2022

BBM PUBLISHING

Roadmap

Quantitative Resources

Hello and welcome: I'm BBM, I run a hedge fund/ prop shop and head the research team/ PSC. This is a collection of resources I have compiled since I have been asked for it so many times. This is in no way a beginner's guide. It is for all and goes from basics to some really advanced materials. Send submissions for next edition!

3rd Edition

Contents (click links)

Chapter 1 Machine Learning and Algorithmic Trading (Textbooks)

Anything highlighted in red is optional since it is more of a repeat with extras of the textbook in black before it. Depends on how hard/ fast you want to learn! You should do the first textbooks then decide whether to do the Machine Learning Section or the Derivatives section first, but you can do them simultaneously. They have crossovers and I love both areas although I am more partial to the former, however they are very much independent and do not require knowledge from the other to learn. All textbooks highlighted in blue are links to the pirate-able version of the textbook.

Disclaimer:

I am not responsible if you commit piracy and I do not recommend you do this because it is wrong or whatever. Also buying the actual book means you get an impressive bookshelf/ some think it is better to read, but I enjoy both PDF and physical.

1) "Quantitative Trading 2nd edition"

1)a "Algorithmic Trading"

1)b "Machine Trading"

2) Trading and Exchanges

Note: Quick note on Ernest Chan Books. They aren't very meaty but are an easy intro so feel free to skim through them. Especially 1) a&b (in red) are very basic to the point where unless you are 100% new to quant they should be skipped.

ML (1/3) Machine Learning for Algorithmic Trading

There is a lot of overlap between machine learning… and mastering python… so start with one of them then read finding alphas then read the other. That is why they are noted as either the 1st or 3rd book to read in terms of machine learning.

ML (1/3) Mastering Python for Finance

Go do this all before doing finding alphas if you "audit" the course then it is free to do them all individually.

https://www.coursera.org/specializations/investment-management-python-machine-learning

- Note: This is a 4 part course so there is certainly a lot to go through, but I think it is one of the best resources because it uses legacy models to build intuition, but unlike most courses then goes on to show you some actual methods that are used and work in the industry. Another BIG benefit is that it uses Python in Jupyter notebook which in my opinion is the best way to do research. Orange is good as well and an R kernel in Jupyter is also a nice alternative (more on that later).

ML (2) Finding Alphas: A Quantitative Approach (2nd edition)

ML (4) Advanced Algorithmic Trading

ML (5) Advances in financial machine learning

ML(6) The elements of statistical learning -general ML knowledge skip if you know ML already

Also, may be useful to go do the deep learning and machine learning courses on Coursera, but of course it won't be finance focused just building a general understanding of what things are and how they work.

Bonus Book :

https://libgen.st/book/index.php?md5=F5D03C3D452EAB8A5F98D141EF183A06

Chapter 2 Derivatives, and Volatility Trading (Textbooks)

Derivatives (1) Hull Options Futures and other Derivatives

Derivatives (1 Alternative) Option Trading & Volatility Trading (both textbooks by Euan Sinclair)

Derivatives (2) Positional Option Trading

Derivatives (3) Trading Volatility

Extra Derivatives:

Option Volatility and Pricing: Advanced Trading Strategies and Techniques

Trading Options Greeks: How Time, Volatility, and Other Pricing Factors Drive Profits, Second Edition

Chapter 3 YouTube Videos

Here are some great videos by Ben Felix. I can honestly recommend all of his videos but these grasp at key point all traders need. Very asset pricing model/ EMH based and whilst I go for EV (Expected Value) it is still important to know. In the podcasts section Vivek Viswanathan on Flirting with Models gives a good explanation of how EV models can work with factor investing and what is wrong/ right about factor models.

https://www.youtube.com/watch?v=jKWbW7Wgm0w
https://www.youtube.com/watch?v=foqswJT3Spc
https://www.youtube.com/watch?v=yco0sC7AJ2U
https://www.youtube.com/watch?v=IzK5x3LlsUU

LEARN VOLATILITY

Patrick Boyle has some great books, but I also recommend his playlists. Especially the last 3 rows, which is a full education in derivatives, and he breaks down financial news in a meaningful and educational way that is fun to watch, but without the narrative:

https://www.youtube.com/c/PatrickBoyleOnFinance/playlists

Leonardo Valencia (Some really great volatility videos, I recommend you watch them)

Tasty Trade (I like their Greek videos, but still prefer Patrick Boyle)

KeyPaganRush (Great visualizations for Charm and Vanna)

LEARN ALGOTRADING

Part time Larry:

Crypto Wizards:

Quantconnect YT channel more on them later

Quantra

Trade options with me (teaches Quantconnect)

Jacob Amaral (Nothing too remarkable, but a few decent trading algo vids)

Algo vibes: (great for people who are just starting)

Note: Algovibes is probably one of my top recommendations. He demonstrates how to build a full system on many occasions and makes it as simple as possible plus also how to research in Jupyter notebook.

