CleverQuant

Stefan Negritoiu
stefann@stefann.com

1. Overview

In today's world the financial industry employs computer software to a great extent, particularly in the field of equity analysis. Unfortunately, most analysts' work still involves much routine manual labor which is not inherently part of the actual analysis, but rather a requirement of the programs used. There is no over-arching software tool that is aware of the analysts' job to help reduce the number of operation involved in data entry and collection and with the subsequent statistical and mathematical investigation.

CleverQuant is a computer software modeling tool aware of the specific procedures involved in equity analysis, from data collection to financial modeling and equity valuation. It is designed to aid financial analysts and researchers by automating the process of gathering input data from disparate sources, and running it through customizable financial models. The key concept behind this tool is that it does not mandate or propose neither specific input factors nor financial models. Instead, it gives the user the flexibility to tap into any relevant data sources and to run those through as many variations of financial models as he or she wants.

An important consequence of using CleverQuant for equity analysis and financial markets modeling is the reduced time needed to confirm or refute a model or investment strategy. This allows for the evaluation of a much higher number of combinations of factors used to form models. Ultimately, it should lead to a better understanding of the market and to more accurate valuations.

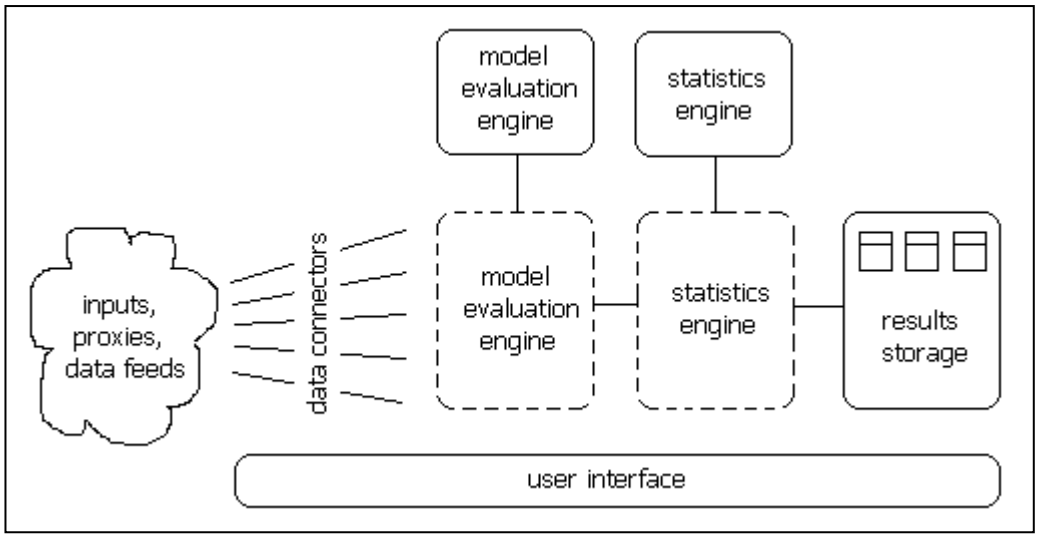
2. Background and motivation

There are two fundamental necessities behind the creation of CleverQuant: realistic and accurate models and a rigorous and less time consuming procedure for validating them. The more input factors a model takes into account, the more realistic it is. Unless we desire to focus on a particular characteristic of the modeled entity, keeping everything else constant or even ignoring everything else will only produce less accurate forecasts.

Although it is beyond the scope of this paper to present the debate around the efficient market hypothesis, empirical evidence suggests that markets have temporary efficiencies at best and that there exist long term deviations from a "recognized" average return. Such deviations might be generated by extrinsic factors like the state of the economy and the investors' degree of trust in the stock market. A realistic model must take into account all such decision factors.

Unfortunately, we are constrained to small degrees of complexity for our models due to the extensive time it would require to verify the model on a relevant sample of data. More often than not however, we realize the necessity to consider extrinsic factors, we rule to its favor, and we fall into the trap of too little testing. This is one explanation for the traditional dichotomy between one researcher publishing a new model and another finding counter-examples after as much as several years.

