# Stock Market Risk Heuristics Using Compressive Sensing

Anurag Pallaprolu (9847112), Aditya Swaroop Attawar (9874876),
ECE Department, University of California, Santa Barbara
March 21, 2020

**Abstract**

This report details the work we have done in partial fulfilment for the course ECE 271A (Principles of Optimization) at UC Santa Barbara. From the point of view of the occurrence of rare-events, there is a fundamental asymmetry in the evolution of stock market time series. We perform a case study on two theories of portfolio allocation [1, 2, 3] and argue that both of them suffer from the monotonocity that they borrow from the historical data used to perform the optimization. We explore an analogy between these rare events and sparse dependence on prior evolution to construct a pipeline that gives us an estimate on the "pre-cognitive window": the amount of time prior to a market movement, so that one can then observe the movement in shorter windows and hedge bets accordingly.

# Contents

# 1 Prelude: Non-Observability and Convexity

Any model that mines historical price data implicitly assumes equivalent future states that are bounded by the largest volatility observed so far, thereby failing to safeguard the portfolio from unknown outliers by definition. This problem is known as the problem of induction in philosophy (as treated by Hume, Russell, Popper) and that of non-observability in statistics [4, 5]. The consequences of non-observability are felt when the predictions we make about the future state of a complex system end up throwing large errors, which can be seen as "overfitting" the estimate to "noise" when viewed from a signal processing perspective. In the case of the stock market ecosystem, one can observe such deviations whenever the trading agent is forced to make a "non-convex" decision, and in the following sections we make this terminology precise.

## 1.1 Convex Preferences

Classical economics defines a preference as an allocation of assets available for acquisition. Therefore, if we purchase $u_1$ units of stock $S_1$ and $u_2$ units of stock $S_2$ and so on, our preference for this bundle would be represented as $(u_1, u_2, ...., u_n)$. One can then establish well-ordering between choices of such bundles based on an agent satisfaction utility (average return). The level curves of this structure define what are called indifference curves [6], and a "convex" or a "well-behaved" preference implies that if we are provided with two bundles: $(u_1, u_2, ..., u_n)$ and $(v_1, v_2, ..., v_n)$, we have:

$$\mathcal{U}(\theta(u_1, u_2, ..., u_n) + (1-\theta)(v_1, v_2, ..., v_n)) \geq \theta\mathcal{U}(u_1, u_2, ..., u_n) + (1-\theta)\mathcal{U}(v_1, v_2, ..., v_n)$$

where $\mathcal{U}$ is the utility/satisfaction function specific to the agent. One of the characteristics of non-observability is the failure of this convexity in the activity of real-life markets (and decision making in general). A safe choice is defined as one which lies within the convex hull of a particular set since the slopes of the indifference curves are normal to the direction of gradient: Generally, when we make decisions, we would want them to lie within the convex hull. However, the occurrence of non-observable events could result in the contrary. An example that clearly depicts

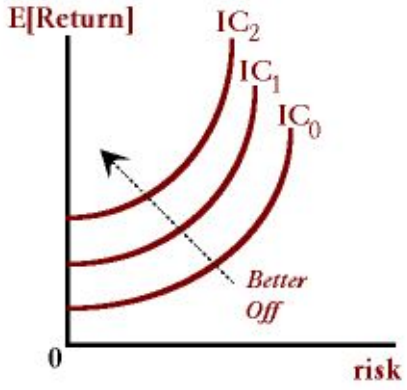


Figure 1: The Indifference Curves.
Source: Digital Economist, Portfolio Analysis and the Demand for Cash Balances

such an event is the recent outbreak of the virus COVID-19. What started off as a localized problem, was unforeseen until almost the entire world was suffering its impact. Furthermore, this unexpected spread of the virus resulted in another catastrophic non-observable event - panic buying at various supermarkets across the country [7]. In this broad scenario, we can notice that each of the 3 agents in the market - the virus, the public, and the supermarket owners - are all trying to be convex with their choices. The virus is being convex (by its very tendency to survive) with its unexpected spread, the people are being convex through panic buying (moving inwards the hull), and the shop owners are trying to be convex by stabilizing their sales to limit the preference set of each customer (pushing outwards from inside the hull).

## 1.2 Failure of Convexity

In Figure 2, $I_1, I_2, I_3$ represent the Pareto fronts or the indifference curves of choices made for goods X and Y. To recapitulate, any choice of combination for X and Y on these curves will result in the same level of satisfaction. The panic buying situation explained earlier can be correlated to this representation. We can visualize the preferences of various agents as those of moving into the convex hull since this would result in a feeling of security. However, there exist infinitely many degrees of freedom that are external to a particular optimizing agent, which can lead to a non-convex region in its feasible set. As shown in Figure 3, any decision based on a convex optimization scheme can lead to ”jumps” in the solution. The magnitude of risk depends on the

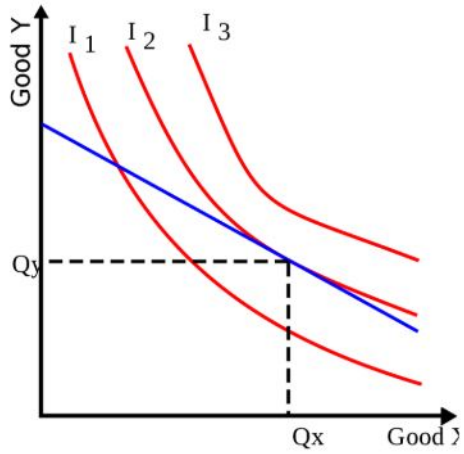


Figure 2: Pareto Fronts
Source: User "SilverStar", Wikipedia

extent of concavity in the objective space (given by the area between the Pareto front and the dashed line). This explains the fact that, although preferences could be convex, the consequences of such choices could backfire if the feasible set contains concavities.

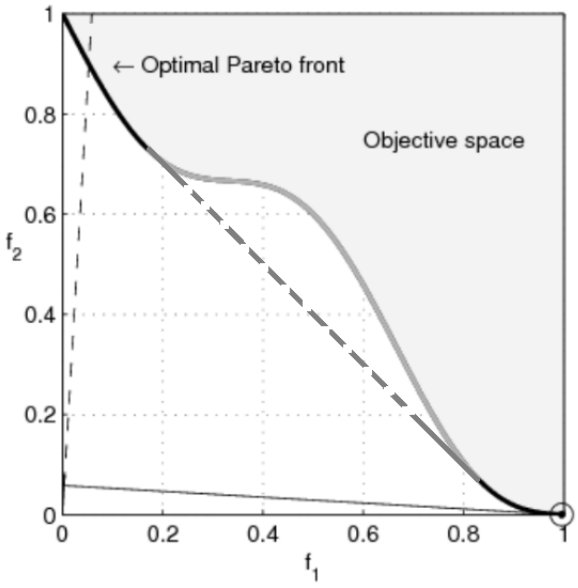


Figure 3: Concavity in the Pareto Front
Source: Guillaume Jacquenot, Wikipedia

## 1.3 Failure of Classical Portfolio Theory

Classical portfolio theory is a framework that was designed to build a portfolio/bundle of assets in such a way that the total risk harbored by the purchase of these assets is minimized. This is done by modeling the said risk as a convex function of the weights of these assets. However, this risk minimization technique does not always result in a good portfolio structure due to its innate non-agility to rare-events. Figure 4 shows the variation of the Dow Jones Industrial Av-


Figure 4: Dow Jones Industrial Average: 6 Month Window
Source: Google Finance

erage over a period of 6 months from September 2019 to March 2020 (The Dow Jones is a stock market index which measures the stock valuation of 30 major companies in the United States). From the graph, we can observe that there has been an approximately steady rise in the value until February 2020, after which, there occurs a sharp fall. If one started investing in September 2019, the classical portfolio theory would suggest continuously increasing investments owing to the increasing convexity in the bundle due to such decisions. But, a sharp drop in the stock values would immediately result in losses if this theory is strictly followed since the agent utility function (will be made precise in the next section) does not adapt to this "noisy" observation "quickly" enough to make changes in the portfolio allocation.

What if we sampled data at a higher frequency? Wouldn't that be one way of making the model agile/sensitive? Unfortunately, large market movements are not triggered unless there is a certain "critical concentration" [8, 9] of traders involved in decision making. In Figure 5, we observe

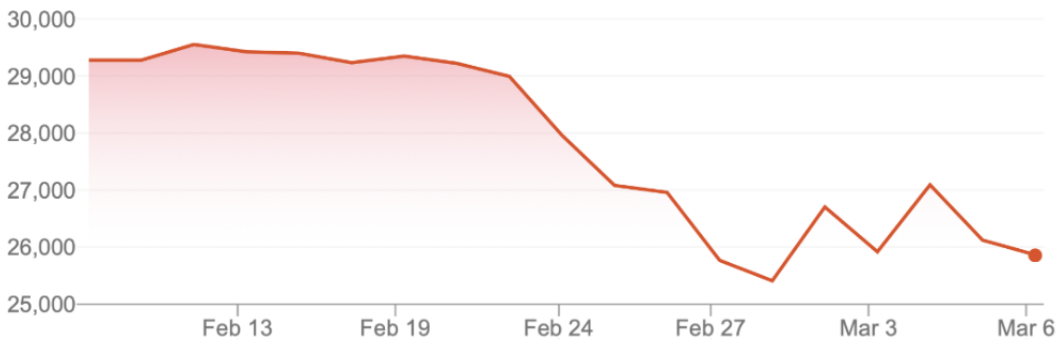


Figure 5: Dow Jones Industrial Average: 1 Month Window

Source: Google Finance

a steady Dow Jones value until February 23rd, after which, there is a sharp decrease. Therefore, no matter how quickly we sampled, we would still have the fleeting sense of stability until the 23rd (although experienced traders would have a ”hunch”, something that is not mathematically quantifiable as of now!). Following the classical portfolio model, continuing to hold onto these assets, say, on the 27th, would result in a lossy position.

The aforementioned examples clearly explain the failure of the classical portfolio theory in designing a robust structure for financial investment. A good portfolio is one which is risk-averse, but also capitalizes on possible returns. The classical portfolio theory is convex to losses, but is not concave to gains, thus clearly failing during the occurrence of non-observable events.

