

- Rodrigo Gordillo: 00:00:06 Welcome to Gestalt University, hosted by the team of ReSolve Asset Management where evidence inspires confidence. This podcast we'll dig deep to uncover investment truths and life hacks you won't find in the mainstream media. Covering topics that appeal to left brain robots, right brain poets and everyone in between. All with the goal of helping you reach excellence. Welcome to the journey.
- Speaker 2: 00:00:28 Mike Philbrick, Adam Butler, Rodrigo Gordillo and Jason Russel are principles of ReSolve Asset Management. Due to industry regulations, they will not discuss any of ReSolve's funds on this podcast. All opinions expressed by the principals are solely their own opinion and do not express the opinion of ReSolve Asset Management. This podcast is for information purposes only and should not be relied upon as a basis for investment decisions. For more information, visit [investresolve.com](http://investresolve.com)
- Rodrigo G.: 00:00:54 Hello and welcome everybody to another episode of Gestalt University. My name is Rodrigo Gordillo, managing partner of ReSolve Asset Management. In today's episode we're actually going to go deep and wide on the topic of machine learning and its applications and live training. We've been wanting to do an episode on this topic for a while now and we're likely going to do many more of these because it's just such a complex topic. We can go in many directions, but for our first foray we thought it was best to recruit the help of ReSolve's head of quantitative research, Mr. Andrew Butler. Andrew's got extensive experience in the practical application of machine learning both in real life projects as you will learn as well as the world of finance and what does work, what doesn't, and how we can apply it to our craft.
- Luckily I'm joined by our CIO, Adam Butler, to help tease out and provide context to some of the more complex topics outlined by Andrew throughout the interview. We begin by spending a bit of time on Andrew's background and he discusses the disparity between academic theory and real life utility of applied mathematics. Because this is a massive and very complex area, we actually tried to go under the hood a little bit by breaking down the field into three major areas of interest. Unsupervised machine learning, supervised machine learning and reinforcement learning and provide examples of real life applications as well as investment applications within each one of these broad categories. And of course there's never truly a well rounded discussion of machine learning or artificial intelligence without spending some time pontificating as to whether machines will take over every single aspect of human utility. Or if we're likely to end up with some sort of hybrid model between man and machine.
- Now we wrap it all up by discussing the dangers of working with a lot of the free tools available online for anybody who's motivated to use them. And the risk of possibly doing more harm than good to your portfolio. We specifically highlight that all the kind of complex patterns that machine learning tools are good at surfacing in simulation, don't really matter at all without a thought out and robust validation process, that really only comes with years of experience in both live trading and the skillful combination of all available tools. It really comes down to

time under pressure. So generally this is the beginning of hopefully a much larger foray into this space. But we went on for about an hour and 20 minutes. We hope that you enjoy the episode and that you come out of it with a bit more clarity in this noisy world of machine learning and investing than you did coming in. Thanks again and we hope you enjoy the episode.

Hello everyone. My name is Rodrigo Gordillo. I'm joined here by our chief investment officer at ReSolve Asset Management, Adam Butler, the one and the only. And for the first time in public, he's coming out of his research hole to go in public talking about machine learning today, our head of quantitative research, Mr. Andrew Butler.

Andrew Butler:	00:03:54	Happy to be here.
Rodrigo G.:	00:03:55	How's everybody doing?
Andrew Butler:	00:03:56	Yeah. Good. ....
Rodrigo G.:	00:03:58	So there's numerous podcasts that Adam has been on, so we know his past and history directory. But I think it's important for us to start with getting to know Andrew a little bit more. Why don't we start with the relationship that you two guys have. It's not a coincidence that you have the same last name.
Adam Butler:	00:04:14	That is true.
Rodrigo G.:	00:04:15	It seems to be a genetic pool that comes out of Newfoundland. Specifically-
Adam Butler:	00:04:22	..... the Butler clan.
Rodrigo G.:	00:04:22	The Butler clan.....
Adam Butler:	00:04:24	Yeah. This is Adam in case you don't know my voice, but Andrew is my cousin. He is my father's brother's son and actually quite a bit younger than me. What are you like 40 or 50 years younger than me?
Andrew Butler:	00:04:37	Yeah, like five, six years.
Adam Butler:	00:04:40	Then the other relations in terms of like intellectual curiosity and stuff we'll cover probably as part of his introduction.
Rodrigo G.:	00:04:48	Yeah. So why don't you tell us a little bit more about your academic background, how you ended up with us, what your trajectory has been to joining ReSolve.
Andrew Butler:	00:04:57	Sure. So I mean probably first and foremost I'm a student of mathematics. I studied math and physics in undergrad. Was always very passionate about kind of the creativity that is involved in mathematics, the elegance of it. Also was very keen

on not having to memorize a whole lot of stuff and regurgitate it. So that was a kind of a key underpinning for me studying math.

Rodrigo G.: 00:05:28 Were you a big fan of Fineman? I know that that's his big thing there. Don't memorize, I understand.

Andrew Butler: 00:05:33 As a student of physics for sure. But I've always been really intrigued in solving real world problems with math. So I studied applied mathematics and I had a great opportunity in undergrad to work on real world applications of mathematics. So we were looking at oil reservoirs, highly computationally expensive oil reservoir simulators, simulators for modeling the tidal flow in the Bay of Fundy. We were working on how to set up wind turbines in a giant field for optimal resource allocation. So these types of, I guess it would be operation research type of problems always really intrigued me. That was my first kind of introduction into applied mathematics that was not so much theory, but also really had tangible consequences and could really be applied in everyday industry.

Rodrigo G.: 00:06:31 Then from there, where did that take you? What was the next evolution in your academic career?

Andrew Butler: 00:06:37 I was working on actually eight sub problem of simulators. Which was, okay, so we've got these highly complex computationally expensive simulators and we're trying to do something with them. We're trying to, for example, optimize the net present value of oil in a 60 dimensional oil reservoir field. The challenge was that these simulators were incredibly computationally expensive. They had to solve partial differential equations nested in 60 dimensions. The challenge with doing that in practice was that in order to get a good solution, so again we're trying to optimize net present value of oil recovery. You had to go and navigate this non-linear, non-convex search space. And it's just even if the one function evaluation of that simulator costs a couple of seconds, it will take you a full day in order to get a reasonable solution. So what I was looking at was, is there a way for us to emulate simulators? So this was kind of my first introduction into applied statistics and machine learning.

And it was how can we make a reasonable model approximation by sampling sparse representations of a very complicated simulator in order to make a reasonable emulation of that simulator, optimize that emulation and then recursively update on that process. We found that to be incredibly productive. It was actually at that time I was tinkering in the math lab that I think Adam you started throwing me a few problems in finance.

Adam Butler: 00:08:25 Yeah, well I think you were still in search mode. You weren't quite sure what direction you wanted to go career wise. You were kicking around maybe going into medicine. We had occasion at some kind of family gathering or something. Or maybe you reached out because you were just kind of curious. I forget exactly what the initial catalyst was, but we did get to talking and I had some interesting

challenges and you were fascinated and sort of came together and tried on a few things. I think you caught the bug to some extent.

