009	0.103***
	(3.04)	(0.61)	(2.35)	(2.34)	(0.26)	(3.03)
Rep	0.117	11.202***	2.110	1.041	12.736**	-0.429
	(0.03)	(2.68)	(0.46)	(0.24)	(2.41)	(-0.13)
Adj R^2	-0.001	0.050	0.000	-0.001	0.064	-0.001

Table V - *Continued*

Panel C: Predicting Booms and Crashes in 1-year Sales Growth Portfolio Returns

	High Sales Growth		Low Sales Growth		Low - High Sales Growth	
	Boom	Crash	Boom	Crash	Boom	Crash
Intercept	0.100***	0.044	0.095***	0.064*	0.005	0.103***
	(3.14)	(1.34)	(2.81)	(1.82)	(0.15)	(3.13)
Rep	-0.113	7.807**	0.667	4.964	13.339***	-0.522
	(-0.03)	(1.99)	(0.16)	(1.33)	(2.63)	(-0.17)
Adj R^2	-0.001	0.023	-0.001	0.009	0.071	-0.001

Panel D: Predicting Booms and Crashes in 1-year Long-Term Growth Portfolio Returns

	High LTG		Low LTG		Low - High LTG	
	Boom	Crash	Boom	Crash	Boom	Crash
Intercept	0.056	-0.038	0.059	0.170**	-0.143***	0.082**
	(1.04)	(-0.6)	(1.16)	(2.46)	(-2.86)	(2.03)
Rep	4.621	14.863**	4.353	-7.870	26.403***	1.784
	(0.79)	(2.05)	(0.73)	(-1.61)	(3.77)	(0.43)
Adj R^2	0.005	0.076	0.005	0.020	0.243	-0.001

Table VI
Out-of-Sample Tests

This table presents results of out-of-sample forecasts of four monthly portfolio returns (CMA, HML,SG, LTG) with *Rep* as the forecasting variable. These forecasts are from an expanding window model that uses the first 20 years as the initial estimation window. R^2 is the out-of-sample R^2 from Campbell and Thompson (2008). MSE-ENC is the forecast encompassing test from Clark and McCracken (2001). These statistics are presented for the full sample as well as the sample that excludes the 'Internet' period from 1997-2003.

	Full Sample			Excluding 1997-2003		
Portfolio	R^2	MSE-ENC	p-value	R^2	MSE-ENC	p-value
CMA	2.8%	11.38	<0.01	1.7%	6.66	<0.01
HML	2.1%	8.26	<0.01	1.8%	5.40	<0.01
SG	1.5%	7.32	<0.01	1.1%	4.53	<0.01
LTG	1.2%	6.40	<0.01	1.5%	5.53	<0.01

Table VII

Subsample Tests of Representativness and Growth Portfolio Returns

This table presents results of regressions of monthly growth portfolio returns on average representativeness over two equal sub-samples. Rep is the 36-month moving average of average representativeness, which is obtained by averaging coefficients obtained from univariate Fama MacBeth regressions of jackpot returns on the four growth variables (AG,MB, SG, and LTG). t-statistics are based on Newey-West standard errors with 2 month lag. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level respectively.

Panel A: Equal Subsample Tests of Monthly CMA Return and Rep		
	1963 - 1993	1993 - 2023
Intercept	0.001	-0.005*
	(0.6)	(-1.88)
Rep	0.453***	0.860***
	(3.41)	(2.63)
Adj R^2	0.021	0.038
Panel B: Equal Subsample Tests of Monthly HML Return and Rep		
	1963 - 1993	1993 - 2023
Intercept	0.001	-0.008**
	(0.69)	(-2.03)
Rep	0.611***	1.087**
	(2.79)	(2.43)
Adj R^2	0.020	0.026
Panel C: Equal Subsample Tests of Monthly SG Return and Rep		
	1963 - 1993	1993 - 2023
Intercept	-0.001	-0.004
	(-1.15)	(-1.44)
Rep	0.397**	0.472*
	(2.53)	(1.68)
Adj R^2	0.015	0.009
Panel D: Equal Subsample Tests of Monthly LTG Return and Rep		
	1982 - 2003	2003 - 2023
Intercept	-0.014**	-0.010**
	(-2.31)	(-2.07)
Rep	1.337**	1.507*
	(2.42)	(1.88)
Adj R^2	0.027	0.010

