**CAPSTONE IDEA**

**Table

Description automatically generated**

**Table

Description automatically generated**

**TITLE**

Building a better equity analyst

**ABSTRACT**

**INTRODUCTION**

Problem Statement: The investment management industry is facing numerous challenges and pressures. Data availability and complexity is growing, the stock market is becoming increasingly efficient, management fees are being compressed and competition continues to rise in the form of active and passive investment products. On the one side of the industry are investment managers that are failing to embrace the technological innovation in the form of AI and data science tools. And on the other side are quantitative managers that are scouring the data universe for new sources of alpha generation. Rarely do fundamental and quantitative managers bridge the gap between their expertise in a collaborative and effective manner as both are entrenched in their respective camps for numerous reasons thus we believe a chasm remains. In order to survive and thrive in this dynamic and challenging landscape, investment firms need tools that are: systematic, understandable, adaptable, complementary, efficient, scalable, multidimensional, innovative and effective. We are proposing the bridging of the gap between fundamental and quantitative insights and tools by bringing together domain knowledge and crowdsourced information with quantitative tools primarily in the form of machine learning and neural networks.

* State of the investment management industry
  + Facts = Data availability and complexity is growing, the stock market is increasingly efficient, management fees are being compressed, competition continues to rise
    - Fundamental managers are not embracing AI fast enough
    - Quants are struggling to find new sources of alpha
  + Needs = Investment firms need tools that are: systematic, understandable, adaptable, complementary, efficient, scalable, multidimensional, innovative and effective
* Addressing needs
  + Systematic
    - Remove biases
      * Gut instincts
      * Emotions
        + Can be driven by job pressures and clients becoming increasingly short term focused
      * Analyst sales pitches
      * Human’s are often irrational beings
      * Numerous behavioral biases: confirmation bias, anchoring, saliency, framing, etc..
    - Instills discipline
      * Challenges the status quo and consensus views which can be affected by aforementioned biases
      * Make it hard to “explain away” new and pertinent information
      * It is uncomfortable to go against opinions, readings, perceived facts, comforts, herd mentality, biases, fears, human psychology
      * PMs struggle when in no man’s land and often aren’t sure which way to swim, as well as when “out of favor” and holding on too long when things “are working”
      * Need discipline not only when buying and selling but in between
    - Repeatable
      * Consistency is key
      * You can’t have consistent results without consistent processes
      * If process and performance are repeatable then client trust builds
  + Understandable
    - Creates and maintains trust
      * Human vs model challenge
        + People don’t trust machines
        + Manager and client must understand output in a manner that can be easily explained so they can trust / reduce the risk of getting blindsided
        + More than black box but rather stats to compare fundamental and technical assessment uniformly by oneself
        + Who drives who? Human or computer?
        + Driven by stats not opinion, biases, etc.
      * Is simple to execute and understand
        + Complex and thorough inputs but simple outputs
        + Einstein – “Everything should be made as simple as possible, but no simpler”
        + KISS principle – “Keep it simple stupid”. Excuse and reason used in my life.
  + Adaptable
    - Though human psychology doesn’t change, the world is always changing
    - Need to be able and discern when to change as inputs warrant change
    - History doesn’t repeat but it rhymes so it is a balancing act between adapting to change will reading the tea leaves from the past
    - Backtests are too static
  + Complementary
    - Change is a four letter word in the investment industry. Rightfully so, b/c Mr Market can pressure change until one capitulates which is often the exact wrong time.
    - Need tools that complement and build onto your process, but do not uproot the core tenets and foundation. The foundation must be set before the storm comes. You don’t pour concrete during a hurricane.
    - Clients invest with you based on what you previously communicated so you must stay true to this
  + Efficient
    - Covers large universe
      * Need to see what is going on throughout universe so opportunities and risks are not missed
      * Able to see the forest AND the trees
    - Best sources
      * After identifying key data sources, such info needs to be consolidated such that it can be easily accessed and explorable for further modelling and analysis
  + Scalable
    - Reduces cost
      * Fees are being compressed
      * Need to expand scope without losing out on execution ability
      * Tools need to be scalable.
  + Multidimensional
    - Breakdown silos between departments and roles
      * PMs, analysts and traders need all be on the same page
      * Often incentives are not aligned between the three thus the outcomes are not optimal
        + Ex. PM wants to buy a stock and tells trader who is then judged on purchase price in the near term as trader is not “allowed” to say no to buying in the near term.
        + Need ways to bridge the skills gap between PMs and traders
    - Costs
      * Costs are also reduced if you can have people, tools, models, etc. that interact with each other and cover more bases thus requiring less manpower
  + Innovative
    - Don’t fight technological change but rather embrace it.
    - Technology is either a tailwind or headwind. If you think it is neutral then it is a headwind as you fight to stand your ground. Choose accordingly.
    - Again, change is often viewed negatively thus what you incorporate should be viewed as being as innocent as a dove but in fact is a shrewd as a snake.
    - Dicotomy = PMs invest in innovative companies b/c of such innovation and the potential it brings but they themselves are slow to embrace it
  + Effective
    - At the end of the day, this business is about winning.
    - The more you “know” in this business, the more you realize that in fact you don’t

. Anthony J. D’Angelo is quoted as saying, “Don’t reinvent the wheel, just realign it.” That is what is being sought in this study is predicated less on coming up with tools that are brand new but rather leaning on and learning from others and seeking to fill in identified gaps and shortcomings in a manner that produces effective results and addresses all the aforementioned needs of investment managers.

* + - D’Angelo – “Don’t reinvent the wheel, just realign it.”
    - Instead of trying to come up with something “new”, rather lean on and learn from the best
      * Don’t try to outsmart others including yourself but rather learn from and take cues from those deemed to be most informed regarding the companies you are investing in:
        + Sell side analysts

Their main job is to intimately study their companies and follow their every move

Don’t think you know more than them either

They can’t see the forest though so have many biases yet they know their companies’ fundamentals inside and out

Pull from their knowledge and avoid / learn from their biases

* + - * + Company management

They feed the sell side analysts info and though they try to sugar coat info they need to be deemed reliable or their stock price / wallet will be punished.

Follow their lead with buys. They put their money where their mouth is when opportunities arise. They sell for various reasons but the main reason they buy is b/c they see value.

* + Summary
    - Identify what is important and avoid what is not
    - Lean on experts and what the numbers say not what the people say. Numbers over narratives.
    - Does it work and can you understand the output so you avoid machine driven blowups
    - Must be able to clearly communicate to colleagues and clients
    - Does it meet all of
* **Solution options: Human and/or Computer**
  + Human
    - Hire more people
    - Analyst coverage, DCFs, meet with management
    - PROS
      * Experience, focus on what is important, not as prone to following spurious / noisy data
    - CONS
      * Subjective, can’t cover universe, emotional,
      * Analyst model – best marketer wins, can a kid really help
      * DCFs – you can make a model say anything you want
      * Don’t adapt well
  + Computer
    - Quant shop – typically premised on numbers over fundamentals, less experience and knowledge regarding fundamentals
    - PROS
      * Objective, unemotional, statistics driven, humans can’t process as much
    - CONS
      * Lost in the weeds, miss the forest
      * Too adaptable, noisy data
      * Not sure what to do when models aren’t working
  + Chasm
    - There is a chasm b/w these two
    - As well as a chasm between fundamental and quantitative
    - Different mindsets, different experience, very different schooling
    - Entrenched approach so difficult to marry the two per se
  + AI in fundamental investment management
    - Why AI is not being used more???
      * Client distrust, misunderstood, blackbox, less personal than human, computer scapegoat
      * Reference - A 2019 study by the CFA Institute that surveyed portfolio managers found that only 10% had used any artificial intelligence or machine learning in their investment process.
      * Reference – 10 reasons ML funds fail
    - Resistant to change
      * Old timers
      * Fear what you don’t know
      * Can’t learn or choose not to
      * Day in and day out in the trenches / daily grind
      * Wrong assumptions = black box, can’t be interpreted
      * Think they know more than they actually do. False sense of security / knowledge.
      * Client demands / fears
      * Uninformed
      * Change is a four letter work but shouldn’t be when viewed in the right context / lens
      * Investment management is not known for adoption of new technologies
      * ……see DDS presentation
  + Fundamentals in quantitative
    - Different schooling
    - Takes significant time to study companies and learn nuances
    - Similarly entrenched and resistant to change
* If all this is believed then what response to make
  + Bury your head in the sand
  + Buy a quant shop / team
  + Build capabilities from the ground up
  + Bridge the gap/chasm between humans and technology = YES
    - Leverage both of their strengths
    - How do we propose to do this?.......read on…..
* Solution
  + What if there was a solution that answers all of the needs above and leverages the benefits of humans and computers
  + Combine domain expertise and AI tools
    - Not just about developing powerful tools but one must understand how to use them
    - Analogy – Racecar vs streetcar: to fully differentiate the two you must know how to drive / apply them
    - Instead of scouring the universe for anomalies, we will take theories / domain knowledge / observations and seek to prove them by relying on individual and collective insights in the form of crowdsourcing
      * Reference – Backtesting protocol – have economic rationale for variables

**LITERATURE REVIEW**

* Hypothesis: The investment management industry will be pressured to embrace the use of data science tools in order to survive and thrive. Humans and computers have their strengths and weaknesses and will be best served to leverage the strengths of both in order to best evolve and compete effectively. We believe the combination of domain knowledge, crowdsourcing of information and data science tools are a recipe for success. Our research review is focused on the following topics:
  + Quantitative investing overview
    - BofA Quant Primer
  + Current usage of machine learning in the investment industry
    - 10 reasons ML Funds Fail
    - ERP ML
    - AI Pioneers in Investment Management
  + Academic research on machine learning in the investment industry
    - A Machine Learning View on Momentum and Reversal Trading
    - A Backtesting Protocol in the Era of Machine Learning
    - Fundamental Analysis via Machine Learning
    - Machine Learning-Based Financial Statement Analysis
    - Machine Learning for Stock Selection
    - Predicting Profitability Using Machine Learning
  + Academic research regarding analyst revisions
    - The Information Content of Financial Analysts Forecasts of Earnings
    - Herding
    - Analyst’s stock views and revision actions
    - Financial Analysts’ Forecasts of Earnings
    - Herding among security analysts
    - Seemingly inconsistent analyst revisions
    - The behavioral basis of sell-side analysts herding
  + Academic research regarding momentum phenomena in the stock market
    - Momentum Strategies
    - The role of analyst forecasts in the momentum effect

**METHODS AND DATA**

* The focus of our analysis will be to develop predictive models that focus on the following aspects of a given investment:
  + Valuation
  + EPS
  + Revenues
  + Price
* Categories on model predictions to be made for a given metric:
  + Level
  + Trend
  + Relative level
  + Relative trend
  + Relative classification
* Initial methods we may use based on preliminary analysis
  + Machine learning algorithms
  + Neural network algorithms
* Data source
  + FactSet database
    - Through the SMU Business Library we gained access to FactSet’s database via an excel add-in.
    - We downloaded specific data on ~652 US large cap stocks going back to 12/31/09.
    - We then created numerous custom metrics based on the downloaded data.
    - Categories of our financial data include:
      * Valuation
      * Analyst revisions
      * Analyst EPS estimates
      * Analyst revenue estimates
      * Price
* Upon developing models we will then determine recommended usage and application of such tools as well as visualization tools for users of our models.
* Then we will create a scenario analysis whereby we apply our models and tools to the year 2020, as we “take on COVID”
* **Proposed Solution = BUILDING A BETTER ANALYST (using data science tools and domain knowledge)**
  + *(Premise is that we can go in many directions with this with various models, ideas, etc. Then try to bring it all together with some insights for a hypothetical portfolio manager via a dashboard or something clear and simple (to use and build). Much more of a playing in the sandbox approach whereby we can work together on things or explore different ideas, thus we won’t be pigeon holing anyone to anything.)*
  + What makes a good equity analyst?
    - Creative, thorough, insightful, to the point, helps to cull through the masses of info, value adding, brings skills to the table that PMs don’t have
    - Human and computer roles that can support an analyst
      * Human
        + Supervised

The human decides what variables are important and what the models will be allowed to learn from

Don’t data mine and search for sources of alpha that are not rooted in underlying philosophy or process

Will always be chasing what is working now

Must go into battle with a strategy and stick to it

Rooting in fundamentals enables this

* + - * Computer
        + Statistics driven
        + What is vs what should be
        + Multidimensional processing
        + Sources

Machine learning, time series, neural networks

* + Sources of information
    - Crowdsourcing those that know best
      * Sell side analysts
        + Reference – Financial Analysts’ Forecast of Earnings

Reference – The information content of financial analysts forecasts of earnings

* + - * + Use innovative tools to analyze

Reference – Fundamental Analysis via ML

* + - * + Momentum – not just in price but also in fundamentals

Reference – the role of analyst forecasts in the momentum effect

* + - * Management – main source of estimates
      * Market – price and valuation
      * Study them so you know their next move
        + Reference - Herding
      * Company – let their numbers speak for themselves
        + Reference – Fundamental Analysis via ML
        + Reference – Predicing profitability using machine learning
  + Stock investment 101
    - Simplistic breakdown
      * Price = valuation applied to fundamentals
      * Price = EPS x P/E
      * All are important
      * Different strategies focus on different components
      * Theory = EPS trends for longer time periods and is less affected by the probability of mean reversion, Valuation is mean reverting
        + Therefore model them separately instead of just focusing on price
      * Important to try to cover your basis
        + Chase EPS but if valuation gets out of line then reversion pain
        + Low valuation but fundamentals don’t support then timing poor
        + Focus on just price – EPS or valuation could be ripe for reversion but have no foresight
  + General plan
    - Bet on 2 while respecting the third = cover all our basis
      * Different underlying theories for each of the variables thus they should be modeled and analyzed separately
      * Theory = Bet on building blocks (fundamentals and valuation) while respecting the price trends, but the results of our analysis will determine proposed strategies and applications of the tools we develop
  + Tools / models to be built
    - **FUNDAMENTALS MODELS**
      * ***EPS growth and revision prediction models***
        + 3 and 5Y lookback
      * Why start with EPS? Major driver of stock price performance over time
      * EPS and Sales theory
        + EPS trumps and trends

Verbiage from ML1 project

Earnings trend just as other things in life…..an object

Larry Kudlow = “Profits are the mother’s milk of stocks.”

Past revisions are serially correlated with future revisions

Key is predicting when past revisions will persist into the future or will not

Trying to predict analyst revisions not stock prices as the theory is that if you get the revisions right then the price will follow / align with

* + - * + Why consensus instead of historical?

What leads and/or correlates with price?

Reference – predicting profitability, fundamental analysis via ML

Except we will focus on consensus b/c more informative and powerful

Reference – The role of analyst forecasts in the momentum effect

Reference – NFJ study

There is power in the revisions / direction of analyst estimates

Reference – the information content of financial analysts’ forecasts of earnings

* + - * + Summary

Must forecast to capture power of EPS

Models to be used: start with ML at least

* + - **VALUATION MODELS**
      * ***Valuation reversion models***
      * Assume that valuation is mean reverting thus trying to model the likelihood and/or amount of potential mean reversion
      * Can get EPS right but if valuation is compressing then returns will suffer
      * FPE = most common metric used
        + Reference – BofA quant primer
      * Assumption = Difficult to model with exact precision thus trying to stack the odds in our favor.
      * Fundamentals are more stable thus easier to model with some precision vs valuation as valuation can ebb in flow given market environment, flows, sentiment, etc.
        + Examples = NEE
      * How to forecast then?
        + 3M reversion models – likelihood of FPE, DY and/or EVS moving in a positive direction in the next 3 months

Goes back to the theory of momentum and trending. Once it starts, then the likelihood of continuing is higher.

Prevents valuation compression from negating EPS strength

* + - * + Mean reversion – percentage return during the next 3 months
    - **PRICE MODELS**
      * ***Price prediction and classification models***
        + 3 and 5Y lookback
      * Why not just P
        + Most try to model P / Price but more info is embedded into P than E and the valuation component results in more volatility / uncertainty with the P
        + Initially focus on modelling the building blocks of Price but will still seek to develop models to aid in Price prediction and/or aid in timing of buying and selling stocks
        + Trying to model fundamentals all the while respecting that the market focuses on various aspects at various times so model the fundamentals and then seek to determine adequate timing (i.e. yes or no)

You can be right on the fundamentals but wrong on your timing

Timing matters

Inputs the same so you are still using the same lens

* + GOALS = create models that predict and analyze the building blocks of a stock’s price as well as the underlying Price and then creating a user interface / dashboard to easily comprehend the output
    - *We can add upon this during capstone and/or recommend other ways to build upon our work and the toolbox we have begun*
    - With these tools in hand we will then use scenario analysis to try to navigate the market
    - Not about creating a blackbox but rather powerful yet understandable tools that can ultimately be used by a PM to aid in decision making / execution of underlying investment process / philosophy
      * If you try to change course during the middle of a battle then likely to fail
    - **What better way to build and put such tools through the fire of investing than to try to navigate 2020 / COVID**
      * Establish a strategy ahead of time and let the model adapt
      * Let the model adapt, don’t try to adapt yourself as emotions will cause you to make the wrong choices
      * The model has no emotions
* **RESULTS**
  + Summarize model building and visualization efforts
  + Summarize 2020 scenario analysis
  + Accept, reject or modify hypothesis based on the results
  + Clarify study scope
* **DISCUSSION & CONCLUSION**
  + Findings
  + Insights
  + Takeaways - Wins and losses
  + Could be improved by:
    - Valuation insights
      * Having more valuation information vs peers using DS to understand best way to determine relative value
    - Insider transactions
    - More technical variables
    - More macro variables that directly or indirectly affect companies
  + Revisit opening statements
* **REFERENCES**
* ***General outline***
  + Background on / state of industry
    - Fundamental managers are not embracing AI fast enough
    - Quants are struggling to find new sources of alpha
  + Potential solutions
    - Human
    - Computer
    - Bridge the gap between human and computer, fundamental and quantitative
  + Proposed solution = building a better analyst using data science tools and domain knowledge
    - Background
      * P = PE x EPS
      * Bet on 2, respect the third
      * PE – FPE is most used yet only 10% use AI, untapped
        + Reference – AI pioneers
      * EPS – consensus vs historical
    - Theory
      * Build models to predict each of these
      * Most focus solely on price
      * Price contains building blocks with different underlying attributes
      * EPS = not mean reverting
      * PE = mean reverting
      * Bet on EPS and/or PE then respect the price
      * Will compare to just price and see if we can beat EW index
    - Why
      * Show stats
    - What / goals
      * Tools to navigate the market and especially “new” environments (COVID 2020)
      * Not just backtest, but adapt in a timely manner
      * Domain knowledge = choose metrics / variables
      * Computer = horsepower
    - Model building
      * What data and why
      * What models and why
      * How to bring it all together
    - Application
      * Dashboard
      * Use cases
      * Human oversight – decide when to use which model(s)
    - Implementation
      * Run hypothetical tests and scenarios
      * Detail findings and insights
      * Discuss how one can use these tools to navigate various market environments
  + Conclusion
    - Revisit opening statements
    - Takeaways – wins and losses
    - Proposed next steps
      * Macro variables
* STEPS
  + Gather data
    - Combine it all
    - Segment for building and testing
  + Prove consensus > historical EPS
  + Preprocess / EDA (just for YE 2019)
    - Outliers
    - Missing values
    - Duplicates
    - Scale
    - Types
  + Cluster
    - EPS (5)
      * EPS stability score = average 3Y and 5Y spearman and pearson R2
    - FPE (5)
      * FPE stability score = average 3Y and 5Y spearman and pearson R2
  + Model framework
    - EPS: response = 3M Fwd EPS Growth AND Revision
      * ML – 5 models
        + Pycaret to determine which model
        + Optuna to tune
        + Finalize
      * Time series?
    - FPE: response = 3M Fwd FPE Change AND classification revision
      * ML – 5 models
        + Pycaret to determine which model
        + Optuna to tune
        + Finalize
      * Time series?
    - Price
      * Baseline: response = 3M Fwd Return
        + ML – 1 model