LEARN SIGNAL PROCESSING

Mike X Cohen (Signal Processing applied to neuroscience, but still great)

Two great playlists by him ^^^^^^^^^^^^ (good alpha here)

https://youtube.com/playlist?list=PLn0OLiymPak2jxGCbWrcgmXUtt9Lbjj_A

https://youtube.com/playlist?list=PLn0OLiymPak2G__qvavn3T8k7R8ssKxVr

Abhishek Agrawal (signal processing stuff in python)
Esther Explains: (Some cool signal processing stuff)

LEARN NICHE QUANT TOPICS

AP Monitor (optimization, very niche/complex, but I like it)

Niche optimization channel (again I like it)

Niche and complex math channel lol

AI Quant (Extremely complex stuff, but holy shit if you understand it, it's GOLD)

H&T (mainly interns who produce no original work, but still exposes you to high quality material)

LEARN DATA SCIENCE/GENERAL CODING

Neural Nine is more general Python, but has some good algotrading vids

Data Science Dojo (great data science stuff)

Ken Gee (general data science)

Coding Jesus (Don't take his career advice, but he knows how to code/ trade well!)

Finn Eggers (java DL stuff)

Keith Galli (good python tutorial)

Gerard Taylor (specifically I recommend his ML in C++ course)

Data Professor (Just general data science):

LEARN STATS/DS MATH

Luis Serrano: Good statistics videos

Complex math channel lol

More easy data science/ stat videos

Stat Quest (really engaging guy for complex stat topics broken down so anyone can get it)

OTHERS

Another GOAT podcast (RCM Alternatives)

Martin Shkreli (The GOAT himself, good for learning how DCFs work (they are a shit model, but they are used loads, so it is good to know how the other half lives))

Mutiny (Listen to every single podcast they have; you won't regret it)

Chapter 4 Courses

On Coursera Robert Shiller has a course called Financial Markets. It is free without the certificate $50 for the certificate. The videos are also on YouTube. This is an amazing start for finance and the markets in general and will teach you the basics of everything in the markets. Coursera link below:

https://www.coursera.org/learn/financial-markets-global

Medium to Higher level:

Andrew Ng has a course on Machine Learning and Deep Learning on Coursera. Those are really good but quite math heavy.

For the math Imperial College London has a Mathematics for Machine Learning course series and it has multivariate calculus, linear algebra and PCA. All of which will be super helpful.

There is also a Financial Engineering Part 1 & 2 course on Coursera as well.

This is also a great one for machine learning in python and has some really great strategies included in there:

https://www.coursera.org/specializations/investment-management-python-machine-learning#courses

https://www.coursera.org/projects/intro-time-series-analysis-in-r

This is a great project you can do in R. Amazing stuff 100% recommend. I really do stress that this is a great resource.

Chapter 5 Podcasts

(One of the best resources DO NOT TAKE LIGHTLY)

Most information especially the most useful is not in textbooks so you need to religiously study podcasts. Think of textbooks as foundational but read them to understand what is talked about on podcasts.

First of all, this one is amazing I have listened to some episodes 6+ times.

Flirting with models

Also, amazing and you NEED to listen to all the episodes

Mutiny fund (YouTube channel shown)

RCM Alternatives, the derivative

Market Champions

Chapter 6 Trading Platforms & Brokerage Firms

I recommend IBKR, but TD is also good. In terms of language, I would say Python is best, but it is up to you.

Quantconnect is hard to learn at first but will take care of server hosting and data which is super annoying and also provides great resources. Handles the hard stuff at first, but also is hard to learn kind of.

Chapter 7 High Frequency Trading & Market Making

Here is a large dump of resource for HFT. In terms of textbooks, I recommend High-Frequency Trading: A practical guide to algorithmic strategies and trading systems 2nd edition.

See the light reading section, but Flash Boys is amazing for HFT (basically the book that made HFT well known) and so are these books (this overlaps with light reading, but I wanted to highlight):

Dark Pools

Broken Markets

The Problem of HFT

Flash Boys: Not So Fast – This is an insiders review of Flash Boys and is really great

Trading at the speed of light

For learning how market makers function:

Empirical Market Microstructure

List of links:

https://sniperinmahwah.wordpress.com/2014/09/22/hft-in-my-backyard-part-i/
https://sudonull.com/post/93403-Online-Algorithms-in-High-Frequency-Trading-Problems-of-Competition-ITI-Capital-Blog
https://www.youtube.com/watch?v=AS7HLtErlI8
Avellaneda strategy: A technical deep dive
https://hummingbot.io/blog/2021-04-avellaneda-tech-deepdown
A comprehensive guide to Avellaneda & Stoikov's market-making strategy
https://hummingbot.io/blog/2021-04-avellaneda-stoikov-market-making-strategy
https://www.youtube.com/playlist?list=PL2F82ECDF8BB71B0C
https://medium.com/@eliquinox
https://www.youtube.com/watch?v=XgFzHXOk8IQ&list=PLQnljOFTspQUGjfGdg8UvL3D_K9ACL6Qh&index=9
https://alexabosi.wordpress.com/2014/08/28/limit-order-book-implementation-for-low-latency-trading-in-c/
https://github.com/rubik/lobster
https://www.youtube.com/playlist?list=PL5Q2soXY2Zi_FRrloMa2fUYWPGiZUBQo2
https://www.youtube.com/watch?v=_OJmxi4-twY
https://www.youtube.com/watch?v=NmarI5ErisE
https://www.youtube.com/watch?v=9nuAjYRbITQ
https://t.co/Rcnw26Bzyr?amp=1
https://web.archive.org/web/20110219163448/http://howtohft.wordpress.com/2011/02/15/how-to-build-a-fast-limit-order-book/
https://www.guru99.com/os-tutorial.html
https://github.com/theopenstreet/VPIN_HFT
https://github.com/hudson-and-thames/mlfinlab/blob/master/mlfinlab/data_structures/imbalance_data_structures.py
https://drive.google.com/file/d/0B4pk0Nap6TZLNTBhblRHcUJUVmM/view?resourcekey=0-G3T886oNA-ZXtLQE7tKZDA

