Statistical Methodology in Alternative Risk Premia

Yang Ming[1], Lim Chinghway[2]

Department of Statistics and Applied Probability,

National University of Singapore

2nd May, 2021

Abstract

Recent strands of literature have found that machine learning techniques have strong potential to outperform traditional regression-based approaches to empirical asset pricing. Since machine learning models are able to effectively extract signals from a vast set of correlated and noisy predictors, capture complex non-linear relations, and regularize itself sufficiently to prevent overfitting. Publications have received considerable attention for proposing recurrent neural networks (Nakagawa & et al, 2019), gradient-boosted regression trees and various other machine learning architectures (Gu, Bryan, & Dacheng, 2018) to use Alternative Risk Premia (ARP) strategies and generate a multi-factor signal to build quantitative investment strategies.

While we do acknowledge these salient points, we retain a healthy dose of skepticism. Portfolios analyzed in literature are not designed to optimize for transaction costs explicitly, and hence using the predictions from such models may be quite transaction cost inefficient. In our work, we describe in detail the trading process of a live trade: how transaction costs play a significant role in overall profitability, and how can one execute trades that limits their market impact in order to reduce transaction costs. We also extend the work of (Rayakar, 2020), constructing a portfolio from his predictions, and modelled the transaction costs explicitly based on market, stock, and trade characteristics. Testing the robustness of the strategy under realistic conditions after the drag represented by transaction costs, we show that a reasonable level of transaction cost erodes the risk-adjusted performance of machine learning driven strategies.

[1] ming@u.nus.edu
[2] lim.chinghway@u.nus.edu

Introduction to Alternative Risk Premia

This paper chiefly addresses whether using forecasts from machine learning models, the subsequent portfolio constructed from these predictions provide a reasonable level of performance, after taking into account transaction costs in realistic conditions.

Traditionally, a linear multi-factor model is one of the most important tools in the industry for stock returns forecasting. The attribute that explains stock returns is coined a factor, a multi-factor model that is simply one contains multiple factors, and the variable of interest is *excess return*, the equity's excess return above the prevailing risk-free return rate of a three-month US treasury bill, presumably a risk-free asset. The excess return, or *risk premium*, is the compensation to the investor for the risk of holding a risky equity.

Empirical asset pricing models seeks to price asset following a cross-section analysis of market-level factors and representative models widely used include:

- Capital Asset Pricing Model (Sharpe, 1964)
- Fama-French Factor Model (Fama & French, 1992)
- Carhart Four Factor Model (Carhart, 1997)

The market practice is to view *Alternative Risk Premia* (ARP) as an extension of the factor-investing approach, which is generally dedicated to long-only equity risk factors. ARP corresponds to long-short portfolios and generally refers to all systematic risk factors that have resulted in positive performance in the past (Hamdan, Fabien, Thierry, & Ban, 2016).

(Rayakar, 2020) extended both (Gu, Bryan, & Dacheng, 2018) and (Nakagawa & et al, 2019)'s work to US equities of the more recent past, modelling individual stocks' returns as time-series using an expanding window validation scheme with ARP into an RNN-OLS ensemble. The ensemble used RNN to estimate factor returns, a set of size, value, and momentum ARP factors, then fed into an OLS regression for the final output. The multi-factor model produced an exceptional out-of-sample $R^2$of 12.2% from 2017-2019.

Our contribution will be as follows:

- We provide a detailed description into real-world transaction costs and present a market impact model that estimates the transaction costs of a proposed trade, ex-ante, based on market, stock and trade characteristics.
- We propose a novel approach to portfolio optimization with quadratic transaction costs, to build an investment strategy minimizing transaction cost.
- We apply the proposed methodologies to an investable universe and test the robustness of the results, feeding the alphas into a portfolio optimizer and assess the performance of the model under realistic constraints.

1. Anatomy of a live trade

To simulate the more realistic scenarios under our proposed methodology, one needs to understand a typical institutional trading process, the role of liquidity and how transaction costs plays in the process. Figure 1 depicts a typical institutional investor's trading process:

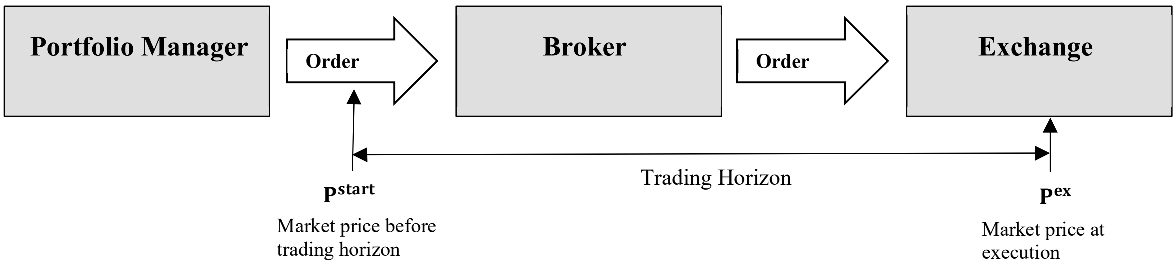


Figure 1 – Typical institutional investor trading process

The trade process starts with a portfolio manager signally the execution of a trade as a result of their investment process: which stocks to buy and sell, the number of the stock to be traded, and the trade horizon: the duration they want the trade to be executed by, before the trade gets executed by proprietary trading algorithms. Any trades will cause its participant to incur a transaction cost, which can be contributed by various factors: bid-ask spreads, market impact and commissions.

