Finding Alphas
FINDING ALPHAS

A Quantitative Approach to Building Trading Strategies

Igor Tulchinsky et al.
WorldQuant Virtual Research Center
Dedicated to All at WorldQuant —

The Future of Trading
Contents

Preface xi
Acknowledgments xiii
About the WebSim™ Website xv

PART I  INTRODUCTION 1
1 Introduction to Alpha Design 3
   By Igor Tulchinsky
2 Alpha Genesis – The Life-Cycle of a Quantitative Model of Financial Price Prediction 7
   By Geoffrey Lauprete
3 Cutting Losses 13
   By Igor Tulchinsky

PART II  DESIGN AND EVALUATION 19
4 Alpha Design 21
   By Scott Bender/Yongfeng He
5 How to Develop an Alpha. I: Logic with an Example 27
   By Pankaj Bakliwal
6 How to Develop an Alpha. II: A Case Study 31
   By Hongzhi Chen
7 Fundamental Analysis 43
   By Xinye Tang/Kailin Qi
8 Equity Price and Volume 49
   By Cong Li
9 Turnover 51
   By Pratik Patel
10 Backtest – Signal or Overfitting  
By Peng Yan  
55
11 Alpha and Risk Factors  
By Peng Wan  
61
12 The Relationship between Alpha and Portfolio Risk  
By Ionut Aron  
65
13 Risk and Drawdowns  
By Hammad Khan  
71
14 Data and Alpha Design  
By Weijia Li  
79
15 Statistical Arbitrage, Overfitting, and Alpha Diversity  
By Zhuangxi Fang  
85
16 Techniques for Improving the Robustness of Alphas  
By Michael Kozlov  
89
17 Alphas from Automated Search  
By Yu Huang  
93
18 Algorithms and Special Techniques in Alpha Research  
By Sunny Mahajan  
97

PART III EXTENDED TOPICS  
101
19 Impact of News and Social Media on Stock Returns  
By Wancheng Zhang  
103
20 Stock Returns Information from the Stock Options Market  
By Swastik Tiwari  
109
21 Introduction to Momentum Alphas  
By Zhiyu Ma  
117
22 Financial Statement Analysis  
By Paul A. Griffin  
119
23 Institutional Research 101  
By Benjamin Ee  
127
24 Introduction to Futures Trading  
By Rohit Agarwal  
145
25 Alpha on Currency Forwards and Futures  
By Richard Williams  
151
## Contents

### PART IV  NEW HORIZON – WEBSIM™  155

26  Introduction to WebSim™  157  
   *By Jeffrey Scott*

27  Alphas and WebSim™ Fundamentals  165  
   *By the WebSim™ Team*

28  Understanding How WebSim™ Works  169  
   *By the WebSim™ Team*

29  API Reference  179  
   *By the WebSim™ Team*

30  Interpreting Results and Alpha Repository  187  
   *By the WebSim™ Team*

31  Alpha Tutorials  199  
   *By the WebSim™ Team*

32  FAQs  211  
   *By the WebSim™ Team*

33  Suggested Readings  223  
   *By the WebSim™ Team*

### PART V  A FINAL WORD  229

34  The Seven Habits of Highly Successful Quants  231  
   *By Richard Hu*

References  235
Index  245
This book is a study of the process of finding alphas. The material is presented as a collection of essays, providing diverse viewpoints from successful quants on the front lines of quantitative trading. A wide variety of topics is covered, ranging from theories about the existence of alphas, to the more concrete and technical aspects of alpha creation.

Part I presents a general introduction to alpha creation, and is followed by a brief account of the alpha life-cycle, and insights on cutting losses.

Part II focuses more on the technical side of alpha design, such as the dos and don’ts of information research, key steps to developing an alpha, and the evaluation and improvement of quality alphas. The key technical aspects discussed in this section are turnover, backtesting, fundamental analysis, equity price volume, statistical arbitrage, overfitting, and alpha diversity.

Part III explores ad hoc topics in alpha design, including alpha design for various asset classes like futures and currencies, the development of momentum alphas, and the effect of news and social media on stock returns.

In Part IV, we introduce you to WebSim™, a web-based alpha development tool. We invite all quant enthusiasts to utilize this free tool to learn about alpha backtesting (also known as alpha simulation), and ultimately to create their own alphas.