INFRA:

https://medium.com/prooftrading/proof-engineering-the-algorithmic-trading-platform-b9c2f195433d#a1f7
https://medium.com/prooftrading/selecting-a-database-for-an-algorithmic-trading-system-2d25f9648d02
https://www.linkedin.com/in/silahian/detail/recent-activity/posts/
http://www.caravaggioinbinary.com/HFT-Simulation-Lab/
https://rickyhan.com/jekyll/update/2019/12/22/how-to-simulate-market-microstructure.html
https://youtu.be/b1e4t2k2KJY
https://medium.com/prooftrading/the-trading-strategy-63183bd231cd

https://medium.com/prooftrading
https://mattgosden.medium.com/tutorial-using-pythons-unsync-library-to-make-an-asynchronous-trading-bot-9ee2ae881272
https://www.youtube.com/watch?v=SOTamWNgDKc

Strats and Backtesting:

https://towardsdatascience.com/application-of-gradient-boosting-in-order-book-modeling-3cd5f71575a7
http://jonathankinlay.com/2021/05/machine-learning-based-statistical-arbitrage/
https://letianzj.github.io/cointegration-pairs-trading.html
https://hudsonthames.org/caveats-in-calibrating-the-ou-process/
https://www.youtube.com/playlist?list=PLv-cA-4O3y95J6xmwSaCILL4FlGJZO0PJ
https://teddykoker.com/2019/05/momentum-strategy-from-stocks-on-the-move-in-python/

Programming:

https://www.youtube.com/watch?v=NH1Tta7purM
https://www.youtube.com/watch?v=pBKwWI56uXc

Market Data:

https://cdn.tun.to/minute/
http://www.kibot.com/free_historical_data.aspx
https://www.dukascopy.com/swiss/english/marketwatch/

3rd Edition Extras:

https://www.youtube.com/c/dYdXprotocol/videos
https://medium.com/open-crypto-market-data-initiative/simplified-avellaneda-stoikov-market-making-608b9d437403
https://www.tastytrade.com/shows/geeks-on-parade/episodes/market-making-with-shelly-geeks-2019-07-19-2019
https://quant.stackexchange.com/questions/36073/how-does-one-calibrate-lambda-in-a-avellaneda-stoikov-market-making-problem
http://proceedings.mlr.press/v128/wisniewski20a.html
https://github.com/valeman/awesome-conformal-prediction
https://medium.com/prooftrading/building-a-high-performance-trading-system-in-the-cloud-341db21be100

Algorithmic and Advanced Programming in Python:

Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 1
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3953589
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 2 https://t.co/uq4I7V4oa3
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 3 https://t.co/1ojWwr2SPO
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 4 https://t.co/Su7OZD1SpU
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 6 https://t.co/hJoREZaOYc
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 7 https://t.co/TBs30cXoND
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 8 https://t.co/kGhjQZMBYX
Algorithmic and Advanced Programming in Python - Syllabus in Computer Science, Decision Making & Data - Masterclass 9 https://t.co/FoSCRn4B5V

Chapter 8 Additional Volatility/Derivatives Resources

This is just one link, but it contains many links within so don't treat it lightly. This is a top recommendation.

https://notion.moontowermeta.com/select-content-from-the-quant-and-vol-community

Options Starter Pack

https://notion.moontowermeta.com/options-starter-pack

Even more of me blatantly copying and pasting Kris's recommendations/ resources, but they're really good

https://notion.moontowermeta.com/moontower-content-by-kris-abdelmessih

Wilmott forum:

I cannot highlight how great of a resource this is. I think it may be one of the only resources where a genuine discussion for pricing exotic derivatives etc can be found. There are some alpha bits in there as well, but in the technical section of the forum there are discussions on models from pricing everything from vanilla European options to Bermudian swaptions, and better yet this is by people who work in the industry.

https://forum.wilmott.com/?_ga=2.79159636.1304736795.1642875642-1589239950.1642875641&_gl=1*3qlrhp*_ga*MTU4OTIzOTk1MC4xNjQyODc1NjQx*_ga_51FRCD57RP*MTY0Mjg3NTYzNy4xLjEuMTY0Mjg3NTY0OS4w

Nuclear Phynance (for some strange reason Wilmott and NP users hate each other) is a great resource. Not so much derivatives only, both NP and Wilmott are diverse, but Wilmott is mainly derivatives. This is an awesome server.

https://nuclearphynance.com/

Chapter 9 Coding Languages Review and Resources

This is mainly opinion, but for research you should know either R or Python very well (I recommend Python as there are more resources) although when you become more advanced basic Python will not be fast enough so you will need to learn advanced Python and Cython (using C/C++ through Python). So here is a textbook for that:

High Performance Python

If you want to do High Frequency Trading (HFT) or Market Making (MM) you will need to learn C/C++ because it is the fastest. Hard to learn, so don't start with it, but rewarding.