# 2 Risk-Averse Portfolio Allocation

## 2.1 Mathematical Model for Risk-Averse Investment

We shall explore the conclusion stated above in more mathematical detail now. The problem statement is as follows: Given historical price data, we must construct a portfolio allocation that gives minimum portfolio risk. Let us assume there exist $N$ assets and a window of $M$ days over which, we must build a portfolio. The inputs to our model that we consider here are the $N$ mean returns of individual stocks represented by the vector:

$$\hat{\mu} = (\bar{\mu}_1, \bar{\mu}_2, ...., \bar{\mu}_N)$$

We also construct a covariance matrix $\hat{\Sigma}$ consisting of variances of stock $S_i$ with respect to stock $S_j$.

$$\hat{\Sigma} = \begin{pmatrix} \rho_{11}(=\sigma_1^2) & \rho_{12} & \rho_{13} & .... & \rho_{1n} \\ \rho_{21} & \rho_{22}(=\sigma_2^2) & \rho_{23} & .... & \rho_{2n} \\ .... & .... & .... & .... & .... \\ \rho_{N1} & \rho_{N2} & \rho_{N3} & .... & \rho_{NN}(=\sigma_N^2) \end{pmatrix}$$

where

$$\rho_{ij} = \frac{\sum_k (S_{ik} - \mu_i)(S_{jk} - \mu_j)}{n-1}$$

The output we must compute is the weight (denoted by $\bar{w} = (w_1, w_2, ..., w_N)$) which determines the optimal portfolio allocation for a given window of evolution. The risk-averse investment strategy then becomes an optimization problem:

$$\bar{w}^* = argmin_{\bar{w}} \bar{w}^T \hat{\Sigma} \bar{w}$$

subject to the constraints:

$$\hat{\mu}^T \bar{w} \geq r_0$$

$$1^T \bar{w} = 1$$

where the scalar $r_0$ determines the minimum level of return we are expecting for the asset bundle. This problem can be further cast as a convex optimization problem provided that the matrix $\hat{\Sigma}$ is a symmetric positive semi-definite matrix. Clearly $\hat{\Sigma}$ is symmetric by construction. Furthermore,

it is known from classical statistics [10] that every positive semi-definite matrix represents a covariance structure for some multivariable distribution, and the converse is not true if and only if $\hat{\Sigma}$ is rank deficient. What this essentially means is that if $\hat{\Sigma}$ is not positive semi-definite then there are correlated assets in the bundle we have chosen, which we can then combine into a single bundle from the point of view of time evolution, and convert it into a positive semi-definite form.

## 2.2 Distribution of Price Changes

In order to implement the risk-averse mathematical model on real data, we consider the daily Open-High-Low-Close (OHLC) quotes for 120 symbols listed on the National Stock Exchange of India [14] for a time horizon of 6 years (2014-2020). According to the standard Brown-Bachelier theory [11], the volatility of a price time series is modeled as akin to that of particle diffusion, that is, the Wiener stochastic process [12]. Using a fundamental mathematical property of these processes known as the Itô lemma [13], we can infer that the price differences follow a log-normal distribution i.e., the logarithmic daily return should be modeled as a Gaussian. However, the occurrence of non-observable events causes significant deviation from this assumption and is clearly depicted by the following histogram charts of a few specimen (the log-normal distribution that is fit to the data is the black curve):

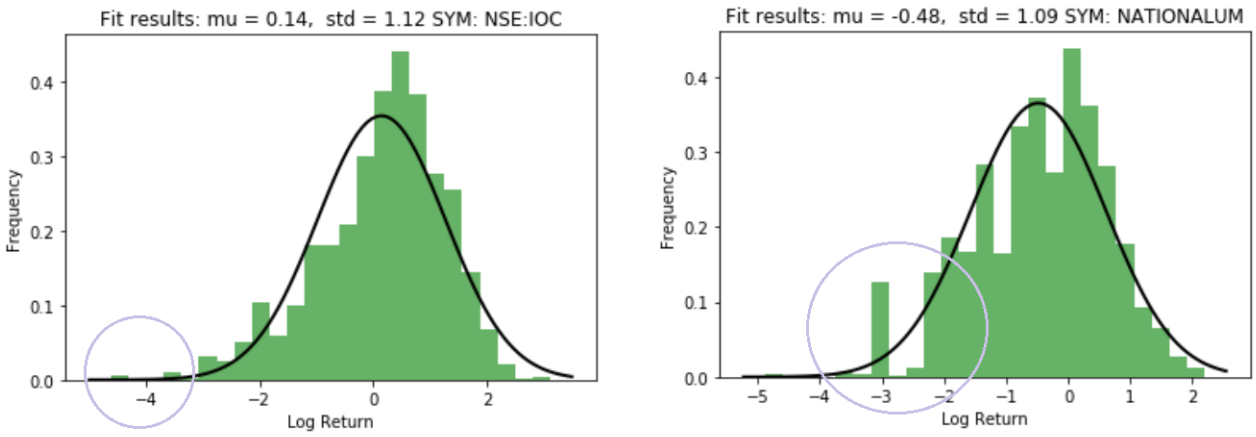


Figure 6: Non-trivial $3\sigma$ and higher deviants. Also peaks are not symmetric.
Source: NSE BhavCopy (see [14])

We included this section just to clarify that the data-sets we are dealing with do include such historical events that had insanely low estimated probabilities according to the de-facto log-normal model. This effect compounds when we deal with portfolio variances and interestingly scales with the time-horizon selected. From the point of view of fitting a more robust distribution, much work has been done in the field of such "heavy-tail" densities, we refer the reader to [15, 16] for more details since it is beyond the scope of the present discussion.

## 2.3 Going Long (Invest and Forget)

With the rather simple looking convex optimization problem for risk aversion, we can actually derive some significant insights with respect to the "non-agility" of classical portfolio theory. We first conduct a very hypothetical experiment, wherein we assume that we are in 2014 and have perfect information about the evolution of the stock prices in the next six years. We then use four different flavors of the original optimization problem and evolve the optimal portfolio over the same six years with an initial capital of $100,000$ units of currency (≈$2000).

**Total (Weak) Risk Aversion**: Minimize $\bar{w}^T\hat{\Sigma}\bar{w} - \bar{w}^T\hat{\mu}$ subject to $w_i > 0, \sum w_i = 1$
**Stochastic Portfolio**: The weight vector $\bar{w}$ is selected randomly over the probability simplex $w_i > 0, \sum w_i = 1$ as a model for an irrational allocation choice. Thus, no actual optimization is performed in this case.
**Break-even Portfolio**: Minimize $\bar{w}^T\hat{\Sigma}\bar{w}$ subject to $w_i > 0, \sum w_i = 1, \bar{w}^T\hat{\mu} \geq 0$. This is a specialization of the classical model where we are demanding that at the least we break even (only lower bounded).
**Returns-only Portfolio**: Minimize $-\bar{w}^T\hat{\mu}$ subject to $w_i > 0, \sum w_i = 1$. This is a degenerate case included here only for the sake of completeness. The minimization is being done over a probability simplex of $w_i$ and the asset with the maximum $\bar{\mu}$ will be selected as the only optimal choice.

When we solve each of the convex problems we described above we will get a portfolio distribution vector, which we can then use to decide initial allocations over the capital provided and evolve these allocations over the entire time-horizon. To clarify the title of the section, we intend to select this allocation in 2014 and wake up in 2020 to check the results, nothing more.

The not-really-surprising result is as follows:

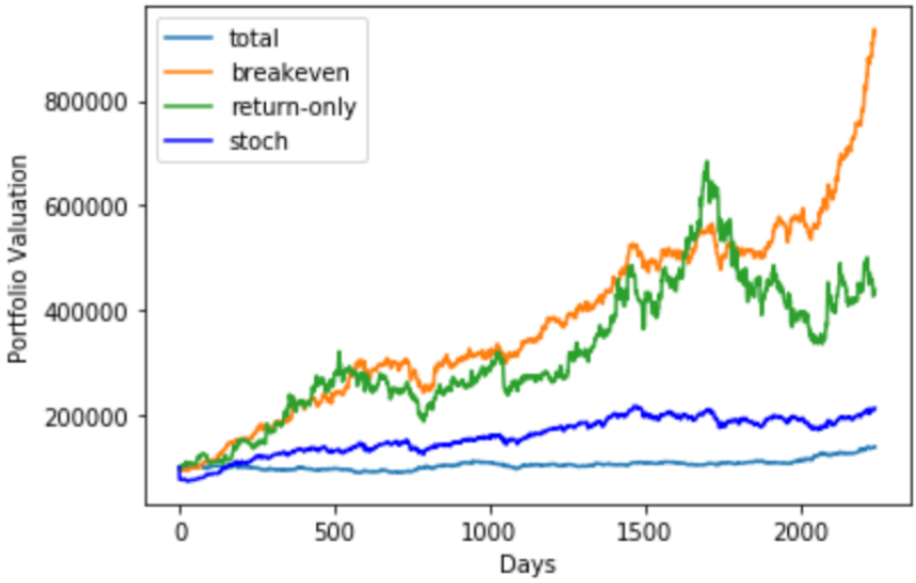


Figure 7: Comparison of long-term evolution of classical portfolios

The weak portfolio optimization by definition has no strict return level guarantee, and therefore would do whatever it takes to minimize the portfolio variance regardless of the amount of profit/loss made due to the evolution. Therefore, it actually performs worse than a randomly selected weightage allocation across the assets, but even with perfect information these two models are unable to capitalize on market movements (flat growth rate). The return-only portfolio obviously mimics the asset with the highest average return and by definition this would do better than the bundle of assets in the previous two models. However, the break-even portfolio does its job really well in the presence of perfect information, making a staggering 800% growth in the 6 year time frame.

At this point, we would like to remind the reader that the above experiment was done just as a zeroth order hypothetical study, and we see that barring from a few days, the break-even portfolio dominates the rest of the strategies in terms of valuation.

## 2.4 Iterative Risk Aversion: Effect of Partial Information

A realistic scenario would be to start investing in 2014 but have only a finite time horizon of $W$ days to acquire information to reallocate/restructure the portfolio. We will first describe the algorithm here and then will discuss its implications and simulation results.

**Iterative Break-Even Optimization**

**1. Initialize** the weight vector randomly over the probability simplex $w_i > 0, \sum w_i = 1$ as a starting point. Granted this might bias the evolution if we randomly end up with a good vector, we repeated the experiment multiple times to see that the chances of this event are very, very low (for a vector on $\mathbb{R}^{120}$).