Finally, in Part V, we present an inspirational essay for all quants who are ready to explore the world of quantitative trading.
Acknowledgments

In these pages, we present readers with a collection of writings on the alchemic art of finding alphas. It is written by WorldQuant’s founder, directors, managers, in-house portfolio managers, and quantitative researchers. The key objectives of this collection are twofold – to present many viewpoints as to how to define an alpha, and how to find one. At WorldQuant, we believe no viewpoint is the best and only answer, and that a variety of approaches is always superior to a single one. We also present our online financial markets simulation tool known as WebSim™, which lets users and consultants create, test, simulate, and track alphas.

WorldQuant would like to thank Rohit Agarwal, Ionut Aron, Pankaj Bakliwal, Scott Bender, Hongzhi Chen, Benjamin Ee, Zhuangxi Fang, Paul A. Griffin, Yongfeng He, Richard Hu, Yu Huang, Hammad Khan, Michael Kozlov, Geoffrey Lauprete, Cong Li, Weijia Li, Zhiyu Ma, Sunny Mahajan, Pratik Patel, Kailin Qi, Jeffrey Scott, Xinye Tang, Swastik Tiwari, Igor Tulchinsky, Peng Wan, Richard Williams, Peng Yan, and Wancheng Zhang for their contributions, polishing efforts, and time invested in making this book a reality.

A special note of thanks to Wendy Goldman Rohm, our literary agent, for her critical and insightful comments on early drafts, her awesome proofing, and many great suggestions throughout the project; and to Werner Coetzee and the teams at John Wiley and Sons for their expert guidance and detailed and helpful advice. Many thanks also to Jeffrey Blomberg and Kristin Chach, WorldQuant’s most dedicated legal team, for their wise counsel and tireless work to keep us on the track and guide us beyond. And thanks to Tracy Tseung for her timely project management and editorial assistance.

Finally, we would like to acknowledge with gratitude the support and faith of every colleague at WorldQuant. Thank you all.
DISCLAIMER

The contents of this book are intended for informational and educational purposes only and, as such, are not intended to be nor should be construed in any manner to be investment advice. The views expressed are those of the various contributors and do not necessarily reflect the view or opinion of WorldQuant or WorldQuant Virtual Research Center.
About the WebSim™ Website

At the time of writing, the WebSim™ information contained in this document is consistent with the WebSim™ website. Since the website is subject to change, in cases where there exist inconsistencies between this document and the website, the terms of the WebSim™ website will govern the most updated and current processes of WebSim™. For the most up-to-date version of WebSim™ and the terms applicable to use of WebSim™, please go to https://websim.worldquantchallenge.com or its successor site.

Registration at WebSim’s™ official website is required to obtain the full functionality of the tool, and to have access to the WebSim™ support team. Successful alphas may, in certain cases, be considered for inclusion in actual quant trading investment strategies managed by WorldQuant.

WEBSIM™ RESEARCH CONSULTANTS

WorldQuant has established a Research Consultant program for qualified individuals to work with our web-based simulation platform, WebSim. This program gives consultants the flexibility to create alphas in their own physical and intellectual environment. This is a particularly ideal pursuit for individuals who are undertaking college education as well as those who are ambitious and highly interested in breaking into the financial industry.

Qualified candidates are those highly quantitative individuals who typically come from STEM (Science, Technology, Engineering, or Mathematics) programs. Actual majors and expertise vary and may
include Statistics, Financial Engineering, Mathematics, Computer Science, Finance, Physics, or other various STEM programs.

You can find more details on WebSim™ in Part IV of this book. Full Research Consultant program information is also available at WebSim’s™ official website.
PART I
Introduction
1

Introduction to Alpha Design

By Igor Tulchinsky

An alpha is a combination of mathematical expressions, computer source code, and configuration parameters that can be used, in combination with historical data, to make predictions about future movements of various financial instruments. An alpha is also a forecast of the return on each of the financial securities. An alpha is also a fundamentally based opinion. The three definitions are really equivalent. Alphas definitely exist, and we design and trade them. This is because even if markets are near-efficient, something has to make them so. Traders execute alpha signals, whether algorithmic, or fundamental. Such activity moves prices, pushing them towards efficiency point.