In terms of research languages some quants will use R or MATLAB specifically because it has lots of statistical functions that are optimized for data analysis. Ernest Chan loves MATLAB but in reality, it isn't very good so stick with R if you want to learn a 2nd/3rd language for researching math heavy topics specifically. Otherwise, Python should fill the role.

Anaconda (100% free) is an easy way to install R and Python and comes with Jupyter notebook and Spyder (I like Spyder as an IDE although this is a big debate), but Jupyter is without fail the best for working in for research, and both R and Python can be used in it.

If you are hopeless at programming, or just want to go fast then Orange (comes with Anaconda) uses a graphical user interface and requires no programming. It uses scikit learn models which can easily be implemented when you do pick up some code as well (for implementation for example). Really great stuff with loads of models, and you can run your own Python scripts in it if the model is not available. There are external packages to install like time series models and NLP specifically, but the standard library is great as well.

Below I have added some resources for learning algorithmic trading skills in alternative coding languages as well as textbooks for generally becoming good at these languages, but these are quite advanced and are for later in your journey. Focus on the textbooks in the start of the roadmap:

High Performance C++ (Textbook)(ADVANCED)

Machine Learning in C++ YouTube Playlist/Course

Machine Learning with C++ (Textbook)

Learning Quantitative Finance with R

Modern Data Mining in C++ and CUDA C

Statistically sound machine learning for algorithmic trading… (Textbook)(C++)

Testing and tuning market trading systems: algorithms in C++ (Textbook)

Chapter 10 Projects

One of the best ways to learn is to find a project. As mentioned before algovibes provides some great projects. Below is a previous post by me regarding the process to be taken to generate a successful ML system through R&D. Most of the work is feature engineering.

A robust + successful + well defined algotrading process:

1) Data - Clean, Accurate, Formatted, and Lots of Data

2) Features - Anything that predicts returns! (can be another models prediction, a filter (Kalman/MA/etc), an indicator (RSI/MACD/etc), a signal (if RSI > 80: 1, if RSI < 20: -1, else: 0).

3) Combination - PCA, Autoencoding, and more! (use some intuition with this, PCA is linear so maybe do that with some momentum indicators and do the momentum indicators together, on the reverse you may want something more non-linear (Autoencoder?) for mean reversion / seasonality / non-linear features

4) Specific Modelling - hierarchical modelling, either time or cluster based. (cluster based might be: model momentum indicators together with linear regressions or something that suits them, or wave indicators with LSTMs or Wavenet which suits them!) & (time based is using models for different timeframes, short term doesn't give AF about financials, but long term does so maybe model separately, on this note the longer you go the more linear your models should be)

5) General Modelling - finally combine everything

6) Risk management/ Regime shift/ Position sizing/ Meta-Labeling - this is a whole lecture in itself (modified kelly for position sizing, regime is Hierarchal HMM or/and GMM, meta-labelelling is just rinse and repeat, (modelling your model btw))

TA, signal processing (search through Wikipedia there is some great content on it), standard price features, microstructural features, entropy features, and data mining are all great ways to generate features. One note on data mining is that it is very easy to overfit so use ridge regressions and check for T-values before incorporating them.

I am also available to hand out research papers and areas to delve into for project ideas.

This is a long list of projects. Have fun checking them all out.

Subject	Structure	Method	Information	URL
Deep Learning of Small Portfolios for Index Tracking	Deep learning of sparse autoencoder (SAE) models	Topology reshaping techniques which help to identify less complex models that capture the essential structure in the data	1) the sparsification techniques force it during training to generalize better, and 2) the robustification using a heavy-tailed noise model diminishes the effect of outliers	https://www.linkedin.com/pulse/deep-learning-small-portfolios-index-tracking-nikolay-nikolaev/
Algorithmic Cryptocurrency Trading with Sharpe-optimal Deep Learning	Sharpe-optimal Connectionist Learning (SCL), Recurrent Reinforcement Learning (RRL)	Recurring Nonlinear Pattern	directional fluctuations in the given series which are sufficient for sending trading signals.	https://www.linkedin.com/pulse/algorithmic-cryptocurrency-trading-sharpe-optimal-deep-nikolaev/
Universal Cointegration for Pairs Trading via Machine Learning	CCM(kNN) - Cross Covergent Mapping (k cluster Nearest Neighbors)	Estimated Price by Pair Nonlinear Modeling(Estimated Single Asset Price by Nonlinear modeling)	Linear combination produces a stationary residual difference series (also called spread).	https://www.linkedin.com/pulse/universal-cointegration-pairs-trading-via-machine-nikolay-nikolaev/?trackingId=nKswdHj1R%2FagxOErDCoMSg%3D%3D
Local Prediction of Mid-Prices in Limit Order Book Markets	kNN Regression	Arrival Time(kNN Regression)	Order Book Mid Price Arrival time	https://www.linkedin.com/pulse/local-prediction-mid-prices-limit-order-book-markets-nikolay-nikolaev/
Machine Learning with Dynamic Time Patterns for Algorithmic Trading	pdf(kNN), CDF(kNN),Adaptive Shape Distance (ASD) [3], Complexity Invariance Distance (CID) [4], Dynamic Time Warping (DTW)	Nonparametric forecasting of time series using dynamic time patterns	Local Density kNN	https://www.linkedin.com/pulse/machine-learning-dynamic-time-patterns-algorithmic-trading-nikolaev/
Nonparametric Machine Learning for Algorithmic Cryptocurrency Trading	Self-adaptive Local Learning Machine (SLLM)	Adaptive repeating nonlinear patterns by proportional degree of similarity reflecting the essential characteristics of the data trajectory	Essential characteristics of the data trajectory	https://www.linkedin.com/pulse/nonparametric-machine-learning-algorithmic-trading-nikolay-nikolaev/
Overfitting Avoidance in Portfolio Construction using Probabilistic	Multivariate General Regression Neural Network (MGRNN)	Network Architecture	Proportionally density of the effective range (stock spread)	https://www.linkedin.com/pulse/overfitting-avoidance-portfolio-construction-using-neural-nikolaev/

Neural Networks				
Generative Adversarial Networks for Machine Learning of Constrained Portfolios	Generative Adversarial Network (GAN)	Generator network and a discriminator network that play an adversarial game	Closeness between the utilities obtained with sampled weights	https://www.linkedin.com/pulse/generative-adversarial-networks-machine-learning-nikolay-nikolaev/
Deep Neural Networks for Prediction-based Portfolio Construction	Robust DNN (RDNN)	Nonlinearity	Rectified Units/Gradient Convergence Speed	https://www.linkedin.com/pulse/deep-neural-networks-prediction-based-portfolio-nikolay-nikolaev/
Regime-based Machine Learning of Green Stock Portfolios	Markov Regime-Switching (MRS) model	Volatility based regime switching	Portfolio weights are calculated dependent on the particular regime, that is the portfolio is conditioned on the regimes.	https://www.linkedin.com/pulse/regime-based-machine-learning-green-stock-portfolios-nikolay-nikolaev/
Efficient Direct Reinforcement Learning of Low-risk Portfolios	Connectionist Reinforcement Learning machine for efficient computation of Online Portfolios (CRLOP)	Multi-output feedforward neural network	Risk-averse investment strategy as a sequence of allocations and re-balancing decisions	https://www.linkedin.com/pulse/efficient-direct-reinforcement-learning-low-risk-nikolay-nikolaev/
Deep Learning of Dynamic Factor Models for Asset Pricing	Deep Dynamic Factor Model (DDFM)	Deep network structure, Neural network pruning technique, using a cardinality parameter to control the degree of sparseness	asset returns from inferred factor realizations, more precisely, the factors are forecasted arrangements of individual asset contributions to the overall portfolio.	https://www.linkedin.com/pulse/deep-learning-dynamic-factor-models-asset-pricing-nikolay-nikolaev/
Support Vector Machine Learning of Sparse Portfolios	epsilon Support Vector Regression (e-SVR) machine	Sparse regression algorithm	Two hyperparameters: the first determines the size of the sensitivity, and the second determines its influence on the magnitude of the fitting error.	https://www.linkedin.com/pulse/support-vector-machine-learning-sparse-portfolios-nikolay-nikolaev/
Online Portfolio Trading by Dynamic Reinforcement Learning	Reinforcement Learning machine for Online Portfolio trading (RLOP)	Recurrent neural network	Momentum and adaptive learning rate for switching between passive and aggressive updating depending on the recent profitability.	https://www.linkedin.com/pulse/online-portfolio-trading-dynamic-reinforcement-nikolay-nikolaev/
Building Smart Beta Portfolios with Large-Scale Machine Learning	Smart Beta portfolios, like the Most Diversified Portfolio (MDP), Risk Parity Portfolio (RPP)	Alternating Direction Method of Multipliers (ADMM)	Statistical characteristics and econometric performanceof the proposed ADMM-based tool for creating Smart Beta portfolios	https://www.linkedin.com/pulse/building-smart-beta-portfolios-large-scale-machine-nikolay-nikolaev/
Bayesian Machine Learning for Robust On-line Portfolio Selection	Bayesian Robust Online Portfolio Selection (BROPS)	Ornstein–Uhlenbeck stochastic differential equation	Heavy-tailed models of returns on prices (based on an approximation of the Student-t density by an infinite mixture of Gaussians)	https://www.linkedin.com/pulse/bayesian-machine-learning-robust-on-line-portfolio-nikolay-nikolaev/
Efficient Computation of Sparse	Most Diversified Portfolio (MDP), Equal Risk	Nonlinear programming algorithms	Maximize the ratio between the weighted average volatility of the	https://www.linkedin.com/pulse/efficient-computation-sparse-risk-based-portfolios-using-nikolaev/