**2. Wait** for the random portfolio to evolve until $W$ days, and also store the price data/market state for these $W$ days in a database $\mathcal{D}$ i.e., $\mathcal{D} = \mathcal{D} + \mathcal{M}_W$.

**3. Optimize** the weight vector based on the database $\mathcal{D}$ with the break-even strategy.

**4. Reallocate** the remaining capital into the new allocation obtained at the end of step 3. Then return to step 2.

What the algorithm essentially does is that it allows the trader to incrementally increase his/her knowledge about the condition of the market evolution to make a decision. This pins the expectation of time-symmetry of market evolution to a window of size $W$ and any sort of deviations from this assumption can be rectified at the end of the window. Ideally, we would expect the agent to make better and better decisions as he/she collects more information, but we also constrain the amount of capital the agent can use. This leads to some interesting dynamics which are shown below for three specimen sizes of $W$ in Figure 8.

There are three key take-aways from the plots. Firstly, we note that as we reduce the sampling window size $W$ (or increase the sampling frequency, this will be our handle to segue into signal processing) we effectively simulate a "paranoid" trader who falls victim to the classic problem of ruin in games of chance [17]. As the time progresses, he/she is almost always left with lesser capital than the earlier iteration, thus successively setting the break-even return level to a lower

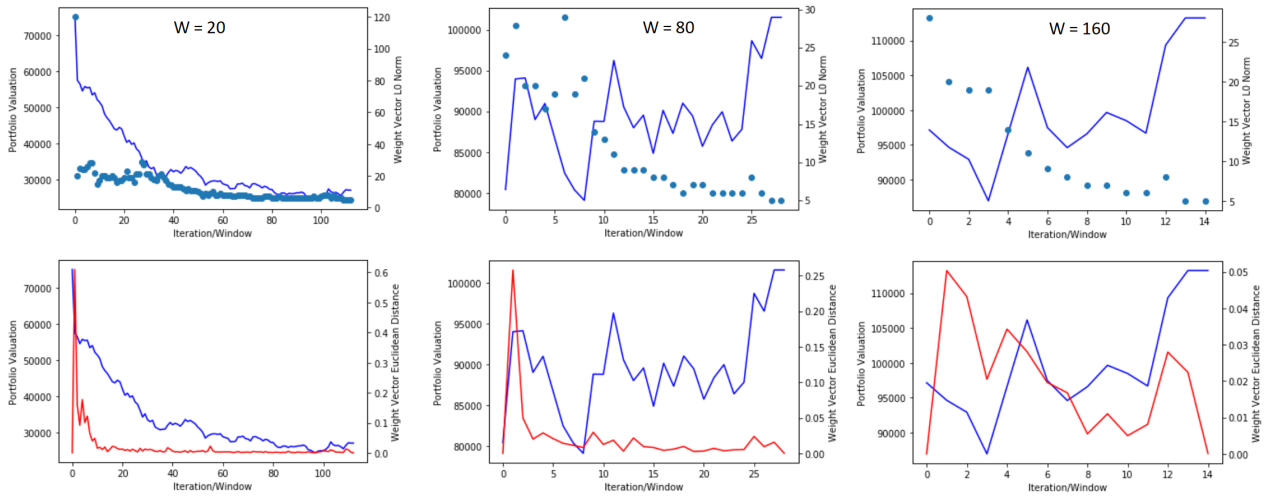


Figure 8: Comparison of short-term evolution of classical portfolios

point. His/her stagnation in decision making is evident from the overlayed charts which represents the number of active assets involved in that window's evolution (top) and the $L_2$ norm between successive portfolio weight vectors (bottom).

Secondly, as the sampling window is increased the investor now has more information to estimate the risk better, thus he/she has a non-zero chance of actually regaining the initial investment despite the turbulence encountered. We find empirically that this probability monotonically increases with $W$ and upon performing a Monte-Carlo evolution over all the possible window sizes $W = 1$ to $W = 2241$ days (or by using the Rolle theorem on monotonic functions) we find that there is a fixed window size $W^*$ for a selection of assets where the portfolio breaks even with high probability.

Lastly, once the sampling window crosses $W^*$ one can actually realize significant positive returns (> 50%) provided one waits for prohibitively long windows ($W > 600$, i.e., re-allocate once every two years!) and this shows the epistemological problem of perfect information in full glory. Contrasting this with the result of the last section (800% in 2241 days) we see that for classical portfolios to truly suppress impending risk, they will need more data than what can be available for a reasonable sampling rate, thus making this policy "non-agile".

# 3 Dynamic Mode Decomposition

From the earlier section, we see the need for a better investment policy when it comes to higher sampling frequencies (lower $W$) and utilize the Koopman mode decomposition [18] concept to see if the portfolio vectors show any mutual oscillatory patterns during small windows. The following sections are our understanding of the paper [3], along with the results/caveats when implemented on our dataset.

## 3.1 Mathematical Model

Dynamic Mode Decomposition (DMD) is an "equation-free" method in which the complex system dynamics can be reconstructed through iterative sampling of the data. The DMD approach utilizes low-rank features in the historical price time series and in [3] the authors subsequently use the least-square fit /minimum-energy model to predict the states of the system in the near future. Let us consider $N$ assets $S_1, S_2, ..., S_N$ over a prior $(W-1)$ day window. We can arrange the historical price information into a matrix of dimensions $N \times W-1$:

$$\hat{X}_1^{W-1} = \begin{pmatrix} \bar{x}_1 & \bar{x}_2 & .... & \bar{x}_{W-1} \end{pmatrix}$$

where each $\bar{x}_i \in \mathbb{R}^N$ is the price vector array for day $i$. Our aim is to predict the the valuation of the stocks on the $W^{th}$ day, which according to our notation would be $\bar{x}_W$. Since we are dealing explicitly with time evolution of multi-agent systems, we introduce a term from the Koopman-Hamiltonian formalism called the Koopman operator $\mathcal{A}$ which is a linear, infinite-dimensional operator used to obtain a linear mapping for the analysis of the underlying nonlinear dynamical system. This was (probably) done to ensure that the DMD method does not explicitly linearize the system, but instead linearizes the observables dependent on the state of the system (moving averages, volatility, et cetera), thereby constructing a linear mapping to approximate the system dynamics directly. Mathematically, for any given day $j$, we want the Koopman operator to give us the vector for the next day, that is:

$$\bar{x}_{j+1} = \mathcal{A}\bar{x}_j$$

Using the conformal representation of $\hat{X}_1^{W-1}$, we can equivalently argue that:

$$\mathcal{A}\hat{X}_1^{W-1} = \mathcal{A}\begin{pmatrix} \bar{x}_1 & \bar{x}_2 & .... & \bar{x}_{W-1} \end{pmatrix} = \begin{pmatrix} \mathcal{A}\bar{x}_1 & \mathcal{A}\bar{x}_2 & .... & \mathcal{A}\bar{x}_{W-1} \end{pmatrix} = \begin{pmatrix} \bar{x}_2 & \bar{x}_3 & .... & \bar{x}_W \end{pmatrix} = \hat{X}_2^W$$

Since the matrix $\hat{X}_1^{W-1}$ is not square in general (more days than assets in the portfolio), there is an expectation that the matrix may not be full rank, and if this is the case, the SVD will have the following structure:

$$\hat{X}_1^{W-1} = \mathcal{U}_{N\times K}\Sigma_{K\times K}V^{\dagger}_{K\times W-1}$$

where the dimensions are mentioned in the subscripts of the respective factors. Substituting this into the previous equation, we have:

$$\mathcal{A}(\mathcal{U}\Sigma V^{\dagger}) = \hat{X}_2^W$$

Pre-multiplying with $\mathcal{U}^{\dagger}$, we bring the LHS to what is mathematically known as a "similarity" transformation:

$$\tilde{\mathcal{A}}_{K\times K} = \mathcal{U}^{\dagger}\mathcal{A}\mathcal{U} = \mathcal{U}^{\dagger}\hat{X}_2^W V\Sigma^{-1}$$

This establishes a similarity equivalence between $\tilde{\mathcal{A}}$ and $\mathcal{A}$, and this essentially entails that they share $K$ eigenvalues (which are our dynamic modes). This can be visualized by working backwards from $\tilde{\mathcal{A}}$, to wit, we have by the very definition of the spectrum:

$$\tilde{\mathcal{A}}\bar{\nu}_k = \mu_k\bar{\nu}_k$$

Substituting the definition in terms of $\mathcal{A}$, we have:

$$(\mathcal{U}^{\dagger}\mathcal{A}\mathcal{U})\bar{\nu}_k = \mu_k\bar{\nu}_k$$

Reordering the terms, we see that this is actually the eigenvalue equation for the original Koopman operator in disguise:

$$\mathcal{A}(\mathcal{U}\bar{\nu}_k) = \mu_k(\mathcal{U}\bar{\nu}_k)$$

Therefore, we have indirectly uncovered $K$ of the (possibly infinite) eigenvalues of $\mathcal{A}$ which are identical to $\mu_1, \mu_2, ...., \mu_K$ (not to be confused with the mean returns notation of the earlier section, we are trying to be consistent with [3]). The eigenvectors clearly live in $R^K$ and we use the $K \times N$ dimensional unitary matrix $\mathcal{U}$ to back-project these to $R^N$:

$$\bar{\psi}_k = \mathcal{U}\bar{\nu}_K$$

In conclusion, we have tamed the problem by reducing the original $N$ dimensional analysis to a (hopefully) smaller $K$ dimensional vector, and in analogy to the Laplace-Fourier series, where we are able to represent a complicated signal as a summation of exponential-sinusoid eigenmodes, we are able to do the same with respect to the price vector evolution:

$$\bar{x}_{DMD}(t) = \sum_{j=1}^{K} b_j\bar{\psi}_j exp(\omega_k t)$$

where $\omega_k = \frac{\log \mu_k}{dt}$ and $dt$ is the sampling rate of the original matrix. In our case it is $1$ day so we set $dt = 1$ and supply the $t$ argument in days instead. Making this substitution one can see that this is nothing but a convolution of the eigenvectors $\bar{\psi}_k$ with the summation kernel given by $b_k\mu_k^t$. In general $\omega_k$ (and therefore $\mu_k$) is complex and if $\omega_k = \alpha_k + \iota\beta_k$, we have:

$$\bar{x}_{DMD}(t) = \sum_{j=1}^{K} b_j \bar{\psi}_j exp(\alpha_k t) exp(\iota \beta_k t)$$

Thus, the real part of the eigenvalue determines the so-called "compounding rate of growth", while the complex part can be related to the oscillatory behavior/volatility of the asset. This distinction was not described in [3], and instead, in that analysis the imaginary part was discarded for a more "growth-oriented" portfolio. The only unknowns we have left to discover are the mode "amplitudes" $b_j$ (which can themselves be complex!). For this we simply substitute $t = 0$ (this is tantamount to solving an initial value problem):

$$\bar{x}_{DMD}(0) = \bar{x}_1 = \sum_{j=1}^{K} b_j \bar{\psi}_j = \hat{\Psi} \bar{b}$$

where the (fat) matrix $\hat{\Psi}$ consists of the $K$ $N$-dimensional vectors $\bar{\psi}_j$ as columns i.e, is of dimensions $N \times K$. The authors in [3] use the Moore-Penrose pseudoinverse to obtain $b = \hat{\Psi}^{+} \bar{x}_1$, which minimizes the $L_2$ norm (more colloquially known as the "minimum energy" solution).

## 3.2 Results and Caveats

If we work backwards in the derivation of the earlier section, we crucially depend upon the matrix $\hat{X}_1^{W-1}$ to be rank deficient for the chosen value of $W > N$. If this condition is not met, say the matrix is full rank, then the authors of [3] claim on page 6 that the method immediately fails. While it may not be so obvious as to why this restriction even exists, it was most unfortunate that many of our empirical asset matrices were full rank. Since we pick the assets randomly, getting a non-full rank price matrix was virtually impossible.[1]. In [3], the authors demonstrated their trading strategy by picking market sectors first and then picking the assets per sector, and therefore would have achieved rank deficiency in this way. But we feel that using extraneous information (i.e., not mathematically integrable) about the market class does not make this a solution based

[1]For the record, we feel that doing anything with price level time series is not suggested since stocks with higher average cost dwarf out the impact of the rest of the portfolio. Instead, utilizing the price difference would have made more sense since focusing on making returns over the initial capital makes the problem more tractable

on pure time-series analysis. However, we could induce rank deficiency by making $X_1^{M-1}$ a tall matrix (column rank deficiency). To this end, we performed this interesting analysis of scanning $W < N$ and $W > N$ and plotting the following "rank distribution" histograms (for 30 randomly selected companies) across all $W$ sized windows in the 6 year data set:

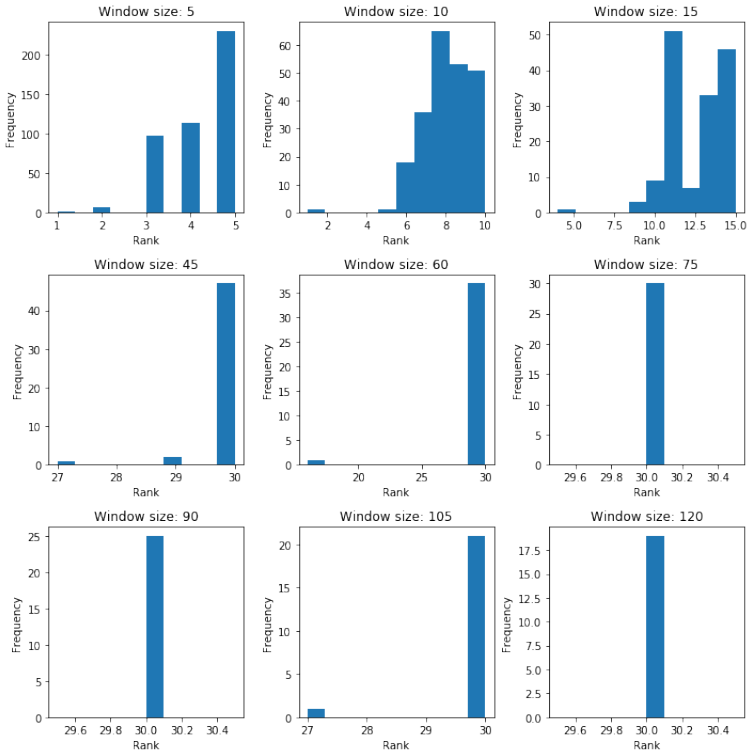


Figure 9: Rank distribution over window sizes for $N$ = 30

Thus, we can see that when the sampling frequency is increased, the likelihood of getting an overall rank deficient structure is higher. Indeed, even contrary to the authors' claim of rank deficiency, we see reasonable estimation for ***certain windows*** where $W > N$ (as shown in Figure 10).[2] We instead encounter an error catastrophe when we try to simulate the evolution for the induced rank-defective matrix with $W = 10$ (as shown in Figure 11). Due to the exponential dependence, we see that the DMD estimate is always either strictly greater or strictly lesser than that of the actual values, even for an evolution over a smaller future window (as can be seen in

[2]For the sake of demonstration we did the computation by taking the complex eigenvalues in the summation and we considered the real part of the final result instead.

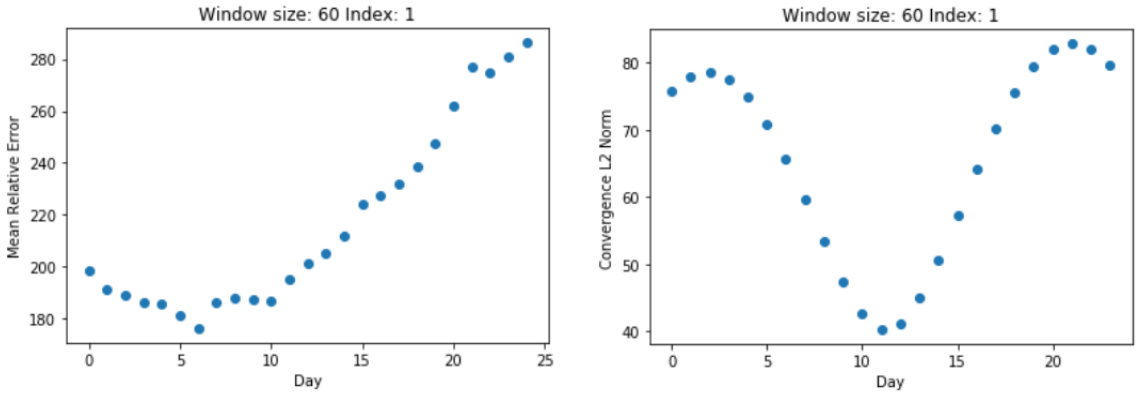


Figure 10: DMD constructed for $W = 60$ and evolved over the next 25 days

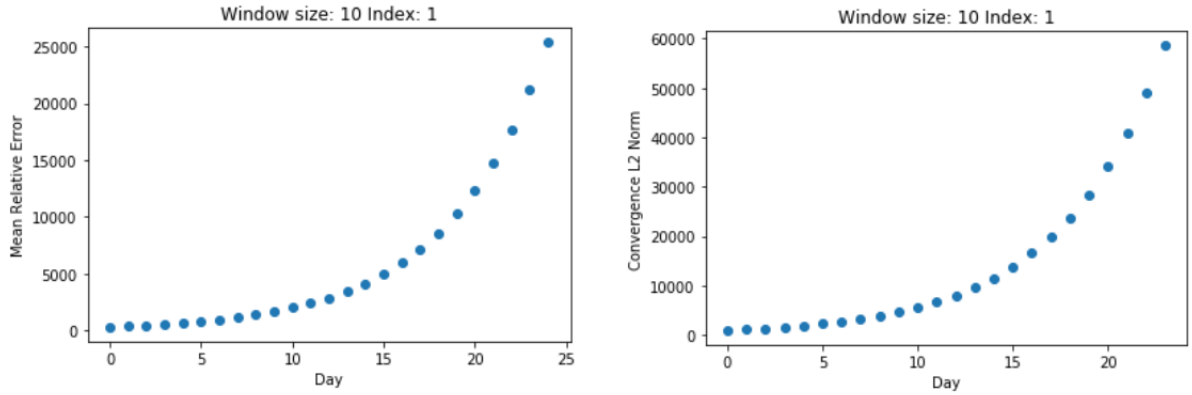


Figure 11: DMD constructed for $W = 10$ and evolved over the next 25 days.

Figure 12) and we are therefore forced to conclude that this technique is in-fact worse than the classical theory if random asset portfolios are to be constructed. One way to fix this mishap is to somehow renormalize the exponential blow-up, and we take inspiration from Benoît Mandelbrot's intuitive idea of "trading time" [15, Chapter 11] to set the denominator of $\omega_k = \frac{\log \mu_k}{dt}$ i.e., $dt > 1$. The results after this fix for $dt = 10$ give much better images (see Figure 13). However, the study was well worth it since it shows the core importance of manipulating the price data as a linear transformation to make any feasible progress towards preventing ruin for smaller $W$. In the last few sections of the report we will detail our investigations into a novel approach for tackling non-observable events so that "agile" decisions can be made using of sharp market movements.

