

An Implementation of Quality Minus Junk

by Anthony Tsou, Eugene Choe, David Kane, Ryan Kwon, Yanrong Song, Zijie Zhu

Abstract The `qmj` package produces quality scores for publicly traded US companies and provides a universe composed of those in the Russell 3000 Index following the approach described in “Quality Minus Junk” (Asness et al., 2014). Quality scores as of January 2016 are included in a precalculated data frame, providing users an overview of stocks in the 2015 Russell 3000 Index. Users can also update these scores themselves with `qmj` functions that download necessary financial and stock data from Google Finance and Yahoo Finance.

Introduction

There is no single, widely accepted definition for the quality of a stock. MSCI’s measure of quality combines the z-scores of three winsorized numbers: Return on Equity, Debt to Equity and Earnings Variability to calculate quality indices for their quality indexes (MSCI, 2013). In “Quality Minus Junk” (Asness et al., 2014), quality is defined as the scaled measure of four components: profitability, growth, safety, and payouts. These components respectively measure earnings relative to costs, change in profits over time, risk in future returns, and “shareholder friendliness” (Asness et al., 2014, p. 4).

`qmj` calculates the quality of publicly traded US stocks. The package uses the Russell 3000, an index composed of the 3000 largest US companies by market capitalization, as a sample universe. This universe was chosen for its relevance and reliable presence on Google Finance and Yahoo Finance, which `qmj` uses to automatically gather data. The size of companies in the Russell 3000 also reduces anomalous data points in our calculations.

Using financial statements and stock price data, `qmj` provides a data frame of quality scores for companies in the Russell 3000 Index, allowing users to immediately access an overview of recent quality scores. Users who desire to modify the data or calculate new quality and component scores can do so with package functions, described in greater detail under the “Recalculating the Data” section.

Data

Companies

The Russell 3000 is an index that tracks roughly 3000 of the largest US-traded companies by market capitalization (Investments, 2015). It is rebalanced annually in June. This package comes with a data frame, `companies_r3k16`, containing the members of the 2015 index by name and ticker¹.

Below is a sample of the first five companies in `companies_r3k16`.

	name	ticker
1	AGILENT TECHNOLOGIES IN	A
2	ALCOA INC	AA
3	AAC HOLDINGS INC	AAC
4	AMERICAN AIRLINES GROUP	AAL
5	ALTISOURCE ASSET MGMT	AAMC

The companies are arranged alphabetically by ticker and were parsed from a text file containing the same content as the online Russell 3000 component list².

Financials

Data from the cash flow statements, income statements, and balance sheets of selected companies are consolidated into a second dataset called `financials_r3k16`.

`financials` is long format data, so for each ticker, there are four rows corresponding to the past four available fiscal years based on 10-K submission date. Since companies have different 10-K submission

¹From <https://www.russell.com/documents/indexes/construction-methodology-us-indexes.pdf>

²That list may be found here: <https://www.lseg.com/en/ftse-russell>

Table 1: Columns in financials_r3k16 (in millions of USD except for per share items)

Abbreviation	Name
AM	Amortization
CWC	Changes in Working Capital
CX	Capital Expenditures
DIVC	Dividends per Share
DO	Discontinued Operations
DP.DPL	Depreciation/Depletion
GPROF	Gross Profits
IAT	Income After Taxes
IBT	Income Before Taxes
NI	Net Income
NINT	Interest and Expense
NRPS	Non-redeemable Preferred Stock
RPS	Redeemable Preferred Stock
TA	Total Assets
TCA	Total Current Assets
TCL	Total Current Liabilities
TCSO	Total Common Shares Outstanding
TL	Total Liabilities
TLSE	Total Liabilities and Shareholder's Equity
TREV	Total Revenue

dates, some companies will occasionally not have data for the past four years, but rather data with some amount of lag.

Here is a sample of the first five rows in financials_r3k16.

	ticker	year	AM	CWC	CX	DIVC	DO	DP.DPL	GPROF	IAT	IBT	NI	NINT	NRPS	RPS
1	A	2015	NA	-328	-98	0.40	NA	253	2041	438	480	401	NA	0	NA
2	A	2014	NA	-204	-205	0.53	NA	384	1976	232	229	549	NA	0	NA
3	A	2013	NA	-89	-195	0.46	NA	372	1907	225	293	734	NA	NA	NA
4	A	2012	NA	-176	-194	0.30	NA	301	3604	1153	1043	1153	NA	NA	NA
5	AA	2014	NA	-1209	-1219	0.12	NA	1372	4769	177	497	268	NA	58	NA
	TA	TCA	TCL	TCSO	TD	TL	TLSE	TREV							
1	7479	3686	976	332.00	1655	3312	7479	4038							
2	10815	5509	1692	335.00	1663	5514	10815	4048							
3	10686	4983	1602	333.00	2699	5400	10686	3894							
4	10536	4629	1893	346.00	2362	5354	10536	6858							
5	37399	8269	5541	1216.66	8852	25093	37399	23906							