There exists sufficient empiric evidence that other "equity attributes" [1] aside from the ones considered by the traditional dividend discount model for instance need to be taken into consideration. Jacobs and Levy show that some ratios like cash flow to stock price and sales to stock price and momentum-like measures such as downward earnings estimate revisions and the number of analysts following a stock give surprising insight into returns variation.



Similar to these findings is the study of French and Fama about the correlation between risk and return. Their work is consistent with the overreaction hypothesis and it concludes that all too often a higher level of return does not yield necessarily from a higher level of risk [2]. Furthermore, when studying the dividends paid by stocks in the S&P Composite Stock Price Index, Shiller finds "no such tendency of stock price to forecast the dividend present value" [3]. All this is rather convincing evidence that no matter how much theoretical sense the efficient market hypothesis makes, it does not hold in real life; possibly because of the markets' inability to be purely and truly efficient or because of exuberant human nature.

But perhaps the most important aspect of today's equity analysis is that we don't have a proven one-size-fits-all model. This means we still need to experiment. It also means there might not exist a model applicable to all businesses, industries and development stages. In his recent book, English presents methodologies for reading financial statements for early-stage and later-stage companies [4]. He also exemplifies how to construct financial models tailored to a specific company.

Such recommendations suggest a departure from generally applicable mathematical formulas and a tendency towards industry or company-specific models. All the more reasons to have a tool like CleverQuant.

3. Architecture

Given CleverQuant's goal to reduce the time required to evaluate financial models it should not come as a surprise that it spans the entire process of equity analysis. It aggregates and automates all operations from data selection and collection to evaluation of models and analysis of results. A graphical overview of the components which make up this modeling tool can be seen in Figure 1. A better illustration of the how CleverQuant works is to show how it eliminates many of the tasks that are not inherent to the job of equity valuation.

In today's world a researcher needs to corroborate data from many distinct sources. Discount rates come from central banks, macroeconomic factors from governmental and non-governmental research institutions, companies' financial statements from commercial databases like Compustat® and historic and real-time data about stock pricing from specialized data feeds. In addition, the researcher might create custom factors to be used in models such as APT. These factors could be proxies for other portfolios in which case the procedure for forming the model involves even more data sources.

CleverQuant provides connectors and data adaptors to the entire range of data sources. The role of the data adaptor is to normalize incoming data. This way, all pertinent information is available in a single place and in a single format making it very easy to include in models.

As an analogy, imagine a library reduced to a single rack of shelves with one book representing each subject heading. Further, let the one book available from each subject heading also be representative of that topic. Thus, if one wishes to build a model of the world or of human life, all one needs to do is reach out and grab the appropriate books and put them in a specific sequence on the floor. Afterwards, a third party can come and read the books (evaluate the model) and get a perception of the world.

The perception could be more or less accurate depending on how relevant the chosen subject headings are to the breadth of human life and the sequence in which they were arranged. It is important however to recognize that all factors that stand as proxies for concepts in real-life were available within hand's reach and that there was a common addressing and locating scheme so that searching through them was easy, fast and straightforward.

Following the data selection and collection process, the researcher chooses a model, represented by a set of mathematical formulas. Upon evaluation, the results are put through a statistics engine and contrasted with other outcomes or with known values. Today, analysts use applications like Excel or Matlab to perform these tasks. Manual tasks arise at data input time and when comparing results obtained from distinct programs. CleverQuant eliminates this part of the process.

4. Implementation

Even though the idea of having an over-arching modeling tool aware of the equity analysis process is perhaps not that revolutionary, there were significant obstacles preventing the its implementation in the past. Key to constructing CleverQuant is a standardized and open mechanism for data-exchange as well as a facility for easily building software from "prefabricated" components.

Both requirements have only been addressed recently by the wide-spread adoption of XML as a common format for data exchange and web services as a cross-platform facility for accessing components over the network. From this perspective it is the fist time in computer software history when one can build an application from existing "building blocks" without necessarily owning them.