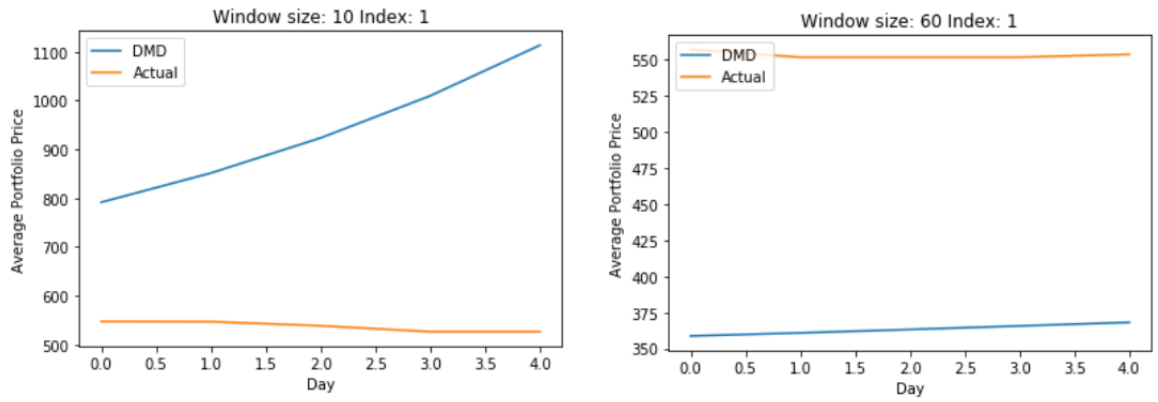


Figure 12: The monotonic growth in Figure 10 from up close, and for $W = 60$

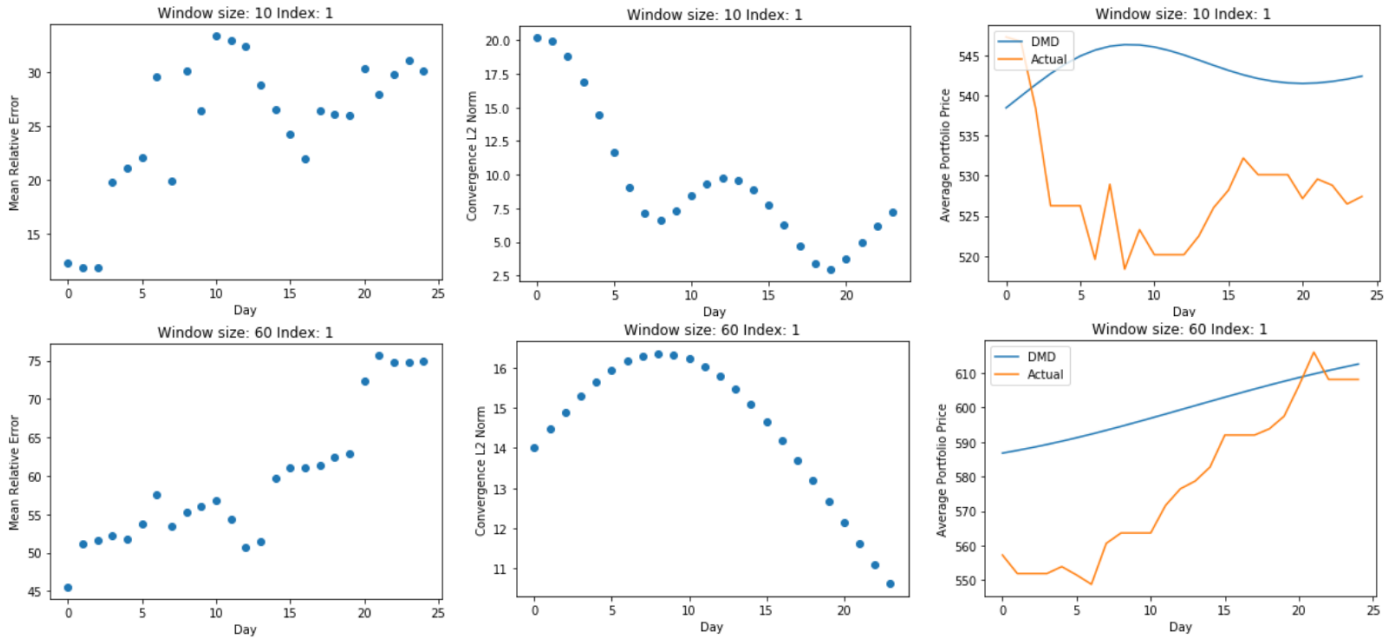


Figure 13: After slowing down time 10X, we see some sensible trend estimation

# 4 Rare Events and Sparse Signals

## 4.1 An Insight

We argue that the concept of a sparse time series plays a central role in any portfolio strategy that wants to make use of rare/low-likelihood events. Due to the lack of any statistical estimator in the presence of non-observability, it makes sense to model the inter-event times of these rare events

by a probability density whose $E[X] = \infty$ [4]. This essentially tells us that once a rare event takes place, by their very definition, we may not be in a position to predict the next occurrence purely based on historical information. In other words, if we were to chop up time evolution in windows of $W$, the number of days which feature a local optimum in the price returns would be far fewer than $W$.

This makes us consider estimating the "impact"/"amplitude" of the window data on the day whose price return is to be calculated, since we already know that the extreme case would be a sparse signal representation. That is, given[3]:

$$\hat{X}_1^{W-1}\bar{b} = \bar{x}_W$$

can we find a solution for $\bar{b} = (\hat{X}_1^{W-1})^{-1}\bar{x}_W$ such that $\bar{b}$ is sparse? The field of optimization that deals with minimizing $|\hat{X}_1^{W-1}\bar{b} - \bar{x}_W|$ for sparse $\bar{b}$ is known as "Compressive Sensing"(CS), best described by the seminal paper [19] but a hazy idea of such a framework was already established in earlier papers (called the Logan phenomenon) such as [20], [21]. For this concept to be applicable, we need $\hat{X}_1^{W-1}$ to obey the so-called Restricted Isometry Condition (RIC), which states that there must exist a $\delta \in [0, 1]$ such that for every sub-matrix $\tilde{X}$ constructed by taking only column subsets of size $p = 1, 2, .., W-1$, and for every $y \in R^p$, we have:

$$(1-\delta)||b||_2^2 \leq ||\tilde{X}b||_2^2 \leq (1+\delta)||b||_2^2$$

In such situations, we say $\hat{X}_1^{W-1}$ is $W-1$ compressible. Since it is not possible to check for every $y \in R^p$ the problem of establishing the RIC is non-trivial, and is infact proven to be NP hard [22]. We therefore consider a milder variant where we accept a weaker RIC when $||\bar{x}_W||_2^2 \approx ||\hat{X}_1^{W-1}\bar{b}||_2^2$. For now, we will defer checking the compliance for various window price matrices $\hat{X}_1^{W-1}$ to a future report and demonstrate the core ideology and the convex optimization form of the CS framework.

[3]We are working with *price return matrices* but we use the same notation as the earlier section for the sake of continuity.

## 4.2 Denoising

Historically, the problem of retrieving sparse signal content from a noisy data set has been solved many times under various names. The classical method of using the Moore-Penrose Pseudoinverse [23] is known as the "minimum energy" solution because the solution $\bar{b} = (\hat{X}_1^{W-1})^+ \bar{x}_W$ minimizes the function:

$$f(\bar{b}) = ||\hat{X}_1^{W-1}\bar{b} - \bar{x}_W||_2^2$$

However, this technique fails to be reliable if either the matrix $(\hat{X}_1^{W-1})^\dagger(\hat{X}_1^{W-1})$ or $(\hat{X}_1^{W-1})(\hat{X}_1^{W-1})^\dagger$ are ill-conditioned. In-order to reduce the dependence on the conditioning of the input, and to penalize overfitting to low amplitude data points, one could change the objective function by adding the so-called "hyperparameter":

$$f(\bar{b}; \lambda) = ||\hat{X}_1^{W-1}\bar{b} - \bar{x}_W||_2^2 + \lambda||\bar{b}||_2^2$$

This is known in the world of statistics as Tikhonov regularization [24, 25], and by looking for the stationary points of $f$, we get the optimal solution as the one which satisfies:

$$\hat{X}_1^{W-1}\bar{b} = \frac{\hat{x}_W}{1+\lambda}$$

In most practical implementations, we see that the function $f$ is convex and we may utilize the CVXOPT Python package [26] and fix various levels for $\lambda$ to see the efficacy of the signal recovery. However, we are now liable to estimate one more parameter: $\lambda$, which is traditionally done using cross-validation over the historical corpus, a bad idea when you have non-observability at the fore. Nevertheless, the effect of the hyperparameter in this case is to make small components of the estimate of $\bar{b}$ smaller by a factor of $1 + \lambda$ thereby somewhat "sparsifying" better than the pseudoinverse.

If one instead switches the penalty term to the following form:

$$f(\bar{b};\lambda) = ||\hat{X}_1^{W-1}\bar{b} - \bar{x}_W||_2^2 + \lambda||\bar{b}||_1$$

that is, we swap out the $L_2$ norm with the $L_1$ norm, we get the following as the optimal solution at the stationary point:

$$\hat{X}_1^{W-1}\bar{b} = \begin{cases} \bar{x}_{W_i} + \lambda & \bar{x}_{W_i} < -\lambda \\ 0 & |\bar{x}_{W_i}| < \lambda \\ \bar{x}_{W_i} - \lambda & \bar{x}_{W_i} > \lambda \end{cases}$$

The "sparsification" process now entails amplifying only those components of the signal that lie beyond a certain amplitude level $\lambda$, analogous to the band-pass filter concept in RF circuit theory. This is known as the LASSO regression technique [21, 27] in classical statistics and like the Tikhonov estimate, it too suffers from the variability in the hyper-parameter $\lambda$.

## 4.3 Basis Pursuit and the L1 Norm

The most natural question that we may now ask is whether we can free ourselves from estimating $\lambda$, and a mathematically feasible way was demonstrated by Chen, Donoho et al. [28] in what is now called as the problem of "Basis Pursuit". Concisely put, for our purposes, we can have a parameter-free denoising by simply minimizing the penalty term itself exclusively. We therefore have our objective function:

$$f(\bar{b}) = ||\bar{b}||_1$$

subject to the following conditions: $\hat{X}_1^{W-1}\bar{b} = \bar{x}_W, b_i > 0 \forall i$. The use of the $L_0$ norm is primarily motivated in [19] as the best choice for sparse recovery but due to the non-analytic nature of the $L_0$ "norm" it becomes computationally (exponentially) expensive to minimize. Instead, the $L_1$ minimization, apart from being convex, is shown [19] to exactly recover the initial signal (of

course, assuming RIC compliance). Moreover, by making the following substitution [29], we see that we can actually reduce this problem to a linear program:

$$f(u_i, b_i) = \sum_i u_i$$

where

$$b_i - u_i \leq 0$$

$$-b_i - u_i \leq 0$$

$$\hat{X}_1^{W-1}\bar{b} = \bar{x}_W,$$

We use the pyCSalgos package [30] as an implementation of the primal-dual technique for solving the LP [31], and also use CVXOPT to compare the performance and behaviour. We select $W = 100, N = 30$ and run this optimization to see the following pattern for the "impact"/"amplitude" coefficients $b_j$ over an arbitrary window in the historical corpus, as shown in Figure 14.