Prices

The prices_r3k16 dataset has the closing stock prices and price returns for the past two years for each company. Here is a sample of the data:

```
# A tibble: 10 × 4
  ticker date      pret close
  <fct> <date>      <dbl> <dbl>
1 GSPC  2015-01-02 -0.0112  2058.
2 GSPC  2015-01-05 -0.00217 2021.
3 GSPC  2015-01-06 -0.0158  2003.
4 GSPC  2015-01-07 -0.00824 2026.
5 GSPC  2015-01-08  0.0124  2062.
6 AAL   2015-02-13  0.0165   48.4
7 AAL   2015-02-17  0.0258   48.0
8 AAL   2015-02-18 -0.00349  49.3
9 AAL   2015-02-19  0.0439   49.8
10 AAL  2015-02-20 -0.0193   51.0
```

In addition to the companies selected, for comparison purposes the Standard and Poor's 500 Index (with ticker GSPC) price data will always be provided.

Quality

The `quality_r3k16` data frame contains the quality scores for each company as well as the four z-scores measuring profitability, growth, safety, and payouts. These z-scores were calculated following the methodology described [below](#), using data from the `prices_r3k16` and `financials_r3k16` data sets.

	name	ticker	profitability	growth	safety	payouts	quality
1	AMERICAN HOMES 4 RENT	AMH	-0.297	26.648	0.387	-0.585	10.631
2	ARATANA THERAPEUTICS IN	PETX	-0.755	-0.127	-1.411	20.469	7.388
3	L BRANDS INC	LB	8.613	-0.004	3.776	0.259	5.139
4	SFX ENTERTAINMENT INC	SFXE	-1.003	13.191	-2.414	-1.035	3.552
5	RITE AID CORP	RAD	6.249	-0.011	2.235	0.208	3.529

Here we have an example showing the top five companies based on quality measurements.

Calculating Quality

We calculate quality scores for publicly traded companies in the Russell 3000 Index by calculating the z-score of the sum of each company's profitability, growth, safety, and payouts z-scores.

$$Quality = z(Profitability + Growth + Safety + Payouts)$$

Profitability

Profitability is a company's profits per unit of book value. It is composed of six variables: gross profits over assets (*GPOA*), return on equity (*ROE*), return on assets (*ROA*), cash flow over assets (*CFOA*), gross margin (*GMAR*), and accruals (*ACC*). *GPOA* is calculated as gross profits (*GPROF*) over total assets (*TA*).

$$GPOA = \frac{GPROF}{TA}$$

ROE is calculated as net income (*NI*) over book equity (*BE*), which is shareholders' equity (the difference of total liabilities and shareholders' equity (*TLSE*) with total liabilities (*TL*)) - preferred stock (the sum of redeemable preferred stock (*RPS*) and non redeemable preferred stock (*NRPS*)).

$$ROE = \frac{NI}{BE}$$

ROA is calculated as *NI* over *TA*.

$$ROA = \frac{NI}{TA}$$

CFOA is calculated as *NI* + depreciation (*DP.DPL*) - changes in working capital (*CWC*) - capital expenditures (*CX*) all over *TA*.

$$CFOA = \frac{NI + DP.DPL - CWC - CX}{TA}$$

GMAR is calculated as *GPROF* over total revenue (*TREV*).

$$GMAR = \frac{GPROF}{TREV}$$

Finally, *ACC* is calculated as *DP.DPL* - *CWC* all over *TA*.

$$ACC = \frac{DP.DPL - CWC}{TA}$$

We then standardize all components of profitability to z-scores and then standardize the sum of the scaled profitability scores into z-scores.

$$Profitability = z(z_{gpoa} + z_{roe} + z_{roa} + z_{cfoa} + z_{gmar} + z_{acc})$$

Growth

Growth is a measure of a company's increase in profits. It is measured by differences in profitability across a time span of four years. Though AQR recommends measuring growth across a time span of five years, public information that is both consistent and well-organized in 10-K forms is only available for a time span of four years from our sources. Thus, we measure growth using a time span of four years, which we will update once this year's 10-K form is submitted for each company in the Russell 3000 Index.

$$Growth = z(z_{\Delta gpoa_{i,t-4}} + z_{\Delta roe_{i,t-4}} + z_{\Delta roa_{i,t-4}} + z_{\Delta cfoa_{i,t-4}} + z_{\Delta gmar_{i,t-4}} + z_{\Delta acc_{i,t-4}})$$