## Momentum and Optimization

- Rodrigo G.: 00:08:55 So do you remember what Adam had you working on at the time? I would imagine given our trajectory, something to do with momentum or optimization.
- Andrew Butler: 00:09:02 Yeah, actually he had a bunch of momentum models that were just traditional moving average momentum models. I think you were pretty keen on how my simulator emulator machine learning methods would apply to the financial world. I think that was the crux of it.
- Adam Butler: 00:09:23 Yeah. And I mean it turns out, I don't think we actually ended up doing much in that domain at all at the time. We just sort of focused more on your expertise in general sort of numerical analysis in the end. And of course you needed to learn ARC because you were proficient in Math Lab, if I recall. That was your sandbox at the time. So it took a long while before we settled on something that really was back in your wheelhouse from before.
- Rodrigo G.: 00:09:50 We'll get into the nitty gritty of the complexities, transitioning knowledge from machine learning in oil reserves to actual finance. So before we get into that, why don't you just tell us a little bit more of how you evolved, what you started practicing when we eventually hired you and you were working on similar projects as one that you just mentioned. What are you working on today? And I know that you also had further academic aspirations that you're working on today.
- Andrew Butler: 00:10:18 Yeah, it was a pretty big transition coming from the physical world where they're very complicated problems, but by and large, what I was looking at was quasi stationary data. So what that means is that if I were to fit a model on my sample distribution, it would translate very well to the out of sample distribution. Immediately as I delved into the finance problems. That was just no longer the case. You would build these highly complicated machine learning models, or at least that's what I was attempting to do on the onset that fits several hyper parameters and none of them generalize to the distribution of out of sample. So immediately it struck me that something was different with the data. I kind of backpedaled out of that and went into simpler methods like linear models, momentum models and their various different transformations of that, ensembles of simple factor models for forecasting asset classes.

Then I've always had a passion for the optimization aspect of it. So I immediately was just captivated with mean variance, optimization. All of the touted flaws of Markowitz's mean variance optimization, how it's error maximizing and how often you see that portfolios that are formed in sample do not perform well out of sample. That was kind of a reintroduction into more complex models. Got this quadratic model that you're building on portfolio assets and how to apply some of

the techniques like regularization, dimensionality reduction clustering in order to make mean variance optimization, attractable solution for portfolio allocation.

Adam Butler: 00:12:20 You spent a couple of years doing a masters in financial engineering. So how did that come about? And looking back on that, what are your takeaways from that direction?

Andrew Butler: 00:12:32 Yeah, I wanted to get kind of more formal training in financial mathematics, financial engineering. So you and I were working on some little small projects at the time and I decided to go ahead and pursue a masters in financial engineering and mathematics. The program was great though I was much more interested in the problems that we were working on at the time. It was really the application of kind of real world simulations that kind of give you insight and a leg up that, "Hey, not all of the stuff that we're learning in the textbooks is really applicable."

Adam Butler: 00:13:16 Yeah. I mean I think that's a general theme that you've encountered all along. Because I sort of came to quant a little later. You came from a real sort of theoretical background. So I think all along that theme of recognizing that the problem is not nearly as clean as most of the other types of problems that come out of the physical sciences or out of other domains that machine learning or other types of applied mathematics and statistics are typically applied to. We've had to overcome those challenges over and over again, which has been a lot of fun.

## Theoretical to Practical

Rodrigo G.: 00:13:49 What would you guys say is the percentage of knowledge that you gained from the academic side of things? I mean, I came from a commerce finance background with a strong focus on statistics and mathematics. I think there's a broad general theme that kind of gets us somewhat to a point where we could use these topics, but when we start practicing it, it has nothing to do with theory. More than anything, it's about what works in live trading. And the real PhD I think anyway is getting your face piled in by the markets when you try to apply a traditional model. Do you guys find that you get something similar or are you applying a lot of the concepts you learned in academia or some of the concepts coming out of a white papers?

Andrew Butler: 00:14:35 Yeah, no doubt. I would say that I take nothing from a textbook or a white paper verbatim, but rather I take bits and pieces from a large collection of ideas. And that seems to be where the magic is. So you take Markowitz mean variance optimization and you hybridize that with L one regularization. Then you apply a secondary cone constraint in order to maximize the number of bets in the portfolio. And you take these three independent pieces, well now you've got something that is actually tractable versus plain vanilla unconstrained mean variance optimization. That way of thinking I think is massively informed how we think about the problem, how we build portfolios, build solutions.

- Rodrigo G.: 00:15:27 Now to put a bow on that, I understand you're pursuing your PhD right now at the University of Toronto. Now that you have a bit more control as to what you want your thesis to be? How is everything that you've learned, everything we've talked about thus far informing the way that you're approaching that part of your academic career?
- Andrew Butler: 00:15:42 Yeah, well I would say if you're interested in doing a PhD and you can do it, go work in industry first. Go get a taste for what it feels like to actually solve some of these problems in practice rather than just popping from say an undergraduate masters to PhD program. It just gives you so much more world contexts. It's a completely different way of thinking about the problem. Things are not formalized like they are in say mathematics where you've got this foundation of principles, and the axioms and everything must be built upon it. Rather it's like I described, you can grab bits and pieces and pool those together in order to build something that actually has practical value. So it's kind of for that like I'm doing my PhD in the engineering department. I think they're very much advocates of that approach. I think engineers have always been good advocates of yes the theory is great, but we are building things and we are doing things that have real applications.
- Adam Butler: 00:16:49 It's worthwhile to I think, share a little anecdote about your experience in presenting it, one of the top theoretical mathematics institutes in the last year or so on machine learning and finance. And just sort of give your observation about some of the other presentations that were shared just thematically. And I think how it really drives home, this idea that so much of the research that is produced in academic finance, in mathematical finance is produced by academics that clearly have very little experience in the real world P&L of trading. Your takeaways from that I think would be instructive.
- Andrew Butler: 00:17:31 Yeah, I mean it was absolute honor to present at the Field's Institute, the conference was on optimization and artificial intelligence. It was a really great two day conference, lots of really good insights. But broadly the theme that was kind of pervasive across a lot of the presentations was that I have a framework of very elegant mathematics and it's very theoretically accurate and constructed in a very elegant way. But when I bring it to practice and compare it to the simplest of implementations, like equal weighting of portfolio for example, it provides zero benefit. So you're kind of dumbfounded because you listened to an hour presentation where the mathematics is just shockingly elegant and complex but provides no perceived value. Not to say that there aren't bits and pieces in there that have value. So my approach when I presented was to kind of take a practitioner's perspective. And rather than talking about elegant mathematics or the latest in neural network architecture, I really talked about a very simple application to meaningfully improve portfolio optimization outcomes.

I related to mean variance optimization and regularization and basically all the tools that we learn in applied statistics in order to attenuate these error maximizing phenomenon, that we get when we over fit data can be equally applied

to the portfolio construction process. But hardly anyone is using it in practice and there's very little written on it in theory. So it was great to kind of just bring that to the forefront and discuss it. I think the feedback was really great.