We see that there are distinct peaks where $b_j$s are locally maximal and almost zero at the remaining places. Intuitively, this gives the motivation of an indicator that estimates the level and timing of such movements, but such an indicator is only useful if it lives in the **causal** proximity of the non-observable event, something we call for the sake of terminology as a "pre-cognition". To clarify, we want our novel framework to give a measure of the time-to-impact from a pre-cognitive event so that we can focus on this smaller window to make a decision. Due to the rather highly localized nature of these peaks, the condition of the "peak" preceding a market movement may not be well defined and may not exist in many of the peaks shown above, and we explore a more popular algorithmic alternative that has, of late, found widespread use in image denoising problems.

## 4.4 Total Variation Technique

We make the switch to the following new objective function that features a hyperparameter $\mu$ which, in this case, does not require any explicit estimation but has significance in terms of

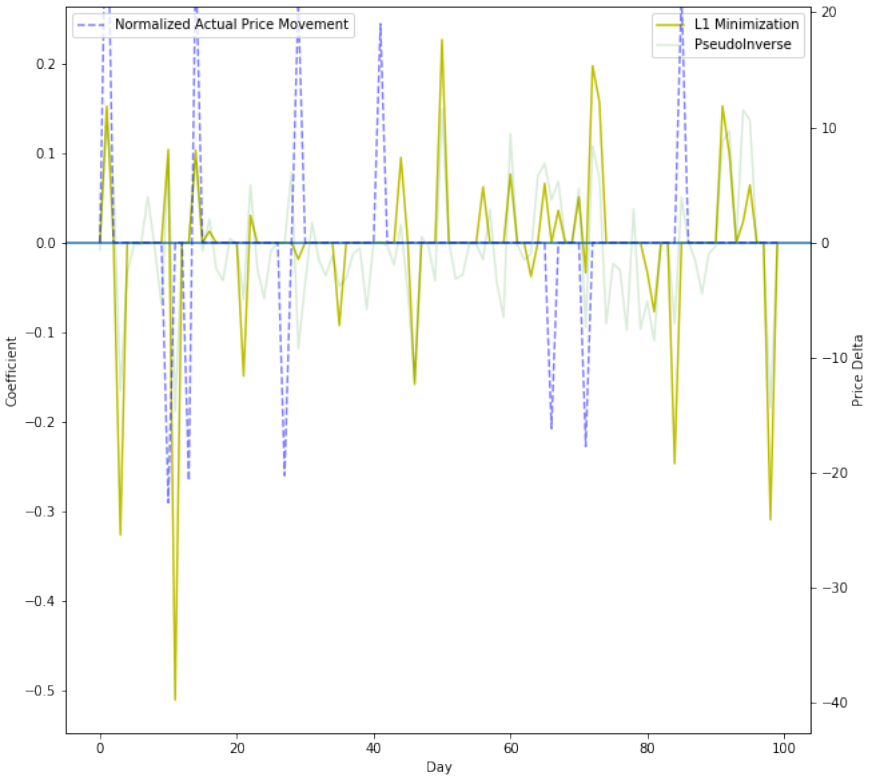


Figure 14: Comparing L1 Basis Pursuit versus actual market movement in the window (right axis is price change level)

smoothing out the "peaks" we discussed in the earlier section.

$$TV(\bar{b}) = \sum_i \mathcal{D}_i \bar{b}$$

where $\mathcal{D}_i$ is the discrete gradient (subtract-after-shifting-index) on the grid described by $\hat{X}_1^{W-1}$. Obviously, we enforce the constraint $\hat{X}_1^{W-1}\bar{b} = \bar{x}_W$. While this construction may not have any significant analogy in terms of the current context, an extension called TVAL (Total Variation with Augmented Lagrangian)[32, 33] has shown success in many recent projects involving de-noising and image reconstruction [34, 35]. This modification converts our objective function to:

$$f(\bar{b}; \mu) = TV(\bar{b}) + \frac{\mu}{2}||\hat{X}_1^{W-1}\bar{b} - \bar{x}_W||_2$$

The authors of [33] have also provided a MATLAB implementation of their algorithm [36] which we utilize to show the effect of increasing $\mu$ (we consider $W = 100, N = 30$ again). As can be seen

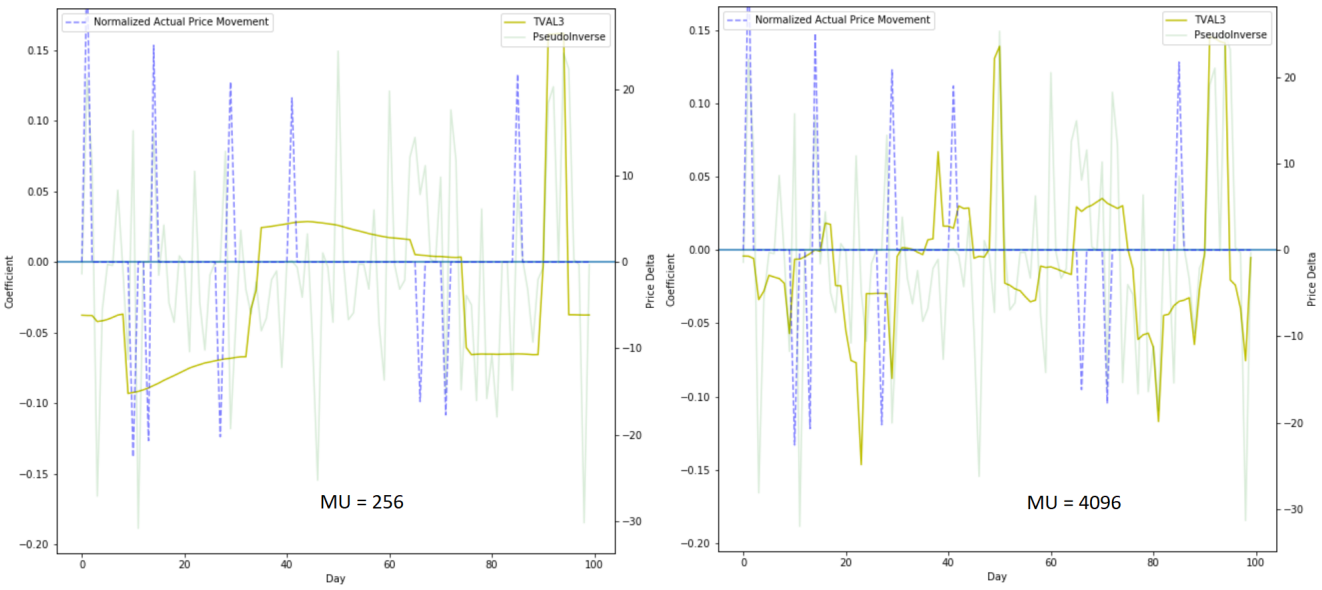


Figure 15: Running TVAL3 Denoising with different hyperparameter values

in Figure 15, the reconstruction gets noisier as the hyperparameter $\mu$ is increased, and actually degenerates to the basis pursuit solution at abnormally large values. However, we can observe that for reasonable settings we actually get some sub-windowing around a market movement that finally allows us make a concrete definition of ”pre-cognition”.

## 4.5 Pre-Cognition and Meta-Distributions

If we zoom in closer into the ”zero-crossings” of the TVAL result, rather than the peaks, we see that we can now definitively state what a pre-cognition actually means. We present Figure 16 to tie up the notation we have introduced so far.

Given a sampling window size $W$, if we are able to estimate when a ”pre-cognition”/”zero-crossing” takes place we can build on this estimate to construct ”meta-distributions”: probability densities that measure the time to impact given a zero-crossing has already taken place. For the sake of segregating market improvements and dips, we construct two distributions for estimating the time to a positive ($P_L$) and negative ($P_S$) impact respectively. In addition, we can construct two more useful densities from the historical corpus: the distribution of the number of zero-crossings in a window of size $W$ ($P_N$) and the time between two pre-cognitive events inside a

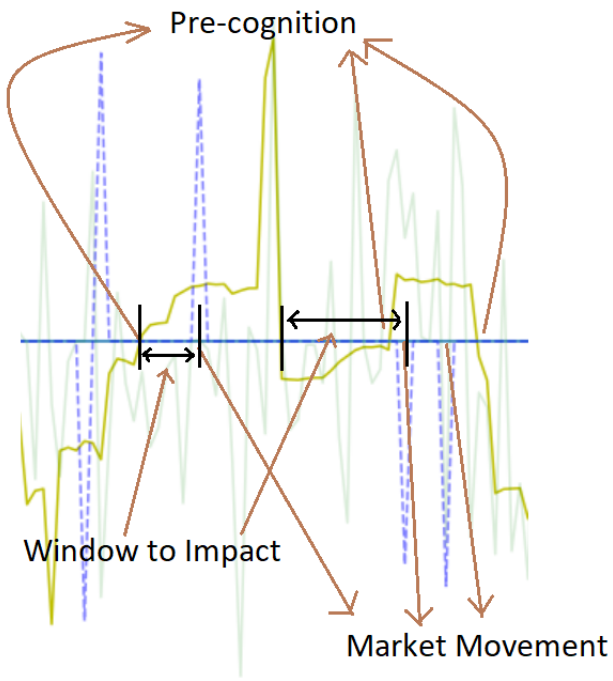


Figure 16: Anatomy of a market movement

window ($P_I$). Sample histograms for these (hidden) densities are shown in Figure 17.

To utilize these objects, we need to construct a generator/analytic approximation to the true empirical CDFs and from [4] we conclude that the next best alternative to a distribution with divergent moments is that of a Gaussian with local perturbations. This concept is known as Kernel Density Estimation, where one wants to construct an envelope distribution given a histogram for the empirical CDF [37]. The idea is quite simple: select a Gaussian with a certain bandwidth $B$, and add it as a "bump" in the final analytic density for every x-coordinate on the histogram, but we scale the amplitude of the "bump" with the height of the histogram bar. Running this routine on our histograms shown earlier, we get the continuous estimations of the densities described in Figure 17, shown in Figure 18.