Safety

Safety is a measure of required return, with safer stocks having a lower required return. Safety is composed of six variables: beta (*BAB*), idiosyncratic volatility (*IVOL*), leverage (*LEV*), Ohlson's *O* (*O*), Altman's *Z* (*Z*), and earnings volatility (*EVOL*). *BAB* is calculated as the negative covariance of each company's daily price returns ($pret_{c_i}$) relative to the benchmark daily market price returns ($pret_{mkt}$), in this case the S&P 500, over the variance of $pret_{mkt}$.

$$BAB = \frac{-cov(pret_{c_i}, pret_{mkt})}{var(pret_{mkt})}$$

IVOL is the standard deviation of daily beta-adjusted excess returns. In other words, *IVOL* is found by running a regression on each company's price returns and the benchmark, then taking the standard deviation of the residuals. Leverage is -(total debt (*TD*) over *TA*).

$$Leverage = -\frac{TD}{TA}$$

$$O = -\left[-1.32 - 0.407 * \log\left(\frac{ADJASSET}{CPI}\right) + 6.03 * TLTA - 1.43 * WCTA + 0.076 * CLCA - 1.72 * OENEG - 2.37 * NITA - 1.83 * FUTL + 0.285 * INTWO - 0.521 * CHIN \right]$$

ADJASSET is adjusted total assets, which is $TA + 0.1 * (\text{market equity } (ME, \text{ calculated as average price per share for the most recent year} * \text{total number of shares outstanding } (TCSO) - BE))$.

$$ADJASSET = TA + 0.1 * (ME - BE)$$

CPI, the consumer price index, is assumed to be 100, since we only care about the most recent year. *TLTA* is book value of debt (*BD*, calculated as $TD - \text{minority interest } (MI) - (RPS + NRPS)$) over *ADJASSET*.

$$TLTA = \frac{BD}{ADJASSET}$$

WCTA is current assets (*TCA*) - current liabilities (*TCL*) over *TA*.

$$WCTA = \frac{TCA - TCL}{TA}$$

CLCA is *TCL* over *TCA*.

$$CLCA = \frac{TCL}{TCA}$$

OENEG is a dummy variable that is 1 if total liabilities (*TL*) is greater than *TA*.

$$OENEG = TL > TA$$

NITA is *NI* over *TA*.

$$NITA = \frac{NI}{TA}$$

FUTL is income before taxes (*IBT*) over *TL*.

$$FUTL = \frac{IBT}{TL}$$

INTWO is another dummy variable that is 1 if *NI* for the current year and *NI* for the previous year are both negative.

$$INTWO = \text{MAX}(NI_t, NI_{t-1}) < 0$$

CHIN is *NI* for the current year - *NI* for the previous year all over the sum of the absolute value of *NI* for the current year and the absolute value of *NI* for the previous year

$$CHIN = \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|}$$

Altman's *Z* is calculated using weighted averages of working capital (*WC*, calculated as *TCA* - *TCL*),

$$WC = TCA - TCL$$

retained earnings (*RE*, calculated as *NI* - dividends per share (*DIVC*) * *TCSO*),

$$RE = NI - DIVC * TCSO$$

earnings before interest and taxes (*EBIT*, calculated as *NI* - Discontinued Operations (*DO*) + (income before tax (*IBT*) - income after tax (*IAT*)) + interest expense (*NINT*)),

$$EBIT = NI - DO + (IBT - IAT) + NINT$$

ME, and *TREV*, all over *TA*.

$$Z = \frac{1.2 * WC + 1.4 * RE + 3.3 * EBIT + 0.6 * ME + TREV}{TA}$$

EBIT is likely an overestimate for a given company due to potentially missing information. *EVOL* is calculated as the standard deviation of *ROE* for a four year span. AQR recommends the past five years, but for the same reason stated in the Growth section, we use a four year span.

$$EVOL = \sigma \left(\sum_{i=t-4}^t ROE_i \right)$$

Likewise, we standardize each variable and then standardize each safety measure, so

$$Safety = z(z_{bab} + z_{ivol} + z_{lev} + z_o + z_z + z_{evol})$$

Payouts

Payouts is a general measure of the company's friendliness to shareholders. It is composed of three variables: net equity issuance (*EISS*), net debt issuance (*DISS*), and total net payout over profits (*NPOP*). *EISS* is calculated as the negative log of the ratio of *TCSO* of the most recent year and *TCSO* of the previous year.

$$EISS = -\log \left(\frac{TCSO_t}{TCSO_{t-1}} \right)$$

Though AQR uses split-adjusted number of shares, we are currently using *TCSO* given available information and will adjust for splits in future iterations of qmj. *DISS* is calculated as the negative log of the ratio of *TD* of the most recent year and *TD* of the previous year.