- Adam Butler: 00:19:28 Well, yeah, it was bits and pieces from your experience, some from the literature in applied statistics, but brought to bear in a very intuitive way if you understand the nature of the problem with very clear advantageous outcomes, in stark contrast to many of the other papers that were presented. Which were much more elegant and had the air of sophistication but were just not useful in practice. I think that that concept generalizes to the difference between academic theory and trader intuition or people who are actually dealing with real P&L on a daily basis. It's just really hard to understand the nature of the problem without actually getting your hands dirty.
- Rodrigo G.: 00:20:13 No question. Well, I think there's also two different objectives in academia versus in practice. What is the objective of an academic? Well, it's to publish papers and to be elegant and to be thought of as the smartest guy in the room. And like you said, it was elegant. It was beautiful, it was very creative, but it did not apply or maybe it did but there's something simpler. And as a practitioner, what you care about is a parsimonious way of getting to the same point. It's this idea of Taleb's Fat Tony versus, what is it ...
- Adam Butler: 00:20:47 Dr Bob.
- Rodrigo G.: 00:20:47 Dr Bob, one is meticulous, one is thoughtful, one's academic and the other one just wants things to work loud a bit more. But he's a practitioner, and I've actually never asked you this, but when you did present, did any of the previous presenters come up to you and say anything to you? Like what was the feedback from them or did you never speak to them?
- Andrew Butler: 00:21:07 No, I did. The feedback was great. I mean a lot of it was kind of, "Wow, I never quite thought of the problem that way and it's so simple." Literally my construct was that mean variance is no different than simple regression. There's literally a plethora of tools in the regression toolkit such as subspace reduction, different forms of regularization terms that massively improve the outcome of regression models. These have been known for 50 years or more, and the problem is identical, but nobody is using it in, or not nobody, but not a lot of it has been broadcasted in both practitioners and academic circles. So it was very positive feedback and just curiosity in general.
- Rodrigo G.: 00:22:07 So we want to get into the topic of machine learning a bit more just to get everybody who's listening, a general understanding of what it is first of all, how it applies out there in the real world and how it applies specifically to finance. But before we get into any of that, Adam, maybe you could give us an idea of this, high level, what is machine learning?

## What is Machine Learning?

- Adam Butler: 00:22:26 Well, at the very highest level you've got the domain of mathematics and beneath mathematics you have the domain of statistics and then beneath statistics you have theoretical statistics and applied statistics. Applied statistics is I think applying the probabilistic way of thinking to real world problems whereas theoretical statistics is more about making assumptions about distributions which may not hold or be very useful in making predictions in the real world. Then I think about machine learning as just a set of tools within applied statistics to make it a little bit easier in order to generate predictions that generalize well out of sample. And you might be able to add some color to that or it may have a completely different take on it.
- Andrew Butler: 00:23:18 No, I think that's great. I would describe it very similarly. I mean it really is at the broadest of level. It's a branch of kind of both mathematics and computer science because I think they kind of both came at it from the same angle that really just studies algorithms for analyzing data, uncovering patterns or relationships in that data. Then learning from that experience in order to meet some objective as a very broad definition of machine learning.
- Rodrigo G.: 00:23:53 Perfect. So I'm just going to go through a graphic here that we're going to post on the show notes that separates machine learning into three broad categories. I want to go through each one of them and discuss them independently to see where it's applied right now and the areas that everybody's used to. So Amazon, Google, all these, Facebook and how it is that we may apply them in finance. The first category is unsupervised learning and there's two sub categories there. Clustering, dimensionality reduction from a general use perspective, what we see be applied to us day to day is things like targeted marketing, customer segmentation. These are things that are used very easily to sell to us. But in finance, how is it that we are applying this unsupervised learning technique or the series of unsupervised learning techniques?
- Andrew Butler: 00:24:41 I mean, we use unsupervised learning in almost every aspect of the algorithms that we design, whether it be in the signal processing aspect of it or in the optimization, the portfolio construction. But I mean at the broadest level, unsupervised learning doesn't have any labels. It doesn't have any direct model representation in terms of input and output, but rather is trying to uncover relationships or groupings or clusters contained within a dataset. That's kind of the broadest definition of unsupervised learning. So when you look at any kind of predictive model, the total error of that model can be broken down into two components, the bias and the variants. So a model for which the total area is dominated by a large bias is typically overly simplistic.
- As you traverse, so if you can think about it on the X axis you have model complexity and on the Y axis you have total error. Total errors decompose into bias and variance. As you traverse across the X axis, increasing your model complexity,



the bias term decreases, but what is replaced by that bias term is a variance term. And the variance speaks to instabilities and over-fitting complexities that take place when you have small sample sizes and highly complex data's for which there are many parameters to fit. Do you have any color to add to that?

Adam Butler: 00:26:27 No. I mean the way I like to think about it is that a model can leave some predictive ability on the table. There's a large amount of error in the bias because you can add parameters or you can add specification to the model and you can get better forecasts. The challenge being that the more complex the model, the higher the risk that it won't generalize well out of sample. So you've got this on the other side, highly complex model that explains a lot of the variants in sample, but then there is a lot of variability in like it just doesn't generalize out of sample. So it's always this trade off between trying to find a model that captures the maximum amount of true predictive information in the data but doesn't capture any spurious patterns in the data. So this is the trade off and in finance this is an especially tricky problem because a model can describe and be highly predictive both in and out of sample in a certain period.

Then something changes in the market and a model that had very low total variance in one sample can then go on and have very high variance in either a new dataset or at a new time step. And you have to embed a buffer there so that you can stay out of trouble if you move into a new paradigm. So that your model is not assuming that it's well fit to the new paradigm, which I think is what makes financial data really such an interesting area of study.

Rodrigo G.: 00:28:11 Does this all just apply to unsupervised learning or is this concept of ...

Adam Butler: 00:28:16 No, this is a general guiding principle of applied statistics really. It obviously applies or machine learning brings to bear a set of tools that helps to find the right trade off between model complexity and out of sample generalization.

Rodrigo G.: 00:28:36 Yeah, it's model design in general. So I've got a model I'm trying to navigate the world with and it's a representation of the real world process. How might I construct the models so that I capture the salient pieces of information that I want to capture without all of the idiosyncratic perhaps irrelevant noise contained within that data. As Adam said, incredibly complicated and financial data because from one day to the next the distribution could change.

Andrew Butler: 00:29:08 You might have a very true signal for 10 years, then the global financial crisis happens and that goes away completely and we're in a new paradigm. So it might've been real. It might've worked with your sample data. You might've done all the validation processes that are required. You might've put it to work for a little bit and then it just completely goes away and this piece of the non stationarity of financial data in contrast to other applications of machine learning. Like you were talking about the oil reservoirs. You said that the data there is quite stationary for the most part, geologists, geology things that are not going to

change as much in those oil reservoirs as it may in the zeitgeists of financial markets. So that's an added level of complexity that we need to zero in and have extreme focus on when it comes to ensuring that we're not creating models that are just 100% complete data mine.

Adam Butler: 00:30:03 One of the great things about finance is there's an abundance of data. But with an abundance of data, you run into this issue, which is the cause of dimensionality. That's where I think this dimensionality reduction and clustering as a subset of the domain of dimensionality reduction can really come in handy.

Andrew Butler: 00:30:22 Yeah, for sure. So, for example, if we think about the problem of constructing a portfolio where we need to define the weights on an asset, well often in order to do that, we need to construct a covariance matrix, which measures both the variance of the individual assets and how they correlate with one another. Now if we have  $N$  assets, then we require  $M$  plus one over two. So call it on the order of  $N$  squared number of parameters in order to estimate that covariance. So this speaks to sparsity because you can imagine if you have an estimator for which you require 100 observations in order to get a reliable estimate. Well in order to get a reliable estimate of something that is on the order of  $N$  squared, well now you need 100 squares. You need 10,000 observations in order to get a reliable estimate.

And we may have 10,000 observations in our daily or intraday data or what have you. But do we have 10,000 relevant observations? Well, probably not. So how might we use dimensionality reduction in order to attenuate this issue? Some of the stuff that we've looked at are things like hierarchical clustering. So for anyone who's aware of de Prado, he came out with a paper, I think it was going on two years ago now talking about hierarchical risk parity. In which he acknowledged the issue of sparsity and the challenges that take place when you have to invert a sparse, unreliable covariance matrix. The solution that he proposed was that well, instead of relying on the covariance matrix directly, instead we can provide a hierarchical clustering. So by hierarchical, I mean we are pairing up together the assets for which move most closely together and they will be tighter together in this hierarchical graph.