Table VIII

Representativeness and Forecast Errors in Analyst Long-Term Growth

This table presents regressions that analyst long-term growth revisions to predict the forecast error in long-term growth and its interaction with representativeness. Forecast error (FEnY) is calculated as (EPSt+n/EPSt)^1/n – LTGt for n = 3, 4, and 5. The sample period is from 1983 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level respectively.

	Panel A: FE3Y					
	Estimate	t	Estimate	t	Estimate	t
Intercept	-0.017***	(-4.44)	-0.016***	(-3.92)	-0.016***	(-3.83)
$LTG_t - LTG_{t-1}$	-0.127	(-1.31)				
$LTG_t - LTG_{t-2}$			-0.175*	(-1.83)		
$LTG_t - LTG_{t-3}$					-0.126	(-1.44)
$LTG_t - LTG_{t-1}$ X Rep	-14.140	(-1.54)				
$LTG_t - LTG_{t-2}$ X Rep			-24.204**	(-2.55)		
$LTG_t - LTG_{t-3}$ X Rep					-36.986***	(-4.15)
Rep	-0.230	(-0.65)	0.001	-0.028	0.219	(0.53)
Adj R^2	0.001		0.003		0.005	
	Panel B: FE4Y					
	Estimate	t	Estimate	t	Estimate	t
Intercept	-0.039***	(-14.09)	0.003	-0.039***	-0.038***	(-12.50)
$LTG_t - LTG_{t-1}$	-0.285***	(-4.43)				
$LTG_t - LTG_{t-2}$			-0.256***	(-4.26)		
$LTG_t - LTG_{t-3}$					-0.263***	(-4.58)
$LTG_t - LTG_{t-1}$ X Rep	-4.341	(-0.65)				
$LTG_t - LTG_{t-2}$ X Rep			-16.811***	(-2.60)		
$LTG_t - LTG_{t-3}$ X Rep					-21.843***	(-3.64)
Rep	-0.254	(-0.99)	-0.017	(-0.06)	0.084	(0.28)
Adj R^2	0.003		0.007		0.010	
	Panel C: FE5Y					
	Estimate	t	Estimate	t	Estimate	t
Intercept	-0.055***	(-24.07)	-0.055***	(-23.27)	-0.054***	(-21.61)
$LTG_t - LTG_{t-1}$	-0.267***	(-5.09)				
$LTG_t - LTG_{t-2}$			-0.287***	(-6.00)		
$LTG_t - LTG_{t-3}$					-0.291***	(-6.36)
$LTG_t - LTG_{t-1}$ X Rep	-8.478	(-1.53)				
$LTG_t - LTG_{t-2}$ X Rep			-12.673**	(-2.41)		
$LTG_t - LTG_{t-3}$ X Rep					-16.416***	(-3.34)
Rep	-0.017	(-0.08)	0.261	(1.15)	0.238	(0.96)
Adj R^2	0.005		0.010		0.014	

Table IX

Representativeness, Sentiment, and Analyst Expected Growth

This table examines the relationship between representativeness, sentiment, analyst expected growth and portfolio returns. Panel A presents the correlation between representativeness, sentiment, and analyst expected growth. Sentiment (Sent) is from Baker and Wurgler (2006) and LTG^{VW} is the value-weighted analyst expected long-term growth of the S&P 500 from BGLS2. *Rep* is the 36-month moving average of average representativeness, which is obtained by averaging coefficients from univariate Fama MacBeth regressions of jackpot returns on the four growth variables (AG,MB, SG, and LTG). Panel B presents the relationship between one-month growth portfolio returns, sentiment, and the representativeness. The sample period is 7/1963-12/2023 for CMA, HML, and SG and 7/1982-12/2023 for LTG^{VW}. Panel C presents regressions predicting, 1-, 12-,36-, and 60-month ahead returns. $L1LTG^{VW}$ lags LTG^{VW} by 12 months and $D1LTG^{VW}$ is LTG^{VW} - $L1LTG^{VW}$. The sample period for Panel C is 1982-2015. *, **, and *** represent 10%, 5%, and 1% significance level