Risk-based Portfolios using Machine Learning	Contribution (ERC)		assets and the total portfolio volatility	
Robust Portfolio Optimization via Connectionist Machine Learning	Mean-Variance Portfolio (MVP), Connectionist Optimization Machine (COM)	Quadratic Programming (QP)	Optimized Mean and Minimum STD, Convergence Rate (Speed)	https://www.linkedin.com/pulse/robust-portfolio-optimization-via-connectionist-machine-nikolaev/
Deep Cleaning of Covariance Matrices for Portfolio Allocation	Covariance Matrix	Autoencoder Machine (AEM)	denoised versions of the eigenvectors of the covariance matrix which help to recover its genuine structure	https://www.linkedin.com/pulse/deep-cleaning-covariance-matrices-portfolio-nikolay-nikolaev/
Finding Structure in the Co-movement of Stock Prices via Adaptive Metric	Centroid	K-means algorithm and the Self-Organizing Map (SOM) networks	mean-reverting eigenportfolio with the stocks from each cluster	https://www.linkedin.com/pulse/finding-structure-co-movement-stock-prices-via-metric-nikolaev/
Deep Learning Autoencoders for Building Principal Component Portfolios	Portfolio arbitrage	Principal Component Analysis (PCA) and the Autoencoders (AE), Variational Bayesian inference	eigenportfolio as a linear combination of all stocks which are allocated contributions according to their corresponding coefficients in the first principal component	https://www.linkedin.com/pulse/deep-learning-autoencoders-building-principal-nikolay-nikolaev/
Machine Learning of Heavy-Tailed Dynamic Spread Models for Statistical Arbitrage	Robust Dynamic Mixture Models	Statistical arbitrage (statarb)	mean-reverting spreads based on a discrete version of the Ornstein–Uhlenbeck stochastic differential equation, Dynamic Hedge Ratios	https://www.linkedin.com/pulse/machine-learning-heavy-tailed-dynamic-spread-models-nikolay-nikolaev/
Self-Tuning Local Learning Machines for Prediction of Stock Market Returns	Self-tuning Local Learning Machine (SLLM)	Local Learning machines (LLMs) based on the nearest neighbours approach	autocorrelations,behaviour patterns	https://www.linkedin.com/pulse/self-tuning-local-learning-machines-prediction-stock-market-nikolaev/
Deep Learning of Heteroskedastic Volatility Models for Risk Estimation	Dynamic heteroskedastic volatility models (D2GARCH)	Deep connectionist structure by unrolling the network in time.	Volatility rank and probability, time relationships in the data	https://www.linkedin.com/pulse/deep-learning-heteroskedastic-volatility-models-risk-nikolay-nikolaev/
Deep-Memory Networks vs. Deep Learning Networks for Stock Market Prediction	Deep-Memory Neural Networks (DMNN)	Dynamical system using sequential Bayesian estimation to accommodate properly the temporal dimension.	Feature Extraction -> long-term dependencies via memory feedbacks and gates for behavioural control	https://www.linkedin.com/pulse/deep-memory-networks-vs-deep-learning-stock-market-nikolay-nikolaev/

Chapter 11 Data

If you ask for it, I can usually provide the historical data. News/ Google Searches/ Fundamental/ Prices/ Meta/ Dividends/ Splits/ Quotes all the way up to EOD data from the 1960s. If you are looking for a more established data source:

Tiingo - $10/mo. Cheap, but the data is only fundamental, EOD, minute data, quotes, news/ NLP data (in fact Tiingo is the best for news feeds for NLP). Live data as well.

Binance – FREE. Only crypto data and the historical data is of course limited to aggregates historically and only a brief window if you want historical sub-minute data. But I have scraped quotes so ask and I can provide.

IBKR - $10-30. Super cheap and you can basically get all the data you could ever want out of it but you need an account and $500 deposited with them since they're a broker (my broker recommendation btw). The API is slow as shit so if you want to download their entire options data library you better make a scraper in AWS Cloud because it will literally take over a month. (AWS SageMaker Jupyter notebooks are an easy way to scrape data without needing to set up servers using anything technical, this method can also be done for live hosting trading algorithms)

Links: (Over a TB of data, worth a fortune, but handed out here for free!)

Numerai uses community sources alpha for running it's fund and gives out loads of free data

https://numer.ai/

G-Research Crypto gives out data as well, but don't submit code. Read the legal docs you are giving them it basically. Blatant code grab. (God this data is awful. I think it was broken on purpose to make the challenge harder)

https://www.kaggle.com/c/g-research-crypto-forecasting/data

Literally 200GB at least… Slow data downloads though

https://b.scsi.to/

22GB of IQFeed Data

https://mega.nz/folder/HUQzDCgK#rc45NqXhRA8SFgK1l2MYcw

1.2GB of CompuStat data.