Those that are moving very differently from each other will be further away. And everything in between, it's kind of like a progression of similarity versus dissimilarity. If you structure your data in this fashion, well then you now have a representation of a covariance matrix, which for those who are familiar with covariance matrices, they're square and it's going to look quasi diagonal. So what that means is that the most important elements are along the diagonal of the matrix emanating outwards are close to zero or negative. So rather than performing a minimum variance portfolio, you can construct a inverse variance portfolio sequentially on each of these different hierarchies. We've loved that way of thinking, of reducing the dimensionality of that space and applied a very similar approach to optimize mean variance portfolios. So for which we also have some

conviction on the estimate of expected return. And have found that by applying these types of hierarchical clustering techniques, the out performance that you would get on a large universe can almost be 2:1 in some cases.

Adam Butler: 00:33:50

Yeah. Just to wrap something concrete around that because there's a lot of ends in that because that's the way you think. But imagine sort of you've got a 100 assets and you want to optimize and you've only got a 100 observations for each of those assets. Well, that's sort of the minimal number of observations per asset that you can actually form a portfolio or else the covariance matrix may be degenerates, like it's not a legitimate estimate that you can feed into an optimizer. Let's assume that our number of meaningful observations is set. Can't find any more observations. So all we can do in order to get a better estimate of covariance is to reduce the number of markets, reduce the number of variables. So how can we reduce the number of variables? Well, we do this for example by clustering and you mentioned a method of clustering called hierarchical clustering.

There are lots of ways of clustering, but the idea is to group assets together that have similar characteristics. Now you've got 100 observations, times 100 assets or 10,000 observations. But when you group them all together, maybe you've really only got 10 different clusters of assets. Now instead of only having 100 observations per market, you've got 1000 observations per cluster. So now you've got a much higher density estimate for covariance, a higher quality covariance matrix for estimation, and therefore you create more reliable, more economically meaningful and stable portfolios. So this is how you reduce the dimensionality of the problem to increase the quality of your estimates of the relationships between variables.

Rodrigo G.: 00:35:31

What's interesting here is that people think about machine learning, artificial intelligence, and then they think applying it to finance. What they're thinking about is, is this new machine learning tool going to predict prices better? So is it going to go up or is it going to go down? Should I go long? Should I short? What you've discussed so far deals with none of them and yet offers, as Andrew said earlier, a tremendous amount of value in terms of the returns risk ratio.

## Returns Risk Ratio

Adam Butler: 00:36:00

It's astonishing really. I talked a little bit about some of what Andrew presented in his presentation at the Field's Institute in the recent webinar. But depending on the universe you're running it on, we observed anything from a 30 or 40% increase in long term Sharpe ratio to well over a doubling of long term Sharpe ratio on unconstrained optimization. So this alone is an unbelievably powerful concept.

Andrew Butler: 00:36:24

No question. And the great thing is that it's also to some extent, universe agnostic. So if you're trying to optimize a portfolio for which the universe is dominated, say three to one stocks to bonds, well then the probability that you will be drawing stock at assets and allocating weight to them is much higher. If you construct the

portfolio in a hierarchical fashion and strategically sample in the kind of the methods that we just described, then we can construct portfolios that reach balance. Irrespective of the initial universe constructions, you can balance stocks versus bonds or bonds versus commodities in a much more parsimonious way.

## Supervised Learning

Rodrigo G.: 00:37:10 So I mean part of this discussion is trying to uncover some of the interesting ways that we can apply machine learning tools to the whole world of finance. This is just one of them. It's an interesting one because it's completely different than what people who have never been exposed to this perceive machine learning to be useful for. And it is, I mean there's just a ton that you can do in that regard that has nothing to do with trying to predict the future movement of price necessarily. Maybe this is a good time to move to another category of machine learning, and this is supervised learning. Once again, supervised learning is used for things like image classification. So the idea of identifying the picture of a cat and having your search engine find the images closest to this definition of a cat. Once you label it and you get, you prime it with data and then you're able to easily and clearly find pictures quickly. That's one way of using unsupervised learning. There's just traditional regression and complex regression analysis. How do we apply supervised or a couple of examples of applying supervised learning in finance?

Andrew Butler: 00:38:12 Sure. So I mean the largest distinction between supervised learning and unsupervised learning is that in supervised learning we are labeling and defining our model in a much more concrete way. So we are labeling our features or input variables to some extent like they have meaning in some context, trend for example, or volatility or skewness what have you. And there's typically some kind of target response for which we are attempting to use our features in some combination in order to hit that target response. That would be kind of a progression context or to classify. So combining the features in such a way that we can distinguish between up days and down days, for example, high volatility versus low volatility. A mixture of a number of these different things. I mean one of the ways that we love to look at the problem is typically through the lens of ensembles where we estimate parameters.

So for example, we estimate trend parameters, we estimate volatility, we estimate skewness, seasonality, carry, using several different ways in order to skin the cat. So trend for example can be specified using look backs of five days or 400 days and kind of everything in between. Are we interested in the sign of that trend? Is it positive, is it negative? Are we interested in the rank of it? Are we interested in some other nonlinear transformation of that factor? So there's so many different ways to express our features and this lends itself really nicely to things like random forest. For those who aren't familiar with random forests, let's think about decision trees. So I think everyone's familiar with regression in which we have features for which we are attempting to map .....variable to features in some linear way, that

would be linear regression. Decision trees kind of walk the path of a thinking model in which you have forks in a tree that separate, for example, a trend factor.

Is it positive or is a negative? If yes, go onto next note. If no go on to no to the left. Is carry positive or negative? If yes, proceed and maybe that leads to a buy. If no, maybe that leads to no action. If trend was negative and carry is negative, maybe that leads to sell just conceptually thinking. So this is a construct of a decision tree. What a random forest is, is an ensemble of decision trees in which we sample around our feature space. We sample around our inputs to those features. So these are the boosting and bagging concepts of machine learning in order to build an ensemble of these decision trees. When you do things these ways, you build thousands or hundreds of thousands of representations of those simple trees that I just discussed. Each of which are potentially valid ways to think about building a system and the collection of all of them can lead to some pretty meaningful predictions.

- Adam Butler: 00:41:44 Each branch of the decision tree will lead to some sort of predictive model. Any one of the predictive models may either be not well specified enough for overly specified. But when you take the average or ensemble or take some sort of guided ensemble of all of that legitimate or predictive nodes of this random forest, then the aggregation of all of those small models is much more robust than any of the individual models on its own. It's more likely to have the right trade off between model complexity and out of sample generalization.
- Andrew Butler: 00:42:22 Exactly. There's actually quite a neat and relevant paper out there, I think published four or five years ago by Quantopian. - "All That Glitters is not Gold", which kind of took this framework to the extreme in which they have several different models that have been added to the Quantopian model aggregator. And wanted to build a validation machine in order to determine whether or not these models are going to hold up out of sample or not. If they are, well then they're investible and if they are not well, maybe we shelve them for the moment. So they defined all of these very intuitive features of the models such as the number of times the model has been back tested, the number of trades of the model, its performance statistics, which seem to be loosely predictive. But all of these kind of meta features that I think are really intuitive in terms of determining whether or not a model will hold up out of sample. These became their features for which they injected into a random forest algorithm and were able to produce results that lived in say the 97th percentile of the distribution of all of the individual models out of sample.
- Rodrigo G.: 00:43:40 Yeah. So this, again going back to the fact the reason it's called supervised learning is because you have to supervise, which you have to put features in place that you then want it to go and fit to a certain outcome. So that's kind of a key distinction between supervised learning and unsupervised learning. The human is very much involved into what you want it to do. It's not like you are seeking a machine learning technique to random data and hoping that it finds patterns, especially in

finance. You have to give it features that aren't oftentimes intuitive. Or whether it's intuition like how many times has been data mined is a pretty good human gut intuition approach to create a label or trend momentum, RSI, MACD. All these things are labels that one can use in a supervised machine learning process. Any more on that before I get into the third category?

Andrew Butler: 00:44:28

I would just add that the only supervision that is kind of not talked about in unsupervised learning is the model framework. So for example, if you're trying to do principal component decomposition in a set of stocks, well that has meaning. What you're trying to do is you're trying to find an orthogonal basis of your data. So this is a projection of your data that is orthogonal such that each component maximizes the residual variance of your total dataset. So that's the only supervision in this unsupervised model. That actually has some good insight because what you want is you want to specify your objective explicitly. So if that is something that is valuable to you, then the user themselves will express that.

Rodrigo G.: 00:45:24

Right. In principal component analysis, you could specify the amount of clusters you want. You have to set that upfront so you're supervising in a way, but then afterwards you're letting it do its thing versus supervised learning, which is you have to use a series of classifiers or labels and then using decision trees and random forest to find ... or whatever technique is available in order to find patterns and complex regressions that may be useful in live trading. Very neat. So let's get to the final one. This is the most popular one out there in the news and that's reinforcement learning. This is where AlphaGo lives. These are where Chess masters are being defeated and sent into a depressive stupor because the machines are taking over. Why don't we talk a little bit about that, what the technique is there and how is it applied in real life or in finance?

## Reinforcement Learning

Andrew Butler: 00:46:15

Sure. So the biggest distinction I think between say, reinforcement learning and supervised learning is that reinforcement learning, there are no explicit input output variables. There's no explicit labeling, there's no explicit labeling of the features for which you are providing some input. There's no kind of an input output classification. In contrast, what you set up is a policy for which an agent is rewarded or penalized for. So an agent, a model is given rules of the game. It's given the flexibility of the actions that it can take. And it is penalized for actions that minimize some objective and it is rewarded for actions that maximize that objective. We see this pervasively in these bounded games like Chess and Go. It's really interesting because you think of a game like Chess and I think the latest consensus is that we do not know an exact number of the number of states that a board of Chess can make. So it's this immensely complex game for all intents and purposes, an infinitely complex game. The challenge is if you were to apply a supervised learning algorithm to a game like Chess, well not necessarily all of the approaches, but an approach that you can take is, well you could take snapshots

of the state of the board and learn through some kind of classification or maybe some convolutional neural net or what have you, but some.....

- Rodrigo G.: 00:48:17 Or like Bobby Fisher's book on Chess where he, I remember reading this book, it's like 200 pages of visuals of different Chess positions or states-
- Andrew Butler: 00:48:27 Exactly.
- Rodrigo G.: 00:48:28 ... and the optimal series of moves that you can make from there. So this requires the machine to depend on certain classifiers established, right? In this case, Bobby Fischer, to learn from them and use that knowledge in order to try to be a human. Where a human can be more creative about it on the spot using similar models, like we were saying earlier in the podcast. So you can have a lot of academic research, but you're grabbing a piece of everything. That is what we're good at as human beings. That is what's difficult to do in a supervised learning scenario. Anyways, so you were saying ...
- Andrew Butler: 00:49:01 Yeah, so that's exactly right. So under that framework it would be very difficult for a machine to beat a human, because they are learning from the representative states of the board that humans have presumably been playing and trying to project what the optimal decision is at that current state. So what the guys at Google and DeepMind have done is they kind of took that problem and flipped it on its head and instead of building explicit features for which in order to attempt to recognize states of boards, they've instead decided to apply reinforcement learning. Which is this whole construct of I have an objective. In this case my objective is to win the Chess game and how might I accomplish that? Well, I have moves. I know how a pawn moves, I know how a king and queen traverse the board and I have a reward system. If I make a move in one direction, I am evaluated on that move based on the current state of this board. If it was a proper move, I get rewarded. If it was a move that is going to lead to me losing the game, well I will be penalized for that action. In this way they're able to have the machine play itself in order to build its own mental model. It's own internal Bobby Fischer.
- Rodrigo G.: 00:50:35 You have a series of levers that you are allowed to pull on and you have a goal and then you just iterate over and over on those levers until you reach that goal. This is where a lot of Go masters or Chess masters are baffled because they've seen moves that they'd never seen before because it's not bounded by all the supervision that has been created by humans, Go players and Chess players that you're building on. But rather it doesn't care about what has happened in the past and what had been built on in terms of knowledge for those games. It is just trying to optimize for that one goal which is winning. It's completely learning it from scratch on its own in this bounded scenario. It's similar for when it comes to reinforcement learning. Robotics is a good example too. You have the levers. These are the pistons that you can use to stand up as a robot and your goal is to not fall. So we could try to supervise that and say, okay, if you're leaning left, use piston two in order to stabilize. But when you give enough pistons or enough areas,

degrees of freedom to the robot, it's really tough to do from a supervised perspective. A lot easier to do it from an unsupervised perspective. Just don't fall.

Adam Butler: 00:51:47 I think it's the closest approximation to how humans learn. When you're learning to walk, you've got the laws of physics, you're going to learn the forces that are acting against you and the tools that you have at your disposal to be able to stand up and remain standing. So balance and then walk forward, walk backwards without falling down. I think no one tells you the steps. No one tells you what the gravitational constant is. You just have to learn the rules of the game by trying it. Then recognizing when you do things that don't cause you to fall and that allow you to move forward. Then recognize those actions that cause you to fall down and hurt yourself.

Rodrigo G.: 00:52:30 And you developed the neural synapses to start strengthening the things that work, things that don't. And all of this works within this, gravity's going to be gravity. It's just not going to change. How does this all change? So people see AlphaGo and they say, "Okay, we're done, we can't beat this machine at Go. We can't beat this machine a Chess anymore." Or maybe they can but we're getting to the point where it's becoming very difficult. We think we can translate this to finance and just make an infinite amount of money with 99% accuracy. Why is this not the case when it comes to reinforcement learning?

Andrew Butler: 00:53:05 Yeah, so I'm not 100% sure, but I would imagine that it would speak to that idea of distributional generalization. So the distribution of returns that you have learned on. So backing up, how might a reinforcement learning algorithm work in finance? Well, there's a pretty good approximate analogy. You've got decisions that you can make, you can buy, you can sell, you can hold. Then you've got reward and penalization and the reward and penalization could come in something very simply like profit and loss. So that would be a kind of the most simplest implementation of a reinforcement agent model. The challenges is that the actions that caused you to get profit in one distribution or even a series of distributions may not at all be the actions that work because the distributions are both heterogeneous across different asset classes and also they are non stationary.

Adam Butler: 00:54:10 Well, imagine learning to walk. When you're learning to walk in your family room, you can walk around and you can maintain your balance and you feel very comfortable. But then imagine you go into your dining room and the gravitational constant changes, the axis of gravity changes. And you're using the same exact rules that you derived from learning to walk in your family room, but they cause you to constantly fall down in your dining room. I mean that is a really good analog for the situation that most traders face.

Rodrigo G.: 00:54:47 Then you go back to the living room and you're kind of able to walk like you used to, but something has shifted and changed in that place as well. So it's not just that you're leaving one, let's say country index, set of stocks and then it's different in Canada than it is in the UK. But when you go back to that space, it's also evolving



and changing. So there are constants in finance that last for sometimes forever. We could theorize that the major factors that have been identified are so ingrained in human nature, like value, momentum, quality and all those things. They may have a risk theory or a behavioral theory that will make it last for a long time. As long as we continue to be human. And not be completely taken over by AI, but they are not as stationary as everything else that we've discussed for this area. So while there are ways and patterns we can identify, there is a faint signal in that room with a lot of noise.

Adam Butler:

00:55:44

No, I wouldn't say again that the problems with reinforcement learning are the exact same problems that you face with other types of learning tools, whether or not we are using our own intuition or some sort of informal reinforcement mechanism. In the end as traders or as investors, ultimately we are focused on learning the features, the patterns, the variables that lead to more profit and less losses. Whether it's through reinforcement learning or traditional methods, we are still faced with a problem of not knowing what the distribution is going to be once we get out of the laboratory and into the markets. So I mean I think there's absolutely ways to use reinforcement learning just as there's ways to use other types of machine learning, informed by the experience of experienced practitioners, traders who understand the pitfalls and opportunities of trading in live markets. I think they're a fantastic tool to add to the arsenal. But just like all of these tools, they're almost useless unless they're paired with somebody that understands the nature of the game that's being played.

Rodrigo G.:

00:57:01

Yeah, and I was just making kind of a highlight that everybody focuses on those areas which has just happened to be reinforcement learning and they have such a high success rate that it was important to bring it back. This is where the Chess beating games are being played. This is where the AlphaGo, the Go's are being played. But it's very different in terms of what outcomes you can expect and finance for in every silo of machine learning, including reinforcement learning. So it comes back to ultimately the reason we wanted to start discussing Andrew's trajectory is because it tends to be the trajectory that everybody that gets into quantitative finance period, goes through. This idea of I have a model, I have a theory, I want to apply it from theory doesn't really quite map perfectly to practice. You start using your imagination, you start collaborating, you start being innovative with whatever quant team you have. It really does come down to time under pressure. It comes down to innovation. It comes down to what actually works in live trading and iterating that way.

And all machine learning really is, is another series of tools aside from the traditional ones that we can use in order to try to succeed in this very competitive field, extracting alpha. Any thoughts on that? I mean we've kind of covered it, but what about you Andrew? As you've kind of reached this threshold that you're at in the intersection between finance, quantitative tools and academia. Do you see a time where it's just all machines?

## The Future

- Andrew Butler: 00:58:35 It's tough to say? I mean I always go back to the idea of pilots on an airplane. By and large today with modern aviation, planes can take off, land, fly totally on their own, just using very robust control systems. What gives comfort is the fact that there are two pilots who are supervising and are able to take the wheel if need be ... So you do get this hybridization of man and machine. I think finance is probably no different. I think a lot of the great insights come from traders who have been in the pits for 20 years and who know the full dynamics of the gold market. They know the full dynamics of energy markets and with that insight are able to provide great guidance for constructing machine algorithms that can take all of these features, put them in various different relationships, non-linear relationships, and provide boosted performance.
- Rodrigo G.: 00:59:54 But the wisdom comes from the human. This is the key distinction. The wisdom does come from that. That initial trader's intuition that comes from years of doing the thing that you are asked to do and succeed at it. Then machine learning comes after, but even then you have to supervise it. This is something that's kind of, we have a new partner coming onboard at ReSolve and his career's been built on prop trading. He's seen these guys that trade a single market over and over with incredible success. He's exclusively focused his career on applying statistical tools and machine learning to extract an insane amount of alpha. But I remember when I said, when you come in, we should talk about how we're just going to build this one thing, it's going to be the machine learning fund, and he almost threw up.
- He was like, absolutely not. I mean, this is not where the magic is at. It is part of the tools that I use, but the true alpha comes in the craftsmanship in every aspect of the process that he's built. There is some machine learning used, but there's also some traditional tools used. But more than anything, it's time under pressure learning what works, crafting the features that you're going to put together, grabbing out of the many tools that you can use, which one's ideal for the particular process you're going to use. How are you going to make sure that you're going to sift out the ones that are just noise and ones that are likely to perform at a sampling? Then how do you prune those techniques after the fact? But there is no, right now in his view, a machine learning algo that's going to a black box that you just kind of plug in, put into the market and make tons of money. It still comes down to that intersection between human and tools, not a tool, but tools. Then you just have to have ton of experience and iterate and learn. Just like walking.
- Adam Butler: 01:01:43 Yeah, I mean, I don't think any of us would say never. I'm convinced that very soon the vast majority of profitable traders in markets will be assisted in some meaningful way by pairing with machines. There may come a time in maybe the not too distant future when it will be mostly machines, building machines, competing with machines. The progress in this space has been blindingly fast. I mean, I know they just announced that they've created a machine that can beat the best players at No-Limit Texas Hold'em in a multiplayer game. Which a year

ago even somebody would have said, there's no way, not in our lifetime will we be able to create a machine that can beat a group in No-Limit Texas Hold'em. Here we are today and the algorithm is so powerful that the team won't even release it because it will eradicate all of the online poker, the entire online poker industry. So I mean the complexity of markets is several orders of magnitude larger. All I'm saying is I don't think we're there yet, I don't think we're close.

- Rodrigo G.: 01:02:59 On that, I actually brought that up with our future partner because he's a professional poker player as well and continues to compete in that space. I agree because I was also in that realm. That it'll be fairly, at the end of the day, what it is, it's learning from recent patterns and it is making some moves based on them. And the true top players if you put them into that room. If you put 10 professional players that change their game as the circumstances evolve in that particular table, he believes strongly that it's just not going to be something they could figure it out.
- Adam Butler: 01:03:32 So I read interviews by some of the top players and they said that usually you can beat top players by mixing up your game, but that this algorithm had an uncanny ability to see through that and be able to dominate. Now, I'm not surprised that you're skeptical as an ex poker player, I'm not surprised that he's skeptical as a poker player. But I think we should all be open to the idea that there's going to be algorithms out there that are going to be able to beat humans at virtually every task, no matter how complex at some point in the future. And it's probably closer than we think.
- Rodrigo G.: 01:04:09 I tend to be more skeptical in that. I think there will always be, when there's human creativity required, anything that's automatable, anything that is repeatable and easily done. And by the way, we're all speculating here, we're going way off tangent, but anything that's repeatable, we need to get that in place and get it implemented. Human beings will be better for it. But anything that requires creativity is in my mind going to be either very, very, very far ahead in a couple of hundred years in the future or not doable at all.
- Andrew Butler: 01:04:43 Well, it begs the question, what is human creativity? Like is art creative? Is music creative? Because by and large, a lot of art, I think has been replicated to some extent by these adversarial neural networks or reinforcement learning algorithms. A lot of music can be transcribed into its harmonic decomposition. So there's a very mathematical relationship with music. Can we just uncover the rules of music and learn what makes a hit and from that replicate that and produce our own number one hit.
- Adam Butler: 01:05:24 I mean, generalized adversarial networks GANs, have already demonstrated a capacity to create brand new Monet's in exactly the same style based on if you show them a photograph, they'll create a brand new Monet in the same style, indistinguishable by experts from the original Monet.

Rodrigo G.: 01:05:48 But that offers no value that will go for \$0 million in an auction house. Again, this is all human contract, but what does it value to human beings and what's a value to human beings is not only what the output is, but where it came from. The story behind it.

Adam Butler: 01:06:01 No, no, I agree.

Rodrigo G.: 01:06:02 Yes, I agree that the circle of fifth is what you're talking about in terms of music. If you grab, you can actually decompose a lot of the pop music and to the circle of fifths and different types of scales that are popular that can.....