HOW ARE ALPHAS REPRESENTED?

An alpha can be represented as a matrix of securities and positions indexed by time. The value of the matrix corresponds to positions in that particular stock on that particular day. Positions in stock change daily; the daily changes are traded in the securities market. The alpha produces returns, and returns have variability. The ratio of return to standard deviation (variability) of the returns is the information ratio of the alpha. It so happens that the information ratio of the alpha is maximized when alpha stock positions are proportional to the forecasted return of that stock.

Expressions and Programs

Alphas can be represented by expressions consisting of variables or programs. Such expressions, or programs, are equivalent to each other, meaning one can always be converted to the other.
HOW DOES ONE DESIGN AN ALPHA BASED ON DATA?

It is simple. A price action is a response to some world event. This event is reflected in the data. If the data never changes then there is no alpha. Thus, it is changes in the data that have the information. A change in information should produce a change in the alpha.

Changes may be characterized in many ways as can be seen in Table 1.1.

<table>
<thead>
<tr>
<th>Table 1.1</th>
<th>Expression of changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A simple difference, A – B</strong></td>
<td>Example: today’s_price – yesterday’s_price</td>
</tr>
<tr>
<td><strong>A ratio, A/B</strong></td>
<td>Example: today’s_price/yesterday’s_price</td>
</tr>
<tr>
<td><strong>An expression</strong></td>
<td>Example: 1/today’s price. Increase position when price is low</td>
</tr>
</tbody>
</table>

All alpha design is the intelligent search of the space for all possible changes. An expression should express a hypothesis. Examples of this can be seen in Table 1.2.

<table>
<thead>
<tr>
<th>Table 1.2</th>
<th>Expressions expressed as a hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression</td>
<td>Hypothesis</td>
</tr>
<tr>
<td>1/price</td>
<td>Invest more if price is low</td>
</tr>
<tr>
<td>Price-delay (price,3)</td>
<td>Price moves in the direction of 3-day change</td>
</tr>
<tr>
<td>Price</td>
<td>High-priced stocks go higher</td>
</tr>
<tr>
<td>Correlation (price, delay(price,1))</td>
<td>Stocks that trend, outperform</td>
</tr>
<tr>
<td>(price/delay(price,3)) * rank(volume)</td>
<td>Trending stocks with increasing volume outperform</td>
</tr>
</tbody>
</table>

QUALITY OF AN ALPHA

An alpha is considered one of good quality when:

- The idea and expression is simple.
- The expression/code is elegant.
- It has good in-sample Sharpe.
- It is not sensitive to small changes in data and parameters.
• It works in multiple universes.
• It works in different regions.
• Its profit hits a recent new high.

ALGORITHM FOR FINDING ALPHAS

Repeat the below steps forever:

• Look at the variables in the data.
• Get an idea of the change you want to model.
• Come up with a mathematical expression that translates this change into stock position.
• Test the expression.
• If the result is favorable, submit the alpha.
An alpha is a model that predicts the prices of financial instruments. And while the idea of modeling the markets and predicting prices was not new back in the 1980s and 1990s, it was during that era that cheap computing power became a reality, making possible both (1) computational modeling on Wall Street trading desks, and (2) the generation and collection of data at a rate that is still growing exponentially as of the writing of this chapter. As computers and systematic data collection became ubiquitous, the need for innovative modeling techniques that could use these newly-created data became one of the drivers of the migration of PhDs to Wall Street. Finally, it was in this climate of technology evolution and exponential data production that the quantitative trading industry was born.

BACKGROUND

Quantitative trading and alpha research took off at the same time that cheap computational power became available on Wall Street. Alphas are predictions that are used as inputs in quantitative trading. Another way of putting it is to say that quantitative trading is the monetization of the alphas. Note that an alpha, as a form of prediction model, is not the same thing as a pure arbitrage. Sometimes the term statistical arbitrage is used to describe quantitative trading that exploits alphas.
Note that one could debate whether alphas ought to exist at all – some of the arguments for and against the existence of alphas can be made as part of an “efficient market hypothesis.” The financial economics academic literature tackles this problem exhaustively, qualifying the markets and the nature of information flow, and deriving conclusions based on various assumptions on the markets, the market participants and their level of rationality, and how the participants interact and process information.