$$DISS = -\log \left(\frac{TD_t}{TD_{t-1}} \right)$$

NPOP is calculated as *NI* - Δ *BE* over a four year span all over sum of *GPROF* for the past four years (for the same reason as explained in the Growth section).

$$NPOP = \frac{NI - \Delta BE}{\sum_{i=t-4}^t GPROF_i}$$

Recalculating the Data

`qmj` is scheduled to update its datasets periodically, at the minimum as the Russell 3000 index rebalances every year. Users however may also directly update, modify, and create comparable data with `qmj`'s functions. This feature is important for those who want to calculate quality scores for their own list of publicly traded US companies, or for those who wish to factor in new data based on recent events or returns.

`get_companies()` will create a companies data set from a text file of space separated company names and tickers. For example, a valid text file input for `get_companies()` would be:

```
APPLE AAPL
```

```
GOOGLE GOOG
```

The user, however, will have to manually create the new text file. `get_prices()` takes a data frame of companies from `get_companies()`, containing the columns "name" and "ticker," and returns the daily prices and returns for the past two years including the most recent trading day. `get_info()` also takes a similarly formatted data frame of companies and grabs the four most recent company 10-K financial statements if available. `get_info()` does not need to be called often since it will only grab new data if a company files a new 10-K statement.

```
> raw_prices <- get_prices(companies_r3k16)
> raw_data <- get_info(companies_r3k16)
```

This raw data is neither easily readable nor usable elsewhere in the package however. We need to tidy the data before it may be used in quality calculations.

```
> clean_prices <- tidy_prices(raw_prices)
> clean_data <- tidyinfo(raw_data)
```

`tidy_prices()` takes as input the result of `get_prices()` and `tidyinfo()` takes as input the result of `get_info()`. The column names of `clean_data` will be the same abbreviations that are described in the [table under the Financials data subsection](#). The quality data frame can then be generated using `market_data()`.

```
> quality_data <- market_data(companies_r3k16, clean_data, clean_prices)
```

Handling Missing Data

A key feature of `qmj` is its ability to automatically retrieve 10-K financial statements and stock price data from public sources for a given data frame of companies through `quantmod` (Ryan et al., 2015). As this information is sourced from Google Finance and Yahoo Finance, key pieces of data may be missing. Consequently, when calling the `market_data()` function, filters exist in order to improve the accuracy of our scaled results.

Specifically, for financial documents, we're interested solely in companies for which we can fulfill the following criteria:

- All companies must have 3-4 contiguous 10-K filings
- Those contiguous 10-K filings must contain a filing that took place within two years (with some leeway) of the present day.

Companies which fail the above criteria will not be considered or included in the resulting data frame.

Similarly, we filter companies which we deem to have inadequate price data to give a good representation of the company. Explicitly, we ensure that **all measured companies have at least 80% of the maximal number of price data points**.

Conclusion

The `qmj` package offers an R implementation of the methodologies found in the paper, "Quality Minus Junk" (Asness et al., 2014), providing tools to calculate quality scores in addition to profitability, growth, safety, and payout subscores for publicly traded US companies. `qmj` also serves as a useful reference tool with its focus on the Russell 3000 and provided data, allowing users to immediately browse regularly updated quality scores for the largest public US companies.

Bibliography

- C. Asness, A. Frazzini, and L. Pedersen. Quality minus junk. *AQR Capital Management*, 2014. URL <http://ssrn.com/abstract=2312432>. [p1, 6]
- R. Investments. Russell u.s equity indexes: Construction and methodology, 2015. URL <https://www.russell.com/documents/indexes/construction-methodology-us-indexes.pdf>. [p1]
- MSCI. Msci quality indices methodology, May 2013. URL https://www.msci.com/eqb/methodology/meth_docs/MSCI_Quality_Indices_Methodology.pdf. [p1]

Anthony Tsou
Williams College
2460 Paresky Center
Williamstown, MA 01267
United States
anttsou@gmail.com

Eugene Choe
Williams College
1104 Paresky Center
Williamstown, MA 01267
United States
ec7@williams.edu

David Kane
Harvard University
IQSS
1737 Cambridge St
CGIS Knafel Building, Room 350
Cambridge, MA 02138
United States
dave.kane@gmail.com

Ryan Kwon
Williams College
1309 Paresky Center
Williamstown, MA 01267
United States
rynkwn@gmail.com

Yanrong Song
Columbia University
116th and Broadway
New York, NY 10027
United States
yrsong129@gmail.com

Zijie Zhu
Columbia University
116th and Broadway
New York, NY 10027
United States
zijie.miller.zhu@gmail.com