It follows then that CleverQuant will use XML as a methodology for automated data input. And since value does not come from reinventing the wheel, the project does not propose a new mathematics and statistics engine. Instead, it will use web services to tap into existing software.

5. User scenarios

The best way to describe CleverQuant is to show how it can enhance current research. In their study on size and book-to-market factors [5], Fama and French devise several rules based on which they rank firms and cluster them together in portfolios. They also derive a model for relating book-to-market equity to expected returns starting from the dividends paid by a company in a certain period. The results are several graphs plotting time series for various ratios plus the results of regressions between the chosen portfolios and several standard measures like the market portfolio and the risk free rate.

The authors disclose that initial company data comes from Compustat. It is reasonable to assume that the actual results are the outcome of plugging in the numbers from the database into a spreadsheet program and plotting the results. This is a tedious, but manageable process. Unfortunately, it becomes unmanageable as the researcher allows for more flexible data acceptance criteria.

In this particular example, the authors decide *a priori* that small cap stocks are the bottom 30 percent, medium stocks the middle 40 percent and large stocks the top 30 percent of all NYSE stocks.

Their conclusions become bound to these "hard coded" numbers. If at a later point someone decides that this is an incorrect division of stocks by size much of their work is wasted.

The same assumptions are made by Daniel and Titman in their study which extends the French and Fama paper [6]. Using CleverQuant, the tedious aspect of their research is eliminated because the process of data selection and input into the modeling engine is automated. A researcher merely has to say "I want to have x portfolios delimited by rules a, b, and c where a, b, and c are permutations of values in this particular set of numbers" and the program generates the appropriate data sets.

The situation complicates when models require inputs that originate in different data stores. In their paper on stocks and bond returns [7], Fama and French corroborate equity and bond "attributes" in their attempt to isolate those factors which determine variations in returns. It is fairly straightforward to imagine that the number of combinations that needs to be tested grows more than proportional with the number of inputs and data sources where they originate. This proves such a significant obstacle in real life that only few combinations are ever tested. CleverQuant could change the odds.

There are several features common to most research in financial economics and equity analysis: data selection, data input, model creation, model evaluation, outcome comparison (regression against known variables) and outcome interpretation. A researcher's or analyst's time is best spent doing data selection, model creation and outcome interpretation. Computer software should do the rest.

6. Conclusion

CleverQuant is an attempt at using the power of computer software and computer networks to build a modeling tool able to capture the breadth of decision factors that ultimately give value to investment opportunities. Such a program should be connected enough to access real-time data feeds so that assumptions can be quickly and easily verified. I strongly believe software like CleverQuant will ultimately make analysts and researchers more knowledgeable because it will allow them to spend more time envisioning proper models and less time dealing with the collection and manipulation of data necessary to sustain or reject their claims.

Looking towards the future we should delegate more and more tasks to carefully crafted software. We can ultimately envision programs that can assist financial officers with hedging decision for instance. Today there is a myriad of options for managing operational risk through hedging. Choices can vary between several financial instruments (underlying asset, futures, options) as well as within a certain financial instrument (term, denomination of contract) to such an extent that it is difficult to argue that a certain choice is the best choice out of all possible ones. Computers however, can evaluate a much higher number of options in a much shorter amount of time thus coming much closer to a guarantee that a certain choice was indeed optimal.

References

1. Bruce I. Jacobs and Kenneth N. Levy, *Equity Management: Quantitative Analysis for Stock Selection*, McGraw-Hill, New York, 2000
2. Eugene Fama and Kenneth French, *The Cross Section of Expected Stock Returns*, Journal of Finance, June 1992
3. Robert J. Schiller, *Irrational Exuberance*, Princeton University Press, 2000
4. James English, *Applied Equity Analysis*, McGraw-Hill, 2001
5. Eugene Fama and Kenneth French, *Size and Book-to-Market Factors in Earnings and Returns*, Journal of Finance, March 1995
6. Kent Daniel and Sheridan Titman, *Evidence on the Characteristics of Cross Sectional Variation in Stock Returns*, Journal of Finance, March 1997
7. Eugene Fama and Kenneth French, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics