

# A Professional Quant Equity Workflow

quant finance By Jonathan Larkin August 31, 2016

Jonathan Larkin, Chief Investment Officer at Quantopian

In my previous post, I laid out a philosophical foundation for producing high Sharpe ratio<sup>1</sup> quantitative investment strategies. Today I'll add substance to that philosophy by giving you a detailed tour of the investment process for a popular and deep area of the quantitative investment world: **cross-sectional equity investing**, also known as equity statistical arbitrage or equity market neutral investing. This approach to equity investing involves holding hundreds of stocks long and short in a tightly risk-controlled portfolio with the goal of capturing transient market anomalies while exhibiting little to no correlation to market direction or other major risk factors. If you want to know how legions of quants at the biggest hedge funds in the world spend their days, then read on.

All cross-sectional strategies can be abstracted into 6 stages: **Data**, **Universe Definition**, **Alpha Discovery**, **Alpha Combination**, **Portfolio Construction**, and **Trading**. Success at each stage is a necessary condition of success; success in one stage alone is not sufficient. Sufficiency arises by making deliberate and thoughtful choices at each stage of the process. A map of this process looks like:



This process is a loop and the cycle time could be once a week, once a day, every 15 minutes, etc.

#### **Data**

Quants are data-driven and data-informed investors. All quant investing starts with data. Fortunately, Quantopian has done the leg work and provides many datasets which have been cleaned, symbol mapped, joined across vendors, and constructed, where possible, point-in-time. At this first step, you need to simply answer the question: "which dataset(s) do I think contain information which would help predict futures returns?" You are mining for gold; first you need to decide where to start mining.

#### **Universe Definition**

Quantopian's historical pricing data today covers about 8,000 US exchange listed securities. Before proceeding to the (perhaps more glamorous) aspects of the strategy, you must first pare down this list to an appropriate set of tradeable securities, i.e., your trading universe. The Q500US and Q1500US have just been released and you can use one of these as is, or leverage the underlying machinery in these to help you generate your own custom universe in this style. You might ask, "why limit ourselves at all? Wouldn't it make sense to use all the available data? Wouldn't that give me the maximum breadth like you discussed in the previous post?" There are some practical matters to deal with first, such as screening out illiquid stocks. However, there is a less obvious, yet absolutely critical, reason for paring down the security list: successful cross-sectional strategies balance a tension between price dispersion and self-similarity in the universe. By definition, cross-sectional strategies extract relative value across securities and, in order to be able to rank something intelligently, there needs to be some degree of uniformity in the characteristics of the things being ranked.

In selecting the universe, you must be mindful of your underlying investment thesis. Consider two examples as an illustration. If you are pursuing a strategy with a thesis based on the information content of overnight versus intraday stock returns<sup>2</sup>, one class of securities that you must be sure to screen out is ADRs. It is not logically consistent to apply a thesis that relies on investor behavior seen through US exchange prices of an ADR when information has diffused into the local share underlying the ADR in another time zone. As a second example, if you are pursuing a strategy in part based on financial statement data, such as the accruals anomaly<sup>3</sup>, you must be sure to screen out stocks to which these measurements or ratios cannot be appropriately applied (in this case, bank stocks).

### Alpha Discovery

An alpha is **an expression, applied to the cross-section of your universe of stocks, which returns a vector of real numbers where these values are predictive of the relative magnitude of future returns**. An alpha could be built from a straight rank or it could be a vector of dimensionless numbers. Alphas are often referred to as factors, and we use this term interchangeably at Quantopian. The Pipeline API is your entry into the world of alpha modeling. At this stage, don't worry about real world details such as trading, commissions, or risk. Create a hypothesis about investor behavior, market structure, information asymmetry or any other potential cause of market inefficiency and see if that hypothesis has any predictive capability. Need some ideas? Try a Google search for "equity market anomalies" or, even better, an SSRN search of the same.

Alpha research is the intersection of art and science and this is where magic happens. Productive alpha research is an iterative process: hypothesize, test, analyze, revise. We recently released a new open source project, currently in beta and available in Quantopian Research, called *alphalens*. You express your alphas with the Pipeline API. You analyze the effectiveness of your alphas with *alphalens*.

### Alpha Combination

It today's markets, rarely is any single alpha significant enough to be the sole basis of an investment strategy. A successful strategy usually includes many individual alphas; if they are strong enough, a few can suffice. The goal at this stage is to implement a weighting scheme which takes as input many normalized alphas and produces a single alpha which is more predictive than the best individual alpha. The weighting scheme can be quite simple: sometimes just adding ranks or averaging your alphas can be an effective solution. In fact, one popular model does just that to combine two alphas. For increased complexity, classic portfolio theory can help you; for example, try solving for the weights such that the final combined alpha has the lowest possible variance. Lastly, modern machine learning techniques can capture complex relationships between alphas. Translating your alphas into features and feeding these into a machine learning classifier is a popular vein of research<sup>4,5</sup>.

### Portfolio Construction

Up until this stage, we've existed in the realm of research, unsullied by the vulgarities of practical implementation. The steps prior are literally in the realm of research as that work is best done in the unstructured Quantopian Research environment, where you can iterate quickly on ideas. The problem changes at this point: we have a final alpha vector and we must introduce this alpha to the real world and structure and trade a portfolio to capture value. At each cycle time of the process, we calculate the alphas, combine the alphas into a final alpha, inherit the portfolio from the previous period, take our final alpha and calculate a new ideal portfolio, and generate a trade list to transition from the previous portfolio to the ideal portfolio.

There are many questions to answer in defining the step to create your ideal portfolio: Which risks do you want to be cognizant of (i.e., your risk model)? What is the objective function in the portfolio construction step? How will the portfolio be constrained?

These three questions are always formally answered, either implicitly or explicitly. Let's work today with only the simplest technique: construct a portfolio based on the quantiles of your final alpha vector; i.e., your longs could be equal weights in names in the top quintile and your shorts could be equal weights in names in the bottom quintile with weights set so the portfolio reaches some total invested value long and short, and such that the value of longs equals the (absolute) value of shorts.

Complexity in portfolio construction builds when the answers to each or all of the three questions become more complex.

### Trading

The output of the portfolio construction stage is an ideal portfolio and a trade list to transition the current portfolio to the ideal portfolio. Trading is the stage of, well, effecting trades in the market. From the characteristics of each choice you have made thus far, you need to answer implementation questions: How fast do I need to trade? How quickly does the predictive power of the alpha decay? Does it make more sense to be passive and execute slowly in the market, or, conversely, does it make more sense to execute aggressively and immediately? You can gain insight into these questions both by viewing the turnover and alpha persistence analysis in the *alphalens* output for your final alpha as well as looking at the turnover and round trip analysis in the *pyfolio* output for your fully formed strategy. If universe selection is a balance between price dispersion and self-similarity, and portfolio construction is a balance between risk and return, then trading is a balance amongst alpha decay, explicit costs, implicit costs, and information leakage about your total intent.

\*\*\*

You might be wondering, "Do I have to use this exact process to be successful? Or to get an allocation for my algo from Quantopian?" Not necessarily. We look for robust high Sharpe ratio strategies that perform well out-of-sample; there is no one perfect way to achieve this. The search space is very large however, and in this post I've outlined a structure that shows how some of the largest and most successful quantitative investors have approached the problem. "To follow the path, look to the master, follow the master, walk with the master, see through the master, become the master." This structure gives you one framework within which you can improvise and innovate.

[1] Sharpe Ratio is a statistical measurement of the risk adjusted performance of a portfolio, and is calculated by dividing a portfolio's average return by the standard deviation of its returns. It shows a portfolio's reward per unit of risk and is useful when comparing two similar portfolios. As the Sharpe Ratio increases, the better its performance.

[2] Aboody, David and Even Tov, Omri and Lehavy, Reuven and Trueman, Brett, Overnight Returns and Firm-Specific Investor Sentiment (April 11, 2016). Available at SSRN: <https://ssrn.com/abstract=2554010> or <https://dx.doi.org/10.2139/ssrn.2554010>

[3] Dechow, Patricia M. and Khimich, Natalya V. and Sloan, Richard G., The Accrual Anomaly (March 22, 2011). Available at SSRN: <https://ssrn.com/abstract=1793364> or [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1793364](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1793364)

[4] Creamer, Germán G. and Freund, Yoav, Automated Trading with Boosting and Expert Weighting (April 1, 2010). Quantitative Finance, Vol. 4, No. 10, pp. 401–420. Available at SSRN: <https://ssrn.com/abstract=937847>

[5] Huerta, Ramon and Elkan, Charles and Corbacho, Fernando, Nonlinear Support Vector Machines Can Systematically Identify Stocks with High and Low Future Returns (September 6, 2012). Algorithmic Finance (2013), 2:1, 45-58. Available at SSRN: <https://ssrn.com/abstract=1930709> or <https://dx.doi.org/10.2139/ssrn.1930709>

The material on this website is provided for informational purposes only and does not constitute an offer to sell, a solicitation to buy, or a recommendation or endorsement for any security or strategy, nor does it constitute an offer to provide investment advisory services by Quantopian.

In addition, the material offers no opinion with respect to the suitability of any security or specific investment. No information contained herein should be regarded as a suggestion to engage in or refrain from any investment-related course of action as none of Quantopian nor any of its affiliates is undertaking to provide investment advice, act as an adviser to any plan or entity subject to the Employee Retirement Income Security Act of 1974, as amended, individual retirement account or individual retirement annuity, or give advice in a fiduciary capacity with respect to the materials presented herein. If you are an individual retirement or other investor, contact your financial advisor or other fiduciary unrelated to Quantopian about whether any given investment idea, strategy, product or service described herein may be appropriate for your circumstances. All investments involve risk, including loss of principal. Quantopian makes no guarantees as to the accuracy or completeness of the views expressed in the website. The views are subject to change, and may have become unreliable for various reasons, including changes in market conditions or economic circumstances.