Adam Butler: 01:06:13 And trace it back to one group in Sweden that's responsible for 80% of all music revenues in the last five years.

Rodrigo G.: 01:06:21 I didn't know that fact. So there's a lot that you can, but there's still, once you dig into what drives human commerce, it's not just all optimality. There are these quirky things that require creativity. In fact, it's demanded from humans who are the ones who have the economic lever and machines won't be able to replicate that.

Adam Butler: 01:06:43 So that's fine. I actually don't disagree. I think that the source of profit in markets will always be the need for humans to interfere. It will always be the fact that a human ultimately will be the one who's incentivized or motivated to generate P&L in markets and humans will disbelieve. They will lose faith. They will have a model or a prediction algorithm, it will work for a time, and then it won't work for a time. The humans that are relying on that to generate a profit, will lose confidence because they don't know what's going on beneath the surface. It's those actions by humans that will continue to create opportunities for machines to generate profits.

So I completely agree that in an agent network and then you've got humans who as you say, don't always exhibit pure wealth- maximizing preferences and they've got other goals and agendas. So by expressing those non wealth- maximizing preferences, it provides an opportunity for machines to take those profits that they're leaving on the table in order to pursue other objectives. But in the end, you will need machines in order to profit from markets. It's just the evolution of what machines will work or not work, I think will be dependent on human frailties and not on machine frailties.

Rodrigo G.: 01:08:07 That makes sense. That's what makes a market right?

Adam Butler: 01:08:09 Mm-hmm (affirmative).

Rodrigo G.: 01:08:09 Unless we get to a point where the value of things are exactly as they should be and the machines have discovered, that is they're telling us how much it's worth.

Adam Butler: 01:08:17 That's right.

Andrew Butler: 01:08:17 Yeah. Well, to some extent there's going to be regulatory and policy of large institutions that have constraints for which are not purely wealth-maximizing, which always leads to-

Rodrigo G.: 01:08:30 Whenever politics are involved ...

Andrew Butler: 01:08:31 That's right.

Rodrigo G.: 01:08:32 There's always going to be opportunity for the wealth- maximizing individuals

Adam Butler: 01:08:36 Politics, regulations, committees, constraints, just humanity, the rules of humanity, I think will for a very long time in the future provide an opportunity for profits.

Rodrigo G.: 01:08:48 Great. I think we've delved very, very deep in this topic. Maybe there'll be an opportunity for us to do a quick 20 minute recap of this. Let's see if we can try to introduce the novice machine learning individual to the topic. But yeah, I think we've covered quite a bit of ground.

Adam Butler: 01:09:05 Being I've had a chance to use your weapons in the dojo.

### Choose Your Weapons Carefully

Rodrigo G.: 01:09:08 Well let's talk about, because you know what? One thing we didn't talk about was the dangers of applying tools like machine learning to your process. One of the ways that I see it is, I come from a martial arts background. I remember getting into the dojo when I first started as a kid and seeing that there were a plethora of weapons that I could use and learn to use fairly well. But every one of those weapons takes years of practice. It takes a year to be able to dominate the nunchucks and which is what I chose because at the time, I think it was a teenage mutant Ninja turtles that were ... What's his name? Donatello, I can't even remember who it was, but point is you can have a wide variety of weapons that are effective in battle. But not only just having access to them like a lot of people have access to right now when it comes to machine learning isn't enough to win in battle. You have to pick them up, use them, test them, practice them, go to war a couple of times and see if they work.

Not only that, even if you become a master at all of them, then there's the extra value of understanding which weapons are going to work in each particular situation, in each battle, in each war. So it becomes a topic very similar to machine learning in that not only are they freely available, they are difficult to master any single one, to master all of them is great, but do you know how to use them in which particular battle? Then if you go at it without any precaution, you might get hurt and you might hurt the people around you. So how does this all come back to

novice applications and machine learning? And also I'd like to get your thoughts in a lot of the white papers that are out there that are getting published on machine learning that we kind of look at and say, "This isn't really applicable." It actually might be the opposite. It might actually hurt anybody who tries to apply this particular tool.

Andrew Butler: 01:10:53 Yeah. So I mean from the very high level, any machine learning algorithm can kind of be decomposed into three parts. There's the model representation, so how you choose to frame it and we discussed supervised versus unsupervised versus reinforcement learning and the different subclasses within. There's the learning or the optimization phase in which you specify an objective in some kind of methodology in order to tune parameters in order to go after that objective. Then the third, which we haven't really touched a whole lot on is kind of validation and regularization.

Adam Butler: 01:11:36 Which is where the real magic is.

Andrew Butler: 01:11:37 That is where the magic is. No question. You can have the greatest neural network in the world unless you're trimming hidden layers and attenuating weights, then it's not going to perform out of sample. Unfortunately, what we see in in academic papers is kind of a number of different things. One being the methodology is applied to a specific regime, say a stock universe and nowhere else. So you have this very limited data set and have no sense of how it generalizes to other forms of data set. You see methodologies for which the complexity vastly outnumbers the perceivable degrees of freedom of your data. So you've got datasets of 20 years, you've got features that are things like 200 day moving average. How many times does that 200 day moving average cross the zero line? I don't know, five. So you've got this super sparse set of data.

Well your trick, do you believe you have 20 years of daily data, for example? When in reality you may truly only have a sample size of on the order of 10. So how might we think about building models that are sufficiently complex because we feel like we have trained ourselves to use these tools but we don't completely bet ourselves in the ass by using them. So this is where the whole regularization and model validation and cross validation comes into play.

Adam Butler: 01:13:26 Then online trimming as well. So observing the models in live trading, this is actually a really interesting point to spend a minute or two because think about a typical asset manager. You typically come out of one of a handful of different schools of thought. Maybe you're a sort of value pure devalue guy. Maybe you're a value quality guy. Maybe you're a GARP, growth at reasonable price. Maybe you're a momentum investor. That's your bias or your style. So you put this framework to work. You select securities, you do this for many years and you spent a long time and it doesn't really work. Now what are your options? You've got one particular style. You know how to do one thing really well. You know how to find

quality companies that are trading for pennies on the dollar. And if you find some reason to believe that that methodology is no longer effective, you will.

Your job depends on you not knowing that fact. You have no other alternative. You run a quality value fund. What's so misunderstood about traditional asset management is just how fragile our confidence is on any particular approach. We can look at momentum and value and quality and low fall and all of these traditional factor approaches. And the standard error on our estimates of their Sharpe ratios and of their means is very large. Our confidence in the persistence of these factors is much lower than most people I think are willing to give credit to. So how do you create robust strategies in an environment where you have low confidence in any particular one? Well, you need to have a wide variety of different strategies in your available basket so that when a strategy stops working, you don't mind getting rid of it.

You don't mind trimming it out of the portfolio because you can then go and rely on five or 10 or 15 or perhaps hundreds of other different types of strategies that you continue to have faith in. Because their distribution in live trading continues to be consistent with what the expectation was from the learning process. So this is what I think most people miss about asset management, that people like to choose their best value manager or the best trend follower, whatever, that's irrelevant. You need to find a large number of trend-followers that do things very differently. A bunch of value investors that do things very differently. Find all kinds of other styles that aren't related to the styles that you're most comfortable with because you just don't know which one of those factors or styles or methodologies are truly legitimate that will generalize. It will continue to persist in performing.

I think that's a realization that we've really begun to internalize over the last few years. So the concentration of the last little while is how can we find new sources of alpha and stop being so focused on type one error. And acknowledge the hurt that you can cause by paying too much attention to type one error and not acknowledging the potential hurt of type two error.