The theoretical justification for using approaches that induce continuity can be done by using well known statistical bounds. Fundamentally, the Glivenko-Cantelli theorem states that any Kolmogorov density constructed on an empirical CDF (sum of indicator functions, the mathemat-

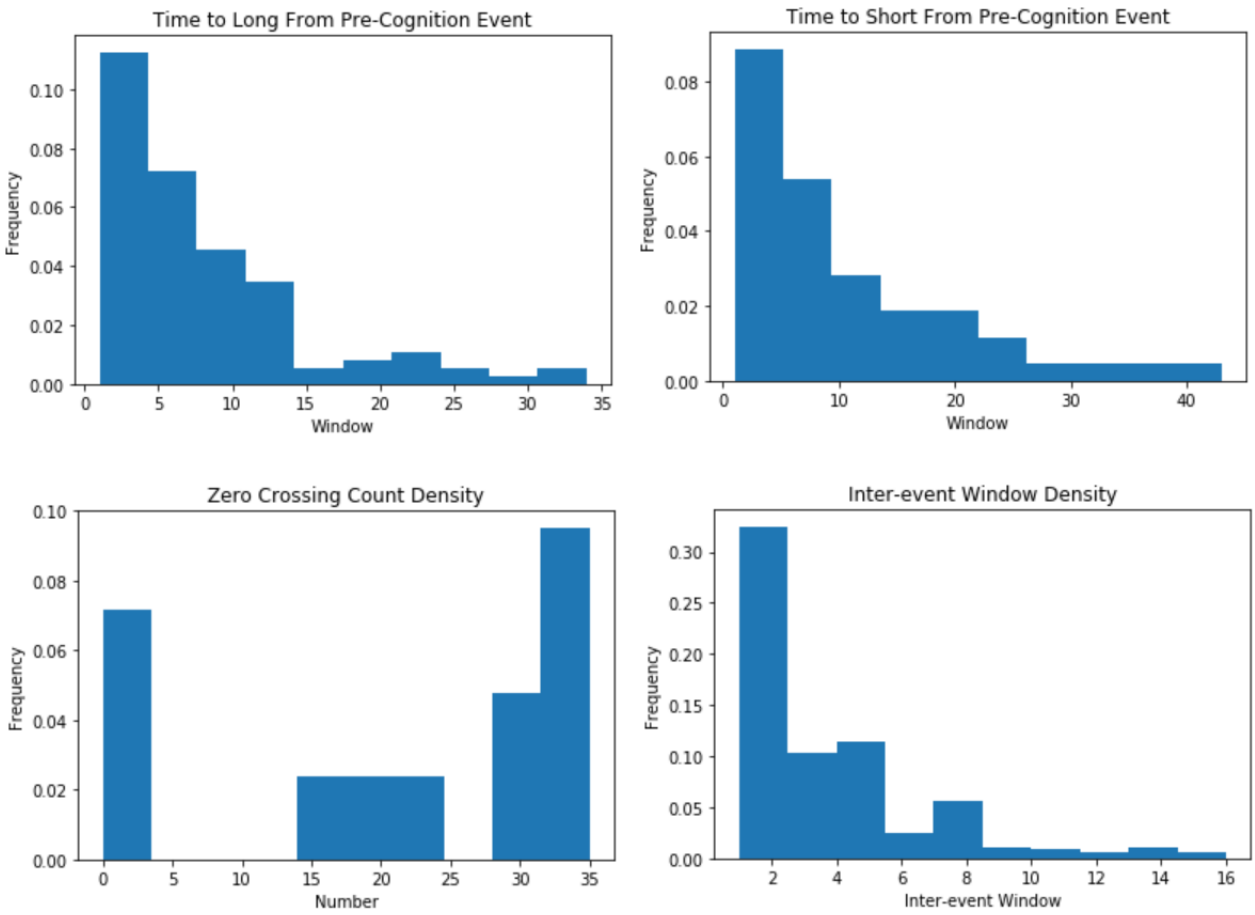


Figure 17: Specimen meta-distributions for $W = 100$ constructed over the first 1000 days of the corpus

ical manifestation of a histogram plot) converges almost surely as the density is made finer. The Dvoretzky–Kiefer–Wolfowitz[38] inequality provides a Chernov-style bound to the Glivenko-Cantelli result, and can be used to estimate the confidence levels.

# 5 Advanced Heuristics

With the KDE estimated generators for the meta-distributions readily available, we can finally describe our attempt at a trading strategy using the notion of "pre-cognition". To ensure that we are not dealing with a situation of perfect information, we split our corpus into a training set of $1000$ days and a test set of $1241$ days. We construct our meta-distribution KDEs over the $1000$ day training set, and use the test corpus for analyzing the performance of our model over a truly

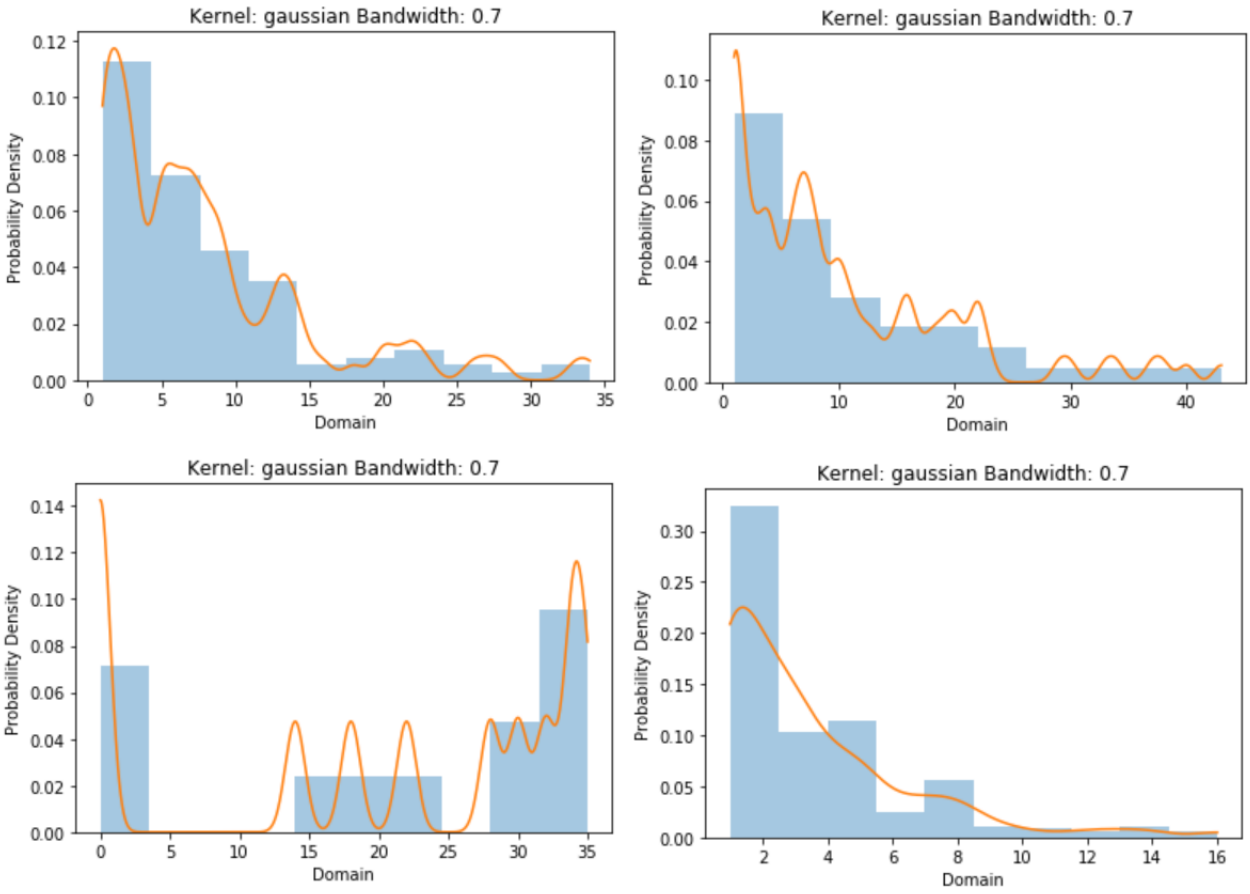


Figure 18: Gaussian KDE over empirical meta-distributions

non-observable future.

**Start** with a fixed window of size $W$. Thus, for the first iteration we would be evolving over days $1000 \rightarrow 1000 + W$, the second iteration over $1000 + W \rightarrow 1000 + 2W$ and so on.
**Estimate** the number of zero-crossings in the window from the $P_N$ meta-distribution KDE. We get a guess $n_g$ for the total number of pre-cognition events we are expecting in this window.
**Initialize** $t = 0, n_{cross} = 0$. We also start with an empty bank account with capitalization $C = 0$.
**While** $t \leq W$ and $n_{cross} \leq n_g$, we estimate the time for the occurence of a pre-cognition event from the $P_I$ inter-event meta-distribution KDE. We get a guess $e$ for the amount of time we need to wait from time $t$ to observe a pre-cognition. Concurrently, we also estimate the time to long movement ($l$) and time to short movement from pre-cognition ($s$). Our timeline now looks like:

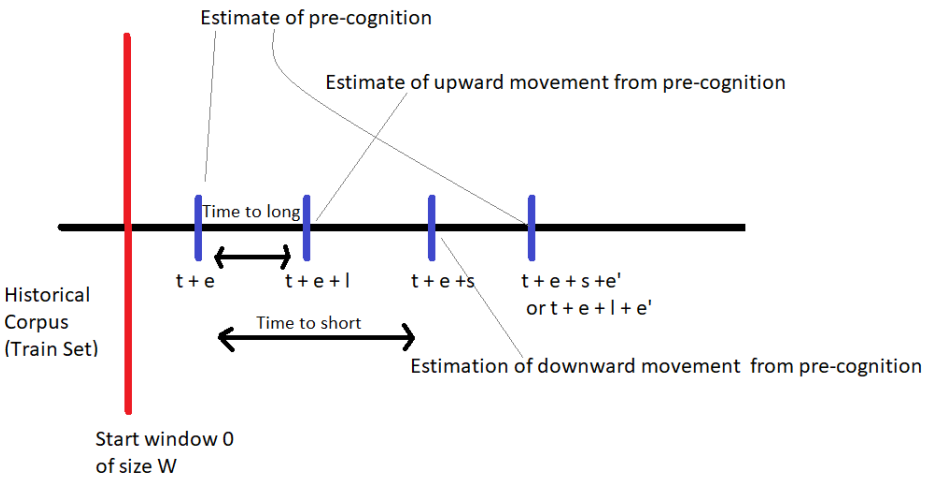


Figure 19: Evolution over test corpus

**Choose** the nearest event estimation i.e, if $l < s$, we will choose to invest instead of short, and vice versa.

**Long** position implies purchasing $1$ unit of the portfolio i.e., $1$ share for each symbol in the portfolio on day $t+e+l$ *provided* the slope of the linear fit between the time-stamps $t+e \to t+e+l$ is greater than $+1$. If $C = 0$, we assume that we can perform a risk-less loan from a financial entity, thus allowing $C < 0$ as well.

