Technical Document: Factors

Kewei Hou*
Ohio State and CAFR

Chen Xue†
University of Cincinnati

Lu Zhang‡
Ohio State and NBER

July 2026§

Abstract

This document details the construction of q-factors and the expected growth factor in the q^5 model posted at `global-q.org` based on Hou, Xue, and Zhang (2015, *Review of Financial Studies*) and Hou, Mo, Xue, and Zhang (2019, 2021, *Review of Finance*).

*Fisher College of Business, The Ohio State University, 820 Fisher Hall, 2100 Neil Avenue, Columbus OH 43210; and China Academy of Financial Research (CAFR). Tel: (614) 292-0552. E-mail: hou.28@osu.edu.

†Lindner College of Business, University of Cincinnati, 2338 Lindner Hall, 2906 Woodside Drive, Cincinnati, OH 45221. Tel: (513) 556-7078. E-mail: xuecx@ucmail.uc.edu.

‡Fisher College of Business, The Ohio State University, 760A Fisher Hall, 2100 Neil Avenue, Columbus OH 43210; and National Bureau of Economic Research (NBER). Tel: (614) 292-8644. E-mail: zhanglu@fisher.osu.edu.

§We thank Taras Zlupko from Center for Research in Security Prices (CRSP) for helpful comments.

1 Overview

We provide factors data for q-factors and the expected growth factor from the q^5 model as well as their underlying benchmark portfolios (Hou, Xue, and Zhang 2015; Hou, Mo, Xue, and Zhang 2019, 2021). This section offers a brief overview, Section 2 details the construction of the q-factors, and Section 3 describes the expected growth factor.

Stock Sample Monthly returns, dividends, and prices are from the Center for Research in Security Prices (CRSP) and accounting information from the Compustat Annual and Quarterly Fundamental Files. Our sample includes all domestic common stocks listed on NYSE, Amex, and Nasdaq.[1] We exclude financial firms (SIC between 6000 and 6999) and firms with negative book equity. Stock returns are adjusted for delisting, as described in Section 3 of "Technical Document: Testing Portfolios." The sample period is from January 1967 to December 2025.

Return Frequencies We provide portfolio returns in various frequencies, including daily, weekly (calendar, Friday close to Friday close), weekly (Wednesday-to-Wednesday, Wednesday close to Wednesday close), monthly, quarterly, and annual. We compute monthly portfolio returns using the end-of-prior-month market equity as weights, and we compute daily portfolio returns using the end-of-prior-day market equity as weights. We then compound monthly portfolio returns into quarterly and annual, and we compound daily portfolio returns into weekly.

Data Format The data files are in CSV format and the returns are in percent. We also provide the number of unique stocks for monthly and daily portfolios. We record missing portfolio-period observations as empty values and the corresponding number of stocks as zero.

Update Schedule The factors data will be updated annually in each February.

[1] In the new CRSP CIZ data, domestic common stocks are identified by the following flags: securitytype='EQTY', securitysubtype='COM', sharetype='NS', usincflg='Y', and issuertype in ('ACOR', 'CORP'). Exchange listings are identified by conditionaltype='RW' and tradingstatusflg='A', together with primaryexch='N' for NYSE, primaryexch='A' for Amex (now NYSE American), and primaryexch='Q' for Nasdaq.

2 The q-factors

We construct the market factor as the value-weighted portfolio return of all domestic common stocks listed on NYSE, Amex, and Nasdaq, minus the 1-month Treasury bill rate. For the market factor, we do not exclude financial firms or firms with negative book equity.

Following Hou, Xue, and Zhang (2015), we measure investment-to-assets, I/A, as the annual change in total assets (Compustat annual item AT) divided by one-year-lagged total assets. We measure profitability as Roe, which is income before extraordinary items (Compustat quarterly item IBQ) divided by one-quarter-lagged book equity.[2]

We construct the q-factors from a triple 2-by-3-by-3 sort on size, I/A, and Roe. In particular, sorting on investment and Roe jointly helps orthogonalize the two factors. Specifically, at the end of June of each year t, we use the median NYSE market equity (stock price per share times shares outstanding from CRSP) to split NYSE, Amex, and NASDAQ stocks into two groups, small and big.[3] Independently, at the end of June of year t, we break NYSE, Amex, and NASDAQ stocks into three I/A groups using the NYSE breakpoints for the low 30%, middle 40%, and high 30% of the ranked values of I/A for the fiscal year ending in calendar year $t-1$.

Also, independently, at the beginning of each month, we sort all stocks into three groups based on the NYSE breakpoints for the low 30%, middle 40%, and high 30% of the ranked values of Roe. Earnings data in Compustat quarterly files are used in the months immediately after the most recent public quarterly earnings announcement dates (Compustat quarterly item RDQ). For example, if the earnings for the fourth fiscal quarter of year $t-1$ are publicly announced on March 5 (or March 25) of year t, we use the announced earnings (divided by the book equity from the

[2]Our measure of the book equity is the quarterly version of the annual book equity measure in Davis, Fama, and French (2000). In particular, book equity is shareholders' equity, plus balance sheet deferred taxes and investment tax credit (item TXDITCQ) if available, minus the book value of preferred stock. Depending on availability, we use stockholders' equity (item SEQQ), or common equity (item CEQQ) plus the carrying value of preferred stock (item PSTKQ), or total assets (item ATQ) minus total liabilities (item LTQ) in that order as shareholders' equity. We use redemption value (item PSTKRQ) if available, or carrying value for the book value of preferred stock.