Panel A: Pearson and Spearman Rank Correlation of Sentiment and Representativeness

	LTG^{VW}	$L1LTG^{VW}$	$D1LTG^{VW}$	Rep	Sent
LTG^{VW}	1.00	0.46	0.66	0.40	0.42
$L1LTG^{VW}$	0.46	1.00	-0.36	0.53	0.41
$D1LTG^{VW}$	0.66	-0.36	1.00	-0.05	0.12
Rep	0.40	0.53	-0.05	1.00	0.60
Sent	0.42	0.41	0.12	0.60	1.00

Panel B: Representativeness and Sentiment

	CMA		HML		SG		LTG	
Intercept	0.003***	-0.001	0.003**	-0.003	0.001	-0.002	-0.002	-0.007**
	(3.3)	(-0.76)	(2.29)	(-1.23)	(0.74)	(-1.22)	(-1.1)	(-2.12)
Sent	0.002**	0.000	0.003*	0.000	0.002	0.000	0.009***	0.006*
	(2.19)	(0.28)	(1.95)	(0.04)	(1.55)	(0.19)	(2.82)	(1.92)
Rep		0.572**		0.786***		0.375*		0.691
		(2.54)		(2.62)		(1.85)		(1.56)
Adj R^2	0.011	0.027	0.007	0.021	0.004	0.009	0.026	0.031

Panel C: Representativeness and Analyst Expected Growth								
	HML				**CMA**			
	1 month	1 year	3 years	5 years	1 month	1 year	3 years	5 years
Constant	0.013	-0.088	-0.954***	-1.170***	-0.003	-0.131	-0.548**	-0.549***
	(0.65)	(-0.38)	(-3.83)	(-4.25)	(-0.27)	(-1.24)	(-2.49)	(-2.76)
Rep	1.446***	9.139***	-3.739	12.638**	0.607***	4.277**	-0.879	4.733
	(4.14)	(3.03)	(-0.53)	(2.53)	(3.01)	(2.39)	(-0.28)	(1.19)
L1LTG	-0.189	0.266	8.502***	8.976***	0.009	1.066	5.522***	5.544***
	(-1.06)	(0.13)	(3.95)	(4.51)	(0.09)	(1.09)	(3.13)	(3.76)
D1LTG	-0.063	0.006	7.330**	10.404***	0.035	0.230	5.951***	8.522***
	(-0.26)	(0.00)	(2.33)	(4.60)	(0.26)	(0.16)	(2.61)	(3.10)
MKTRF	-0.140***	-0.022	0.120	0.152*	-0.161***	-0.096*	-0.121	-0.084
	(-3.37)	(-0.20)	(0.87)	(1.66)	(-6.50)	(-1.76)	(-1.12)	(-0.72)
Adj R^2	0.097	0.193	0.352	0.582	0.176	0.328	0.555	0.534

	LTG				**SALEG**			
	1 month	1 year	3 years	5 years	1 month	1 year	3 years	5 years
Constant	0.032	0.147	-0.540**	-0.733***	0.011	0.027	-0.145	-0.295**
	(1.17)	(0.55)	(-2.15)	(-3.19)	(0.84)	(0.25)	(-0.62)	(-2.35)
Rep	1.356***	9.171**	-2.511	10.547**	0.550**	3.993**	-1.394	-1.684
	(3.29)	(2.21)	(-0.32)	(1.99)	(2.22)	(2.24)	(-0.43)	(-0.43)
L1LTG	-0.338	-1.675	4.838**	5.207**	-0.115	-0.416	1.720	2.957***
	(-1.36)	(-0.67)	(2.16)	(2.55)	(-0.98)	(-0.43)	(0.91)	(2.64)
D1LTG	-0.210	-2.316	2.464	3.909	-0.018	-0.785	0.951	3.839
	(-0.64)	(-0.71)	(0.56)	(1.14)	(-0.12)	(-0.61)	(0.36)	(1.39)
MKTRF	-0.466***	-0.352***	-0.121	-0.051	-0.146***	-0.065	-0.140	-0.035
	(-9.99)	(-2.69)	(-0.91)	(-0.49)	(-5.18)	(-0.96)	(-1.27)	(-0.33)
Adj R^2	0.306	0.254	0.184	0.453	0.104	0.102	0.175	0.162