**Short** position implies selling all units on day $t+e$ to pick them back up on day $t+e+s$. Therefore, in this style of short selling we ensure that the number of unit stay constant for simplicity.

**Restart** the inter-event density generator and compute the new estimate for $e$ from the timestamp of either the long or short decision made previously. Continue the same till the end of the window.

When this strategy is implemented over a random choice of $30$ assets with a sampling window size of $W = 70$, we have the following results for the variation of $C$ and the number of units held: While the bank balance may make the situation look bleak, we must also remember that we are accruing stock units as well and our net (unrealized) profit will actually be calculated as:

$$\mathcal{E} = U \sum_i \bar{x}_{W_i} + L$$

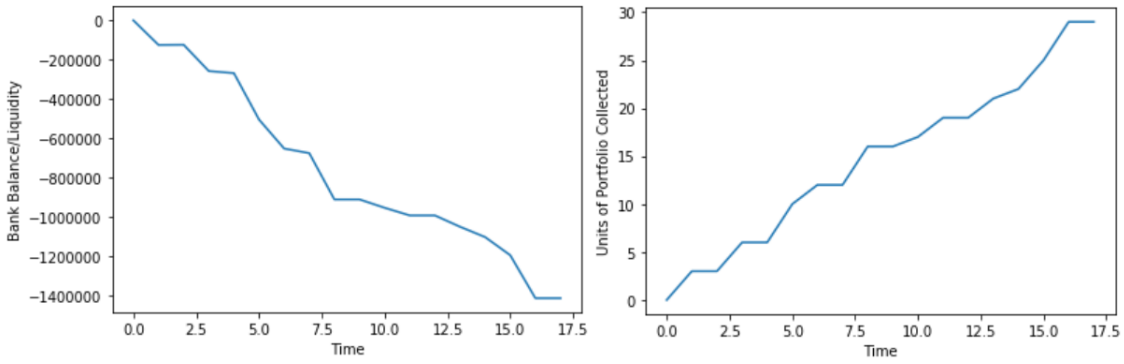


Figure 20: Liquidity and Unit Count Variation for $W = 70$

where $U$ is the number of units of the portfolio held at the end of the evolution and $L$ is the amount of liquid cash available/spent. In the specimen evolution described above, we compute approximately a 4X growth compared to the initial amount loaned.

# 6 Caveats and Future Scope

Our current formulation for preventing ruin involves analysis performed on small window sizes. In addition to this, we have considered several idealizations in order to have a generic view of the utility of this approach. The effect of taxation and commission plays a crucial role in modifying strategies that result in better returns (This aspect has not been included in our project for simplicity). Generally, financial organizations charge a finite rate of interest on any borrowed capital, which is yet to be modeled in this approach. Also, we have restricted the purchase and sale of assets to one per transaction instance, which largely differs in reality. In addition, we plan on extending the heuristics developed in this project to the domain of high-frequency trading.

Furthermore, good strategies for making financial decisions necessitate the design and usage of better financial instruments. In addition to TVAL3, we plan on investigating the consequence of using a framework based on Fourier projection onto convex sets[39, 40]. This approach will potentially lead to better de-noising since it is compliant with the Restricted Isometric Conditions and provides a coherent methodology to implement Dynamic Mode Decomposition with Compressive Sensing.

# 7 Acknowledgements

We are really grateful to Dr. Yasamin Mostofi for providing us with the opportunity to explore such exciting topics in optimization theory, and we would like to thank her for the valuable suggestions during the presentation of our project.

# 8 Bibliography

1 H. Markowitz, *Portfolio Selection*, The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77-91.

2 W.F. Sharpe, *Capital Asset Prices: A Theory Of Market Equilibrium Under Conditions Of Risk*, The Journal of Finance, Vol. 19, No. 3. (Sep., 1964), pp. 425-442.

3 J. Mann, J. N. Kutz, *Dynamic Mode Decomposition for Financial Trading Strategies*, arXiv.org, q-fin, arXiv:1508.04487, August 2015.

4 N. N. Taleb, A. Pilpel, *I problemi epistemologici del risk management (On The Unfortunate Problem of the Nonobservability of the Probability Distribution)*, Economia del rischio, Milan, 2004.

5 I. Levi, *The Enterprise of Knowledge*, MIT Press, 1980.

6 H. R. Varian, *Intermediate Microeconomics*, W. W. Norton & Company, New York, 1987.

7 G. Borter, *No milk, no bleach: Americans awake to Coronavirus panic buying*, Reuters, March 2020.

8 R. Cont, J. P. Bouchard, *Herd behavior and aggregate fluctuations in financial markets*, Macroeconomic Dynamics, Vol. 4, Issue 2, pp 170-196, 2000.

9 S. Bikhchandani, D. Hirshleifer, I. Welch, *A theory of fads, fashion, custom and cultural changes as informational cascades.* Journal of Political Economy 100, pp 992–1026, 1992.

10 M. R. Chernick, *What does a non positive definite covariance matrix tell me about my data?*, https://stats.stackexchange.com/q/30466, Cross Validated StackExchange, June 2012.

11 L. Bachelier, P. H. Cootner (Ed.), *The Random Character of Stock Market Prices*, MIT Press,

Cambridge, MA, 1964.

12 J. C. Hull, *Options, Futures and Other Derivatives*, Prentice-Hall, New Jersey, pp. 226-228, 2003.

13 K. Itô, *On Stochastic Differential Equations*, Memoirs of the American Mathematical Society, Vol. 4, pp. 1-51, 1951.

14 NSE BhavCopy Database, available for free online: https://www1.nseindia.com/products/content/equities/equities/archieve_eq.htm

15 B. Mandelbrot, R. Hudson, *The (Mis)behaviour Of Markets: A Fractal View of Risk, Ruin and Reward*, Profile Books, London, 2004.

16 N. N. Taleb, *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*, STEM Academic Press, 2020.

17 W. Feller, *Introduction to Probability Theory and Applications, Volume I*, Chapter 14, page 342, Wiley, 1973.

18 H. Arbabi, *Koopman Spectral Analysis and Study of Mixing in Incompressible Flows*, PhD thesis, University of California, Santa Barbara, 2017.

19 E. Candes, J. Romberg, T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, arXiv:math/0409186v1, June 2004.

20 D. L. Donoho, P. B. Stark, *Uncertainty Principles and Signal Recovery*, SIAM Journal on Applied Mathematics, Vol. 49, No. 3, June 1989, pp. 906-931.

21 F. Santosa, W. W. Symes, *Linear Inversion of Band-Limited Reflection Seismograms*, SIAM Journal of Scientific and Statistical Computing, Vol. 7, pp. 1307-1330., 1986.

22 A. M. Tillmann, M. E. Pfetsch, *The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing*, in IEEE Transactions on Information Theory, vol. 60, no. 2, pp. 1248-1259, Feb. 2014.

23 R. Penrose, *A generalized inverse for matrices.* Mathematical Proceedings of the Cambridge Philosophical Society, 51(3), pp 406-413, 1955.

24 A. N. Tikhonov, *On the stability of inverse problems*, Doklady Akademii Nauk SSSR. 39 (5),

pp. 195–198, 1943.

25 M. Lustig, F. Ong, J. Tamir, *Compressed Sensing Tutorial*, EE369C (Medical Image Reconstruction), Novemeber 2007.

26 M. S. Andersen, J. Dahl, L. Vandenberghe. *CVXOPT: A Python package for convex optimization*, Version 1.1.6, Available at cvxopt.org, 2013.

27 R. Tibshirani (1996), *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological). Wiley, Vol. 58 (1), pp. 267–88, 1996.

28 S. S. Chen, D. L. Donoho, M. A. Saunders, *Atomic Decomposition By Basis Pursuit*, SIAM Journal of Scientific Computing, Vol. 20, pp. 33-61, 1999.

29 E. Candes, J. Romberg, *L1-MAGIC: Recovery of Sparse Signals via Convex Programming*, Caltech, October 2005.

30 N. Cleju, *Python Compressed Sensing algorithms*, github.com/nikcleju/pyCSalgos.

31 S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

32 C. Li. *An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing.* Masters thesis, Rice University, 2010.

33 C. Li, W. Yin, H. Jiang, Y. Zhang, *An efficient augmented Lagrangian method with applications to total variation minimization*, Computational Optimization and Applications, 56(3), 507-530, 2013.

34 S. Depatla, L. Buckland, and Y. Mostofi, *X-Ray Vision with Only WiFi Power Measurements Using Rytov Wave Models*, IEEE Transactions on Vehicular Technology, Special Issue on Indoor Localization, Tracking, and Mapping, Volume 64, Issue 4, April 2015

35 Y. Mostofi, *Compressive Cooperative Sensing and Mapping in Mobile Networks*, IEEE Transactions on Mobile Computing, Volume 10, Issue 12, December 2011

36 C. Li, W. Yin, Y. Zhang, *TVAL3: TV minimization by Augmented Lagrangian and ALternating direction ALgorithms*, www.caam.rice.edu/ optimization/L1/TVAL3/, November 2010.

37 J. VanderPlas, *Kernel Density Estimation in Python*, Pythonic Perambulations, December 2013.

38 A. Dvoretzky, J. Kiefer, J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Annals of Mathematical Statistics, 27 (3), pp. 642–669, 1956.

39 M. Tofighi, K. Kose. A. E. Cetin, *Denoising Using Projection Onto Convex Sets (POCS) Based Framework*, arXiv:1309.00700v1, September 2013.

40 L. Bregman, *The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming*, Computational Mathematics and Mathematical Physics, Vol. 7, No. 3, pp. 200-217, Leningrad, 1967.