[3]In the CRSP CIZ data, the monthly price (item MTHPRC) is not always observed on the last trading day of the month. In those cases, we set the monthly price to missing and exclude those observations from the calculation of month-end market equity.

third quarter of year $t-1$) to form portfolios at the beginning of April of year t. In addition, for a firm to enter the factor construction, we require the end of the fiscal quarter that corresponds to its most recently announced quarterly earnings to be within six months prior to the portfolio formation. We impose this restriction to exclude stale earnings information.

Taking the intersections of the two size, three I/A, and three Roe groups, we form 18 portfolios. Monthly value-weighted portfolio returns are calculated for the current month, and the portfolios are rebalanced monthly. The Roe portfolios are rebalanced monthly at the beginning of each month, and the size and I/A portfolios are rebalanced annually at the end of each June. The size factor, R_{Me}, is the difference (small-minus-big), each month, between the simple average of the returns on the nine small size portfolios and the simple average of the returns on the nine big size portfolios. Designed to mimic the common variation in returns related to I/A, the investment factor, $R_{\text{I/A}}$, is the difference (low-minus-high), each month, between the simple average of the returns on the six low I/A portfolios and the simple average of the returns on the six high I/A portfolios. Finally, designed to mimic the common variation in returns related to Roe, the Roe factor, R_{Roe}, is the difference (high-minus-low), each month, between the simple average of the returns on the six high Roe portfolios and the simple average of the returns on the six low Roe portfolios.

Formally, let $R_{ijk,t}$, with $i = 1, 2$ and $j, k = 1, 2, 3$, be the month-t return of the value-weighted portfolio that contains all the firms in the i_{th} size group, in the j_{th} I/A group, and in the k_{th} ROE group. For instance, $R_{132,t}$ is the portfolio containing all the firms that reside simultaneously in the small-size portfolio, the high-I/A portfolio, and the median-Roe portfolio. We calculate the

month-t returns of the size, the investment, and the Roe factors as, respectively:

$$R_{\text{Me},t} \equiv \left(\sum_{j=1}^{3}\sum_{k=1}^{3} R_{1jk,t} - \sum_{j=1}^{3}\sum_{k=1}^{3} R_{2jk,t}\right)/9, \tag{1}$$

$$R_{\text{I/A},t} \equiv \left(\sum_{i=1}^{2}\sum_{k=1}^{3} R_{i1k,t} - \sum_{i=1}^{2}\sum_{k=1}^{3} R_{i3k,t}\right)/6, \tag{2}$$

$$R_{\text{ROE},t} \equiv \left(\sum_{i=1}^{2}\sum_{j=1}^{3} R_{ij3,t} - \sum_{i=1}^{2}\sum_{j=1}^{3} R_{ij1,t}\right)/6. \tag{3}$$

Hou, Xue, and Zhang (2015) start their sample in January 1972, limited by earnings announcement dates and book equity in Compustat quarterly files. We follow their procedure from January 1972 onward but extend the sample backward to January 1967 following Hou et al. (2019). To overcome the lack of quarterly earnings announcement dates, we use the most recent quarterly earnings from the fiscal quarter ending at least four months prior to the portfolio formation month. To maximize the coverage for quarterly book equity, whenever available we first use quarterly book equity from Compustat quarterly files. We then supplement the coverage for fiscal quarter four with book equity from Compustat annual files.[4] If both approaches are unavailable, we apply the clean surplus relation to impute the book equity. If available, we backward impute beginning-of-quarter book equity as end-of-quarter book equity minus quarterly earnings plus quarterly dividends.[5] Because we impose a four-month lag between earnings and the holding period (and the book equity in the denominator of Roe is one-quarter-lagged relative to earnings), all the Compustat data in the backward imputation are at least four-month lagged relative to the portfolio formation month.

[4]We measure annual book equity per Davis, Fama, and French (2000) as stockholders' book equity, plus balance sheet deferred taxes and investment tax credit (Compustat annual item TXDITC) if available, minus the book value of preferred stock. Stockholders' equity is the value reported by Compustat (item SEQ), if available. Otherwise, we use the book value of common equity (item CEQ) plus the par value of preferred stock (item PSTK), or the book value of assets (item AT) minus total liabilities (item LT). Depending on availability, we use redemption value (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for the book value of preferred stock.

[5]Quarterly earnings are income before extraordinary items (Compustat quarterly item IBQ). Quarterly dividends are zero if dividends per share (item DVPSXQ) are zero. Otherwise, total dividends are dividends per share times beginning-of-quarter shares outstanding adjusted for stock splits during the quarter. Shares outstanding are from Compustat (quarterly item CSHOQ supplemented with annual item CSHO for fiscal quarter four) or CRSP (item SHROUT), and the share adjustment factor is from Compustat (quarterly item AJEXQ supplemented with annual item AJEX for fiscal quarter four) or CRSP (item MTHCFACSHR).