Table X

Representativeness, Factor Momentum, and Additional Economic Factors

This table examines if the relation between representativeness and future growth portfolio returns survives controls for factor momentum and commonly used predictors of market returns (CAPE, CAY, DFS, and TMS). Panel A presents the regressions that use average representativeness to predict growth portfolio returns while controlling for past 12-month (Ret_12) and 36-month (Ret_36) factor momentum. Rep is the 36-month moving average of average representativeness. Panel B presents regressions that use average representativeness to predict growth portfolio returns, while controlling for additional economic factors. CAPE is the cyclically adjusted price-to-earnings ratio available on Robert Shiller's website. CAY is the consumption, wealth, income ratio from Lettau and Ludvigson (2001). DFS is the default yield spread calculated as the difference between BAA and AAA rated bond yields. TMS is the term spread calculated as the difference between long-term yield government bonds and Treasury bill. t-statistics are based on Newey-West adjusted standard errors with a 2-month lag. The sample period is from 1963 to 2023. *, **, and *** represent 10%, 5%, and 1% significance level

Panel A: Representativeness and Factor Momentum								
	CMA		HML		Sales Growth		Analyst Growth	
	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	-0.002	(-1.03)	-0.002	(-1.1)	-0.002*	(-1.68)	-0.008**	(-2.42)
Rep	0.547***	(3.26)	0.678***	(2.79)	0.406***	(2.69)	0.976**	(2.22)
Ret_12	0.011	(0.71)	0.013	(1.03)	-0.004	(-0.28)	0.001	(0.02)
Ret_36	-0.001	(-0.12)	-0.002	(-0.33)	-0.003	(-0.38)	-0.013	(-1.39)
Adj R^2	0.026		0.021		0.009		0.024	
Panel B: Representativeness and Additional Economic Factors								
	CMA		HML		Sales Growth		Analyst Growth	
	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	0.011	(1.36)	0.018	(1.33)	0.014	(1.61)	0.035	(1.61)
Rep	0.802***	(3.81)	1.048***	(3.4)	0.662***	(3.1)	1.552***	(3.11)
CAPE	-0.000	(-1.5)	-0.000	(-1.32)	-0.000*	(-1.87)	-0.001*	(-1.73)
CAY	-0.028	(-1.17)	-0.049	(-1.1)	0.008	(0.27)	0.017	(0.26)
DFS	-0.410*	(-1.84)	-1.145***	(-2.69)	-0.338	(-1.3)	-1.287**	(-2.06)
TMS	-0.042	(-0.75)	-0.003	(-0.03)	-0.091	(-1.39)	-0.216	(-1.61)
Adj R^2	0.030		0.035		0.014		0.035	

Appendix A. Variable Definitions

Variable	Explanation
AG	Total Assets$_t$/Total Assets$_{t-1}$ -1
BE	Book value of equity (BE) is calculated as shareholders' equity (depending on data availability SEQ, CEQ+PS or AT-LT) plus deferred taxes (TXDITC) minus preferred stock (PS: depending on data availability can be PSTKRV,PSTKL, or
LTG	Analyst long-term growth estimate (LTG) is average percentage growth rate for earnings that analysts made at December of last year.
MB	Market to book ratio (MB) is calculated as market value of equity at December of last year divided by book value of equity (BE).
ME	Market value of equity at December of last year.
Sent	Investor sentiment from Baker and Wurgler (2006).
SG	Sales$_t$/Sales$_{t-1}$ -1
Size	Natural log of firm market capitalization at the end of June of year t.