https://mega.nz/file/6IwnQKQL#Xb1PQja8veVCRWy7nJ_o45ZKeDyDy4IYV7QAHQnv7A4

Cryptocurrency data for Bittmex

https://www.kaggle.com/tencars/392-crypto-currency-pairs-at-minute-resolution

3k stocks, decade+ of financial data

https://www.kaggle.com/miguelaenlle/parsed-sec-10q-filings-since-2006

More data. Also not great, but hey anything helps, and it can be fixed of course

https://www.kaggle.com/c/ubiquant-market-prediction/data

Loads of NLP data. Headlines and analyst reports mainly

https://www.kaggle.com/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests

Google search data

https://www.kaggle.com/miguelaenlle/google-trends-history-for-4000-stocks

Optiver IV prediction data

https://www.kaggle.com/c/optiver-realized-volatility-prediction/data

Chapter 12 GitHub Repositories

One of the best ways to find good examples are on GitHub and one of the best repositories for algorithmic trading is the one that accompanies the machine learning for algorithmic trading textbook referenced earlier.

https://github.com/stefan-jansen/machine-learning-for-trading

Another repository that comes from a textbook is the repository that comes from Mastering Python for Finance. The models in the textbook are quite good especially LSTAR models which aren't usually in time series courses, but are great models.

https://github.com/jamesmawm/Mastering-Python-for-Finance-source-codes

This is a github repository that links to other notebooks and has quantitative resources in itself. It has a lot more risk models/ pricing models especially compared to the first repository referenced which is purely about generating alpha, and still has loads of purely alpha based model so is loaded with resources.

https://github.com/letianzj/QuantResearch

This github provides some basic examples toward applying signal processing in Python which can be used as features in the feature engineering process rather successfully.

https://github.com/SparkAbhi/SignalProcessingWithPython

This is an interesting project that applies one of the most important pairs trading papers in the literature in a detailed manner with up to date code examples as well. The use of PCA to generate multivariate portfolios for both portfolio optimization and mean reversion trading is a key advancement here.

https://github.com/alexdai186/Eigenportfolios

Generating features/ finding examples of great features to use is always a good thing to have so the next two repositories give great examples of basic feature engineering. The second one doesn't make as good features as these are only basic price features, but it shows how to use PySpark which is used for big data applications such as with HFT data (a couple months of quote data can be 50GB compressed -> ½ TB uncompressed and 50TB if you engineer 500 features, so distributed computing is needed!)

https://github.com/hjeffreywang/Stock_feature_engineering/blob/master/Feature_generation.ipynb

https://github.com/MiaDor12/Advanced_Feature_Engineering_of_Raw_Data_of_Stocks-with_PySpark/blob/master/Advanced%20feature%20engineering%20with%20pyspark%20on%20raw%20data%20of%20stocks.ipynb

Here is another example of good feature engineering and the use of fractional differencing to make the data stationary which then lets you use models that assume stationarity such as FFT (Fast Fourier Transform) although there are non-stationary signal processing models as well. More details are in a thread I wrote on this subject on twitter

https://github.com/alexbotsula/Price_direction_forecast

Continuing my point on signal processing one of the best models is an EMD (more on twitter) and this is an awesome repository that applies it.

https://github.com/Cy743652/CEEMDAN-LSTM/tree/main/Code

Here is a github full of microstructural models for high frequency trading (the next few repos will be HFT/MM).

https://github.com/gjimzhou/MTH9879-Market-Microstructure-Models

VPIN is an important model for market making and is one of the latest models in the literature so here is an implementation repository.

https://github.com/jheusser/vpin

High frequency trading statistical arbitrage example github repository.

https://github.com/clfrenchgit/gdax-bot

HFT using DL models for statistical arbitrage.

https://github.com/scibrokes/real-time-fxcm

This one isn't quite github but is a great resource for finding example C++ code for developing low latency C++ systems.

http://dlib.net/

This is a great example of an implementation that genuinely has alpha

https://www.kaggle.com/yamqwe/1st-place-of-jane-street-adapted-to-crypto

Find this table in the notebook and you can view all the other great notebooks in the series. If you can manage to compile this into one system and utilize other resources in this guide you will find alpha.

All notebooks in the series 👇

CV + Model	Hyperparam Optimization	Time Series Models	Feature Engineering
Neural Network Starter	MLP + AE	LSTM	Technical Analysis #1
LightGBM Starter	LightGBM	Wavenet	Technical Analysis #2
Catboost Starter	Catboost	Multivariate-Transformer [written from scratch]	Time Series Agg
XGBoost Starter	XGboost	N-BEATS	Neutralization
Supervised AE [Janestreet 1st]	Supervised AE [Janestreet 1st]	DeepAR	Quant's Volatility Features
Transformer)	Transformer		⏳Target Engineering
TabNet Starter			⏳Fourier Analysis
Reinforcement Learning (PPO) Starter			⏳Wavelets

These are more just other people's resources, but they are all in drives and I enjoyed most of them, but obviously I prefer my own resources as I have vetted them more.

https://github.com/beimingmaster/quant-resources

Chapter 13 Light Reading

This section includes books that are not quite textbooks but build a general knowledge of how the industry works. This is good for showing you know the industry well in interviews, and just really great common sense in modelling.