If data are unavailable for the backward imputation, we impute the book equity for quarter t forward based on book equity from prior quarters. Let $\text{BEQ}_{t-j}, 1 \leq j \leq 4$, denote the latest available quarterly book equity as of quarter t, and $\text{IBQ}_{t-j+1,t}$ and $\text{DVQ}_{t-j+1,t}$ be the sum of quarterly earnings and the sum of quarterly dividends from quarter $t-j+1$ to t, respectively. BEQ_t can be imputed as $\text{BEQ}_{t-j} + \text{IBQ}_{t-j+1,t} - \text{DVQ}_{t-j+1,t}$. We do not use prior book equity from more than four quarters ago ($1 \leq j \leq 4$) to reduce imputation errors.

3 The Expected Growth Factors

Following Hou et al. (2021), we estimate monthly cross-sectional forecasting regressions of one-year-ahead investment-to-assets change, $\text{d}^1\text{I/A}$, on the natural log of Tobin'q, $\ln(q)$, operating cash flows, Cop, and the change in return on equity, dRoe. At the beginning of each month t, we measure current investment-to-assets as total assets (Compustat annual item AT) from the most recent fiscal year ending at least four months ago minus the total assets from one year prior, scaled by the 1-year-prior total assets. The one-year ahead investment-to-assets change, $\text{d}^1\text{I/A}$, is the investment-to-assets from the first year after the most recent fiscal year end minus the current investment-to-assets.

At the beginning of each month t, Tobin's q is the market equity (from CRSP) plus long-term debt (Compustat annual item DLTT) and short-term debt (item DLC) scaled by book assets (item AT), all from the most recent fiscal year ending at least four months ago. For firms with multiple share classes, we merge the market equity for all classes. Following Ball, Gerakos, Linnainmaa, and Nikolaev (2016), we measure operating cash flows, Cop, as total revenue (Compustat annual item REVT) minus cost of goods sold (item COGS), minus selling, general, and administrative expenses (item XSGA), plus research and development expenditures (item XRD, zero if missing), minus change in accounts receivable (item RECT), minus change in inventory (item INVT), minus change in prepaid expenses (item XPP), plus change in deferred revenue (item DRC plus item DRLT), plus change in trade accounts payable (item AP), and plus change in accrued expenses (item XACC), scaled by book assets, all from the fiscal year ending at least four months ago.

Missing annual changes are set to zero. The change in return on equity, dRoe, is Roe minus the four-quarter-lagged Roe. Roe is income before extraordinary items (Compustat quarterly item IBQ) scaled by the 1-quarter-lagged book equity. We compute dRoe with earnings from the most recent announcement dates (item RDQ), and if not available, from the fiscal quarter ending at least four months ago. We winsorize all variables at the 1st and 99th percentiles of their distributions each month. Finally, missing dRoe values are set to zero in the cross-sectional forecasting regressions.

At the beginning of each month t, we construct expected one-year-ahead investment-to-assets changes, denoted $E_t[\mathrm{d}^1\mathrm{I/A}]$, by combining most recent winsorized predictors with the average slopes estimated from the prior 120-month rolling window (30 months minimum). The most recent predictors, $\ln(q)$ and Cop, are from the most recent fiscal year ending at least four months ago as of month t. dRoe is computed using the latest announced quarterly earnings, and if not available, the earnings from the most recent fiscal quarter ending at least four months ago. To avoid look-ahead bias, the average slopes in calculating $E_t[\mathrm{d}^1\mathrm{I/A}]$ are estimated from the prior rolling window regressions, in which $\mathrm{d}^1\mathrm{I/A}$ is from the most recent fiscal year ending at least four months ago as of month t, and the regressors are further lagged by 12 months.

At the beginning of each month t, we use the beginning-of-month median NYSE market equity to split stocks into two groups, small and big. Independently, we split all stocks into three groups, low, medium, and high, based on the NYSE breakpoints for the low 30%, middle 40%, and high 30% of the ranked $E_t[\mathrm{d}^1\mathrm{I/A}]$ values. Taking the intersection of the two size and three $E_t[\mathrm{d}^1\mathrm{I/A}]$ groups, we form six benchmark portfolios. Monthly value-weighted portfolio returns are calculated for the current month t, and the portfolios are rebalanced at the beginning of month $t+1$. Designed to mimic the common variation related to $E_t[\mathrm{d}^1\mathrm{I/A}]$, the expected growth factor, R_{Eg}, is the difference (high-minus-low), each month, between the simple average of the returns on the two high $E_t[\mathrm{d}^1\mathrm{I/A}]$ portfolios and the simple average of the returns on the two low $E_t[\mathrm{d}^1\mathrm{I/A}]$ portfolios.

References

Ball, Ray, Joseph Gerakos, Juhani Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *Journal of Finance* 55, 389–406.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2019, Which factors? *Review of Finance* 23, 1–35.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An augmented q-factor model with expected growth, *Review of Finance* 25, 1–41.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.