Rodrigo G.: 01:16:35

So just describe type two error for those people listening.

Adam Butler: 01:16:39

Yeah, so most of the factor literature has been focused on identifying, first of all, very simple, usually univariate features like book to market or PE or what have you. That when you apply a linear regression that over a very long time horizon. So in theory you have a large number of samples. So you're looking for a strategy that exhibits up very high T statistic. So a very low probability that the strategy has produced excess returns purely through random chance. But what that ignores is the fact that there may be other strategies where you run these simulations and you evaluate the T statistics, but the strategies themselves are actually highly legitimate. But during the sample period that you have used to evaluate their effectiveness, they have underperformed just due to bad luck. So here you have a

strategy that actually is legitimate, but in sample looks like it isn't legitimate because it has just had a bad string of luck.

So by ignoring these many, many, many other strategies that might actually be legitimate. Then zeroing in only on those strategies that have been legitimate in sample, you are profoundly truncating your opportunity to produce sustainable alpha. Because you've probably rejected a very large number of strategies that are just as effective as the ones that you've honed in on. It's just that the ones that you've honed in on have done particularly well in sample.

Rodrigo G.:

01:18:18

Well, yeah, and I think it's important to understand the layers here. So when I speak to allocators, I go through this thought experiment. How many managers are you meeting a year? If you're lucky you're having quality meetings with 50 managers. If you have, out of those 50 managers, you have to allocate to how many? How many can you actually legitimately allocate to given the minimum AUM constraints, giving the requirements from your committee and so on? Two, three, maybe four. What you're trying to do is you're trying to identify those with hopefully like you're looking at high T statistics, a lot of the times it comes down to that. And you are rejecting a wide variety of managers that might be valuable, that have lower T stats, might have gone through a poor period in sample and you just career risk or whatnot. You're not able to allocate to them.

Then you go into these high T stat guys and they may not be as robust. Now, let's say that you got free reign from the committee and you could use all 40. Now you're including all of them as ensemble that would, we've kind of proven this. As long as they're all somewhat trying to solve a financial problem, trying to solve the problem of picking better at asset classes or stocks as an ensemble, you're more likely to be broadly correct rather than specifically wrong. You're more likely to have a higher Sharpe ratio, lower draw down and so on. But what's beautiful about the layer below that is that as a quant investor, you don't have to be constrained by the 40 managers that you mean. You can create your own features, your own managers, you can create thousands of those from your own pure imagination that come kind of broadly around that, the same trying to solve the problem.

Then if we bring it to the machine learning level, you can ask the machine, you can give it parameters or ranges of things that it can do to create even more thousands of things that make sense in that. You can use those to test, now you can really do significantly more than the allocator can. Then the individual manager that maybe out of those 40 they have three or four techniques. So it becomes a much more robust. It's robust versus complex, it's robust, it's a lot of small edges. Then the key of all this is that you can funnel all of these strategies through this validation process and this validation process does most of the work. When you think about machine learning is like it's trying to come up with techniques. Yes, that's great, but let's see if we can filter a lot of these while what still comes out the other side is enough ensembles that even if a few bad players got through, there's enough bad players that if it's a random walk, you're not getting ... a random walk



is a zero return. So you're getting a bunch of guys that might over a certain sequence of times make money luckily, a lot of them lose

money luckily, they cancel each other out. And what's left is the true signal, the true strategies that have gone through and actually made enough of an impact to the portfolio to capture that alpha. So once it comes down to quantitative investing, comes down to machine learning, you can just do a lot more than what the average allocators constrained by. And maybe we can talk a little bit more about the validation process and dig in that.

## Validation Process

Adam Butler: 01:21:13 Well I don't know. I mean the validation process is where the magic happens. So I like honestly don't really want to go too much into that other than to say that any machine, any data scientists worth his salt can find patterns in the data. That's not the hard part. The hard part is finding, well first of all features that we know have merit, preferably a large number that we know have merit for different reasons. Then having the knowledge and understanding of how markets work in order to create validation layers that do the job in the right way. Because you need a different type of validation for financial markets than you use for other markets that are either stationary in their distributions or bounded or where it's not a zero sum game where you have multiple agents competing for the same alpha source. Here's the really fun thing, I think.

So I would rather have a handful of reasonably significant different alpha sources. Like I'm talking T scores in the neighborhood of one and a half to two than have one alpha source that's got an in sample T stat of three. Because first of all, in an adversarial agent model, the strategy with a T stat of three is going to be the one that everybody else zeroes in on. They're all going to find that strategy. That's going to be the one that most of the capital is deployed to. What you want instead is a group of weak but significant classifiers that all do things for different reasons or identify different effects for different reasons that work well together. Now you're not relying on any one strategy to create P&L. As you say, if you've got a very large number of strategies and only a handful of them are legitimate, are generating true P&L. Well at the limit, all of those other strategies are just errors.

They're error terms and at the limit they cancel each other out, their trades cancel each other out. So the variance from your noise strategies goes to zero and all you're left with is the signal from your alpha strategies. So I mean this is the dimension of the game that is not at all dealt with in classical finance. And that most of the investors who are focused on the factor literature and on zeroing in on the best single factors or the best single specifications completely miss. I think it's going to be the difference between long term out performance and constantly searching for that one holy grail that you never quite find.

Rodrigo G.: 01:23:50 Maybe that's the reason that you don't hear about it.

Adam Butler: 01:23:52 Absolutely.

Andrew Butler: 01:23:52 So true.

Rodrigo G.: 01:23:54 I mean it is that validation process is where all the magic is and it is what people see as a mere component of it all.

Adam Butler: 01:24:01 Exactly.

Rodrigo G.: 01:24:02 Then even the idea of doing enough of that where whatever does fall through that filter, even if you've got a few wrong, they should cancel themselves out. **You just have to keep on feeding that machine.**

Adam Butler: 01:24:13 Exactly.

Andrew Butler: 01:24:14 Very cool.

Rodrigo G.: 01:24:16 I think we've covered quite a bit of ground here today. Are we satisfied? Has the world's learned enough about machine learning today and AI to dispel some of the myths and recognize that it's not a silver bullet.

Adam Butler: 01:24:26 I think we did a good job scratching the surface.

Andrew Butler: 01:24:29 That's right.

Rodrigo G.: 01:24:30 Oh my God, you don't understand how complex this is going to be for everybody that's hearing this. I wanted to make this 20 minutes. Ladies and gentlemen, it's two hours into it roughly, but thank you gentlemen for coming in. Andrew, thank you for joining us and your inaugural public appearance and we will bring you on again for sure. Adam, thanks again.

Adam Butler: 01:24:50 Thank you. Consummate host as always.

Rodrigo G.: 01:24:53 We will try to put as much of the content in the show notes as possible that we have on this topic and as always reach out if you have any questions.

Adam Butler: 01:24:53 Thank you.

Rodrigo G.: 01:25:04 Thank you for listening to the Gestalt University podcast. You will find all the information we highlighted in this episode and the show notes at [investresolve.com/blog](https://investresolve.com/blog). You can also learn more about ReSolve's approach to investing by going to our website and research blog at [investresolve.com](https://investresolve.com), where you will find over 200 articles that cover a wide array of important topics in the area of investing. We also encourage you to engage with the whole team on Twitter by searching the handle [@investresolve](https://twitter.com/investresolve) and hitting the follow button. If you're enjoying the series, please take the time to share us with your friends



through email, social media. If you really learn something new and believe that our podcast would be helpful to others, we would be incredibly grateful if you could leave us a review on iTunes. Thanks again and see you next time.