That said, from a simple intuitive perspective, it makes sense that a very complex system such as the markets would exhibit some level of predictability. Whether these predictions can form the basis of exploitable opportunities is the argument that the quantitative trading industry is making every day, with more or less success.

**CHALLENGES**

Even if one can make an argument in favor of the existence of alphas under various stylized assumptions, the details of prediction in the real world are messy. A prediction with low accuracy, or a prediction that estimates a weak price change, may not be interesting from a practitioner’s perspective. The markets are an aggregate of people’s intentions, affected by changing technology, macro-economic reality, regulations, and wealth – which makes the business of prediction more challenging than meets the eye. Thus, to model the markets, one needs a strong understanding of the exogenous variables that affect the prices of financial instruments.

**THE LIFE-CYCLE OF ALPHAS**

A fundamental law of the markets is that any potentially profitable strategy attracts attention and attracts capital. Since the markets are a finite size, when more capital chases a strategy or employs a particular alpha, this implies that the fixed-sized pie that constituted the original opportunity needs to be sliced into multiple thinner slices. The end result is that, while alphas are born from the interaction of market participants, when they are (1) strong enough, (2) old enough, and (3) consistent
enough to be statistically validated and provide the basis for profitable trading strategies, they will begin to attract capital. This capital flow will ensure that the alpha will shrink and become more volatile, until there is so much capital chasing the idea that it will stop working. However, this process will affect the markets in ways that create other patterns, perpetuating the cycle of birth and death of alphas.

**DATA INPUT**

In order to predict the price movement of financial instruments, alphas need data. This data can be the prices themselves or a historical record of those prices. Most of the time, however, it helps to have more information than just the prices. For example, how many shares of a stock were traded, its volume, etc., can complement the historical price–time series.

A simple diagram to represent what an alpha is doing is as follows:

```
DATA (E.G. HISTORICAL PRICES) → ALPHA → PRICE PREDICTION
```

Note that data quality can have a large effect on the output of an alpha. So it’s important to evaluate data quality before it is used and address shortcomings then. Issues that may affect data quality can be technical, e.g. hardware problems, or related to human error, e.g. unexpected data format change, extra digits, etc.

**PREDICTIVE OUTPUT**

An alpha model’s output is typically a prediction. In many markets, it’s easier to predict the relative price of a financial instrument than it is to predict the absolute price of a financial instrument. Thus, in stocks, many alpha models predict the movement of the prices of various stocks relative to other similar stocks.

Typically, alphas are implemented using a programming language like C++, Python, or any other flexible and modern language. In larger organizations, a software environment developed in-house can abstract the alpha developer from many book-keeping and data management issues, letting the developer focus on creative research and modeling.
EVALUATION

What is a good alpha? What is a bad one? There is no single metric that will answer that question. In addition, the answer depends in part on how the alpha is going to be used. Certain investment strategies require very strong predictors; others benefit, marginally, from weak ones. Some pointers to alpha evaluation are:

• Good in-sample performance doesn’t guarantee good out-of-sample performance.
• Just like in academic statistics, outliers can ruin a model and lead to erroneous predictions.

It takes a lot of in-sample and out-of-sample testing to validate an idea. The more data one has, the more confidence one can have in an alpha. Conversely, the longer the period one considers, the more likely that the alpha will exhibit signs of decay and the more likely fundamental market changes will make the alpha unusable in the future. Thus, there is a natural tension between developing confidence in an alpha and its usefulness. One must strike the right balance.

LOOKING BACK

When developing alphas, one has the opportunity to look back in time and evaluate how certain predictive models would have performed historically. And, while evaluating an alpha with backtesting is invaluable (providing a window into both the markets and how the alpha would have performed), there are a few important points to remember:

• History never repeats itself exactly, ever. So while an alpha idea may look great on paper, there’s no guarantee it will continue to work in the future. This is because of the perverse power of computation, and the ability of creative modelers to miss the forest for the trees. With computational resources, one can evaluate a very large number of ideas and permutations of those ideas. But without the discipline to keep track of what ideas were tried, and without taking that into account when evaluating the likelihood of a model being a true model versus a statistical artifact only, one will end up mistaking lumps of coal for gold.