Liars Poker

Flash Boys (This and Dark Pools are a great resource of understanding HFT)

Irrational Exuberance

The intelligent investor

Pragmatic capitalism (there is a great list of books to read at the end of the book)

The black swan – tail risk

Fooled by randomness – tail risk/ understanding randomness

The quants (Scott Patterson) (so much common sense, and lessons in this book)

Dark Pools (Scott Patterson)

More money than god – A full history of hedge funds

Black edge – Steve Cohen Biography. What a legend

The man who solved the markets – Great book but here is a summary of the lessons from a friend (but do read the book as well, it's interesting)

A man for All Markets: Beating the Odds, from Las Vegas to Wall Street

The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin, and Reward

When Genius Failed – A very important lesson emphasized by the quants

A Random Walk Down Wall Street

Billion Dollar Whale

The Predators' Ball

Broken Markets

The Problem of HFT

Flash Boys: Not So Fast – This is an insiders review of Flash Boys and is really great

Trading at the speed of light

Investing for adults – Easily the greatest book for passive investing (It's a mini-series, but basically as long as a single book)

The rise of carry

Chapter 14 Careers

The usual career path:

1. Attend Target University (this means top university)
2. President of Finance Club / Student Consultancy / Interest Shown
3. University Internships at Investment Banks
4. Spend at minimum 3 months memorizing every question in a quant interview question textbook and learning every brain teaser you can find.
5. Cold emails/ coffee chats/ networking
6. Get an analyst role at an investment bank
7. After 2 years transfer to a hedge fund/ prop shop

The first thing to make this better is to get a top internship in high school. This is hard, but resources are provided below. Once an internship is obtained it is referred to as a conveyor belt because it becomes far easier to get another. 85% of those with internships come back. When you receive a summer internship do not wear a Rolex, Gucci sleds, etc. You are there to wear a plain Casio, always leave later than your boss, and get that return offer.

Once you get the analyst role it is not done. This is basically hell on earth and is designed to filter down people. You will spend every hour of the week working or meeting your basic needs. You can do the math, but anything over 80 hours a week comes out of your sleep. You will work 70-110 hours… Going straight to a hedge fund or prop shop is the best route as such but requires years of religious studying to go straight to "buy-side".

The reason everyone goes to "sell-side", or investment banking beforehand is because it is the best training ground. I have done my best to provide resources and I do believe a

motivated person could learn the content without going to "sell-side", but it is a training ground and sometimes analysts leave after just 6 months even to go to buy-side.

One resource is Wall Street Oasis. Just as a warning you will get shit on for not having a GS (Goldman Sachs) internship in pre-school because everyone is a massive HARDO (finance term for those who treat this like they're in the NBA: eat, sleep, quant), but the career advice is great.

Watch Alpesh Patel on Tiktok (@greatinvestments) and check out the "internship" he offers. They are a 100m+ fund and offer an open internship. It's basically a course and you won't learn much, but anyone can do it and it will +1000 points your resume.

Same deal with "theforage.com". You can get virtual GS, and JPMC internships (no application/rejection it's open to all) which aren't really actual ones, but hell they're the one calling it one so smack it on the resume. +1000 points.

Here is a YouTube channel all about investment banking and hedge funds/ private equity. This is coming from the non-quantitative side of things and probably refers to Macro or ELS (Equity Long/Short) funds more than quant funds and is also more M&A (Mergers and Acquisitions) than S&T (Sales and Trading), but that doesn't really matter because the recruiting processes are basically the same. Some really good videos about coffee chats. The fact is that you will be sending 1000s of cold emails/LinkedIn messages. Kris Sidial would follow senior people to work and give them his resume, and he's a volatility legend now. Don't be embarrassed to do this because otherwise someone else will. The competition is massive so the process for recruiting is ruthless. The usual process is:

- Cold email/LinkedIn/twitter message (Be thoughtful there are guides in the YT channel for all of these parts btw)
- Attempt to get a call. This is your chance to shine and why you've been reading piles of textbooks. You need to be impressive because otherwise they won't want the next step
- Coffee chat. Try and get an in person meeting. You should be subtle about it but once it is going well ask about internships. The entire product of all this effort is to get a single comment from them to HR in the break room along the lines of "If you see an application from xx, he really knows what he's talking about and is really interested".

https://www.youtube.com/c/PeakFrameworks

Another tip I will give is that showing that you can learn fast and are willing to put in those hours to get there is just as important as knowing what you are talking about.

A great article about careers from some top industry characters. I will highlight a quote that I really took away from it and personally agree with. It is a lot easier to teach a mathematician to trade than a trader to solve PDEs. The math you learn in your degree is a way of thinking as much as it is useful

https://notion.moontowermeta.com/career-advice

For those looking to learn derivatives and volatility and work on an exotics desk at an investment bank (one of the best ways to learn), in addition to the resources posted earlier Benn Eifert on Twitter often posts interview questions that can't be found elsewhere.

This is a legendary quant interview questions book.

https://b-ok.lat/book/3711947/517cb4

Here is a good channel for careers for extra.

https://www.youtube.com/c/DimitriBianco

Credits:

If you received this document, you probably have my contacts anyways so feel free to ask questions, but I also have a few worthwhile threads on twitter at:

https://twitter.com/TerribleQuant

-BBM