

An Implementation of Quality Minus Junk

by Anthoney Tsou, David Kane, Ryan Kwon

(Campbell et al., 2007) (Gray and Carlisle, 2013) (Kane et al., 2011) (Kane and Lu, 2013)
Abstract The **qmj** package produces quality scores for companies based on the work of Asness et. al (2013). It measures the quality of each of the 3000 largest US companies from the Russell 3000 Index based on profitability, growth, safety, and payout, using the latest available data from Google Finance. The package includes tools to automatically gather relevant financial documents and stock return data, allowing users to update their data whenever desired. The package also provides utilities for analyzing the scores of individual companies, various plotting and filtering tools, and generally helps separate the list of companies into “junk” stocks, which are expected to underperform relative to the market, and “quality” stocks, which are expected to outperform.

Introduction

A stock’s quality represents how much investors should be willing to pay for it, all else equal. Thus, the return of a high-quality stock is intended to exceed its purchase price. Naturally, the return of a high-quality stock is at its highest when its purchase price is low. When a high-quality stock yields a high return, however, investors become more willing to pay a higher price for it, decreasing its return and creating a cycle of undervaluing and overvaluing the stock.

A Quality Minus Junk approach (Asness et al., 2013), which involves investing long on high-quality stocks and shorting low-quality stocks, produces high risk-adjusted returns. The **qmj** package evaluates the quality of stocks by measuring the profitability, growth, safety, and payouts of a company relative to other companies in the Russell 3000 Index. We define each of these characteristics in the following section of the paper.

Drawing data from financial statements, the **qmj** package uses regression based analysis and calculations with closed form expressions to produce a quality z-score for each company of interest. We demonstrate the use of a variety of tools for both casual and diligent investors to not only view the quality of stocks, but also to understand the underlying factors that compose the scores.

Calculating Quality

We calculate quality scores for publicly traded companies in the Russell 3000 Index by summing the z-scores for each company’s profitability, growth, safety, and payouts. We attempt to perform the same calculations as Asness, but we have a few adjustments given the availability of data from public sources.

Profitability

Profitability 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 $NI + \text{depreciation } (DP.DPL) - \text{changes in working capital } (CWC) - \text{capital expenditures } (CX)$ 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 all profitability scores into z-scores.

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

Growth

Growth 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, and it is still too early in the most recent year (2015) for most companies to have submitted a 10-K form. 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. As of now,

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

Safety

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 $-(\text{total debt } (TD) \text{ over } TA)$.

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

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

$ADJASSET$ 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 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 $NI - \Delta BE$ 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).

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

Data

The package **qmj** comes pre-compiled with data, using a universe comprised of all companies in the Russell 3000 Index as of February 1, 2015. The Russell 3000 Index is a list of the 3000 largest US companies, according to market cap, updated yearly, usually around May or June. The Russell 3000 Index was chosen due to the US-centric nature of our financial sources (Google Finance and Yahoo Finance), in addition to providing reliable, interesting data (i.e. no anomalies such as Gross Profits over Assets (GPOA) doubling for a tiny company due to a small absolute change in profit occurs).

For this data, **qmj** is able to automatically retrieve the most recent list of companies, retrieve the previous four 10-K's when possible.

```
> library(qmj)
> data(companies)
> data(financials)
> data(prices)
> data(quality)
> head(quality, n=5)
```

	ticker	name	profitability	growth	safety	payouts	quality
1	ANGI	ANGIES LIST INC	-0.158	24.558	-0.891	-1.870	21.639
2	SBCF	SEACOAST BANKING CORP F	-0.955	21.177	0.236	-1.0821	19.376
3	UHAL	AMERCO	-0.335	19.256	-0.161	-1.986	16.775
4	GUID	GUIDANCE SOFTWARE INC	0.225	13.618	0.0957	-1.559	12.379
5	BRO	BROWN & BROWN INC	0.148	-0.0406	11.766	0.225	12.0991

'companies' stores names and tickers from the Russell 3000 Index component list, 'financials' stores, whenever possible, the four most recent years of financial data (balance sheets, income statements, and cash flow statements) which are relevant to our calculations. 'prices' stores closing stock prices and calculated price returns for each company as well as the S&P 500 for the past two years, and 'quality' is the list of quality scores, as well as the component scores, of each of our companies.

Updating Data

Though **qmj** keeps datasets updated, it has a few functions that can extract information directly from Google Finance to grab the most recent data.

```
> #raw_prices <- get_prices(companies)
> #raw_data <- get_info(companies)
```

get_prices() takes a data frame of companies, organized by 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 data frame of companies, organized by name and ticker, and grabs the most recent company 10-K financial statements. Thus, **get_info()** does not need to be called often since it will only grab new data once per year. Both functions will return a data frame that can be organized easily. An easy way to make the data more readable is through tidy functions in the **qmj** package.

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

tidy_prices() takes as input the result of **get_prices()**, which is assigned here as **raw_prices**, and organizes the data into columns for ticker, date, price, and price return. **tidyinfo()** takes as input the result of **get_info()**, which is assigned here as **raw_data**, and organizes the data into columns for ticker, year, and various items found in company financial statements such as total assets and net income. The column names themselves are abbreviations that are used in the Appendix.