Pycaret to determine which model

Optuna to tune

Finalize

* + - * Timing: response = classification will / will not beat benchmark median during next 3 months
        + ML – 1 model

Pycaret to determine which model

Optuna to tune

Finalize

* + - Summarize model performance and compare
      * Choose model (ML vs times series) based on loss function OR average results based on loss function
  + Strategy
    - HUMAN KEY ELEMENTS
      * Create “buckets” / strategies based on various criteria which is premised on being applicable for different market environments. This is a way to go back to leveraging the human. You build models based on fundamental and theoretical beliefs, then let statistics drive. Then it comes back to the human as to how to implement:
      * Some times you may be more willing to take risk and go after highest growth regardless of quality thus you’d pick more from the lower quality basket and bet on mean reversion or something like this
      * Hero trade = do this coming out of COVID…..now may be a good time to do this too.
      * Other times you want to focus on quality and best growers there
      * Does the human believe in mean reversion for the market or a given stock…..then choose accordingly
      * Thus create time series models that differ on their assumptions? Not sure if this can be done in scale or not????
      * This is where we expound on not just making “good models” but understanding the models being built and what their assumptions are. Then force the human to intervene. Thus the human decides what info is most important (beginning), the models do their thing (middle), then human is ultimately accountable to determine best application (end)
      * The marketplace is fearful of blackbox models thus this avoids this and gets back to the point of leveraging the strengths of both and understanding the models outputs, drivers, etc.
    - Baseline
      * ML model of Price
      * R1000
    - Proposed strategies
      * “EPS trends and trumps”
        + Top 2 or 3 clusters (depends on model loss function), pick top 50 that have highest EPS growth and yes price classification
        + Top EPS growth
        + Top EPS revision
        + Combine
      * “Valuation matters most”
        + FPE – top mean reversion potential with yes price classification

Used when you trust the denominator

* + - * “Price is what you get paid on”
        + Standalone
        + Combo
      * Combo
        + Bet on 2 and respect the 3rd
        + Theory = If you can get valuation and fundamentals right then you will win big

½ EPS and ½ valuation candidates

* + - 2020 scenario
      * Rerun models at end of each quarter
      * Decide on strategy at each rerun
      * Assess performance during next 3 months
        + Return – mean, median
        + Volatility
        + Compare with baseline price models and EW index
      * Output / Dashboard
        + Predictions
        + Drivers
        + Confidence / loss function
        + Line graph with historical and predictions
  + Conclusion
* Findings
* Insights
* Takeaways - Wins and losses
* Could be improved by:
  + Valuation insights
    - Having more valuation information vs peers using DS to understand best way to determine relative value
  + Insider transactions
  + More technical variables
  + More macro variables that directly or indirectly affect companies
* Revisit opening statements