Analyzing the Universe (Of Companies)

In the quality data set, it can quickly be seen that the growth score for Angies List Inc. is abnormally high, and accounts for virtually all of its quality score. In many cases, it is undesirable to consider companies with high quality scores that are “driven” (here defined as composing at least half the quality score) by a single component score. `qmj` provides a filter.

```
> data(quality)
> head(quality)
```

	ticker	name	profitability	growth	safety	payouts	quality
1	ANGI	ANGIES LIST INC	-0.158	24.558	-0.891	-1.870	21.639
2	SBCF	SEACOAST BANKING CORP F	-0.955	21.177	0.236	-1.0821	19.376
3	UHAL	AMERCO	-0.335	19.256	-0.161	-1.986	16.775
4	GUID	GUIDANCE SOFTWARE INC	0.225	13.618	0.0957	-1.559	12.379
5	BRO	BROWN & BROWN INC	0.148	-0.0406	11.766	0.225	12.0991
6	CFFN	CAPITOL FEDERAL FINL IN	5.754	-0.0490	5.738	0.229	11.672

```
> sans_growth <- filter_companies(quality, filter="growth")
> head(sans_growth)
```

	ticker	name	profitability	growth	safety	payouts	quality
5	BRO	BROWN & BROWN INC	0.1484	-0.0406	11.766	0.225	12.0991
6	CFFN	CAPITOL FEDERAL FINL IN	5.754	-0.0490	5.738	0.229	11.672
8	CENX	CENTURY ALUMINUM CO	-0.191	3.440	6.177	-0.196	9.230
9	CXW	CORRECTIONS CORP OF AME	-0.328	3.220	4.101	0.191	7.184
10	RSE	ROUSE PROPERTIES INC	3.621	0.0470	1.888	0.986	6.541
11	PDCO	PATTERSON COS INC	-0.741	-0.0530	0.166	6.781	6.152

If desirable, we may also select specifically for those companies which are driven by a particular component. Note that ‘remove’ is, by default, set to TRUE, and ‘isolate’, is set to FALSE.

```
> cpayouts <- filter_companies(quality, filter="payouts", remove=FALSE, isolate=TRUE)
> head(cpayouts)
```

	ticker	name	profitability	growth	safety	payouts	quality
11	PDCO	PATTERSON COS INC	-0.741	-0.0530	0.166	6.781	6.152
18	KTWO	K2M GROUP HLDGS INC	0.166	-0.0443	0.789	3.925	4.836
23	CMCO	COLUMBUS MCKINNON CORP	1.482	-0.0389	-0.0215	3.169	4.591
30	WIFI	BOINGO WIRELESS INC	-0.534	-0.0299	0.929	3.663	4.0280
32	SXT	SENSIENT TECHNOLOGIES	0.531	-0.0250	0.443	3.0387	3.988
36	CALD	CALLIDUS SOFTWARE INC	-0.136	-0.0434	-0.413	4.510	3.917

Or, we can select for all companies which are not driven by any component score.

```
> well_rounded <- filter_companies(quality, filter="all")
> head(well_rounded)
```

	ticker	name	profitability	growth	safety	payouts	quality
6	CFFN	CAPITOL FEDERAL FINL IN	5.754	-0.0490	5.738	0.229	11.672
21	HASI	HANNON ARMSTRONG SUSTAI	1.217	-0.0277	1.736	1.728	4.652
22	ADMS	ADAMAS PHARMACEUTICALS	1.692	0.500	2.226	0.211	4.629
24	OPLK	OPLINK COMMUNICATIONS	0.999	1.898	1.211	0.218	4.326
33	WSO	WATSCO INC	1.940	-0.0182	1.876	0.171	3.969
34	PCCC	P C CONNECTION	1.499	-0.0200	0.502	1.979	3.959

It may also be desirable to look at quality scores specific to a subset of our extant universe. For example, it may be desirable to focus on a specific industry, instead of the entire market.

```
> data(companies)
> data(financials)
> data(prices)
> subset_companies <- companies[1:35,]
> subset_qualities <- market_data(subset_companies, financials, prices)
> head(subset_qualities)
```

	ticker		name	profitability	growth	safety	payouts	quality
1	ACN	ACCENTURE	PLC IRELAND	1.497	0.102	1.470	0.684	3.752
2	AAON		AAON INC	1.201	0.125	0.839	0.776	2.940
3	AAPL		APPLE INC	0.562	0.387	0.859	1.0367	2.844
4	AAOI	APPLIED	OPTOELECTRONICS	0.494	1.616	-1.0127	1.226	2.323
5	AAMC	ALTISOURCE	ASSET MGMT	-0.389	2.846	-0.625	0.458	2.290
6	ABBV		ABBVIE INC	1.570	0.0751	0.353	0.197	2.195

Conclusion

In the **qmj** package, we automate AQR's method of assigning quality scores for publicly traded companies in today's market. The package itself provides convenient datasets and utility functions, and it also takes advantage of R's robust nature to allow seamless interaction with functions in the base R package and other packages.

Bibliography

- C. Asness, A. Frazzini, and L. Pedersen. Quality minus junk. *Journal of Computational and Graphical Statistics*, 2013. [p1]
- K. Campbell, J. Enos, D. Gerlanc, and D. Kane. Backtests. *R News*, 7(1):36–41, 2007. [p1]
- W. Gray and T. Carlisle. *Quantitative Value: A Practitioner's Guide to Automating Intelligent Investment and Eliminating Behavioral Errors*. John Wiley & Sons, 2013. [p1]
- D. Kane and Y. Lu. Performance attribution for equity portfolios. *The R Journal*, 5(2):53–61, 2013. [p1]
- D. Kane, A. Liu, and K. Nguyen. Analyzing an electronic limit order book. *The R Journal*, 3(1):64–68, 2011. [p1]

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

David Kane
Williams College
Williamstown, MA 01267
United States
dave.kane@gmail.com

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