Let us illustrate with an individual investor's trades. While an individual, from his investment process, have have the intention to purchase 20,000 MSFT[3] shares, the market's current ASK[4] price of $151.08 contains only 3900 shares, but the individual can place a market order to fulfill the order as quickly as possible, at worse market price offerings, and Table 1 illustrates how the individual's order gets fulfilled:

Market Order for MSFT
Buy 2800 @ $39.24
Buy 1100 @ $39.24
…
Buy 500 @ $39.34

Table 1 – Market Order

This difference between when the individual investor signaled the trade and the price point is referred to as the bid-ask spread, and it erodes profitability. It stems from the lack of market depth when the order cannot be executed at the theoretical price point.

[3] MSFT: Stock ticker for Microsoft Corporation
[4] ASK Price is the price a seller states they will accept for a security
BID price is the price a buyer states they will buy for a security

For an institutional investor, the problem is compounded further because their trades can cause significant market impact. In the financial markets, any actions of a market participant has an effect on the market price, and using the MSFT example, the individual's order to buy MSFT shares must be fulfilled by a counterparty. However, the market price of MSFT did not get impacted significantly as many people trade MSFT, providing liquidity[5].

Due to the large size of their trades, which can take place at a significant percentage of an asset's trading volume, any trade actions an institutional investor perform leaves a trading footprint on the market. The market impact can be several percentage points between the ex-post Execution Price and the ex-ante Trade Price.

Therefore, market impact is a key consideration before an institutional investor commits to a trade, as it can erode a theoretically profitable strategy. Institutional investors spend considerable resources into proprietary trading algorithms, that seek to minimize market impact from raising the execution price too much causing significant slippage.

The largest part of the transaction cost of a large institutional trader faces is the market impact (Frazzini, Ronen, & Moskowitz, 2018), with commission and bid-ask spreads small in comparison to market impact at large trade sizes, since the latter do not effectively scale with size. In the following section, we go into detail on how we modelled the transaction cost of a large institutional investor's trades through estimating its market impact using an econometric causal model.

[5] Market liquidity refers to the ease with which an asset can be sold on the market without causing a significant change in its market price

2. Modelling transaction costs

We define *market impact* as the difference between the start price that exists before a trade begins, and the actual traded price at the time the trade has executed. We adapted the econometric causal model presented in (Frazzini, Ronen, & Moskowitz, 2018) to measure the ex-ante market impact of any arbitrary trade. The Market Impact Model takes the following form[6]:

$$MI = a + b \cdot x + c \cdot sign(x)\sqrt{|x|}$$

The model is similar to other transaction cost models in literature, such as (Scherer, 2010)

$$tc = Commision + \frac{Bid}{Ask} - Spread + \theta \sqrt{\frac{TradeVolume}{Dailyvolume}}$$

(Boyd & et al., 2017)[7]

$$tc = a|x| + b\sigma \frac{|x|^{3/2}}{V^{1/2}} + cx$$

The three transaction cost models share the same functional form, using some combination of total market volume traded of the asset, idiosyncratic volatility of the asset and market expectation of volatility in determining transaction cost. The rationale for using (Frazzini, Ronen, & Moskowitz, 2018) model was that it was calibrated from a trade database worth $1.7 trillion of live executed trades across a 19-year period from August 1998 to June 2016 across 21 equity markets of a large institutional trader, a significant size of data in across any scale. In comparison, our own data source only had data available from January 2005 to December 2019 in the US equities market. Additionally, the model itself matches economic theory, which we shall describe in greater detail in the following Table 2, as well as statistical measures of market impact such as Kyle's Lambda (Kyle, 1985) and other estimates (Hasbrouck, 2009).

---

[6] a = $\theta_2$time-trend + $\theta_3 log$ (1+ market cap) + $\theta_6 \sigma^{IV}$ + θ_7VIX,
a controls the general level of the trading and the market characteristics
b = θ_4 (fraction of daily average volume)
c = θ_5 (sqrt of fraction of daily average volume)
a and b controls the characteristics of the trade with respect to the stock
x: 100*m/dtv signed dollar volume of the trade (m) as a fraction of stock's average one-year dollar volume

[7] a: one-half bid-ask spread ex-ante
V: daily average volume traded
σ: idiosyncratic volatility
x: signed dollar of the trade

Theta	Name	Coef	Economic theory
θ_2	time-trend: June 1926 is 1, June 1927 is 2, ...	-0.01	Transaction costs have declined over time, largely driven by technological events such as moving to decimalization in traded places. (Frazzini, Ronen, & Moskowitz, 2018)
θ_3	market capitalization: log(1 + Market Capitalization)	-0.14	Larger stocks face lower price impact costs, which is also intuitive as the liquidity large stocks is typically much greater than for smaller stocks. A large trade size in illiquid stocks will result in much larger transaction costs, consistent with both literature (Kyle, 1985) and many empirical estimates (Hasbrouck, 2009)
θ_4	fraction of daily volume	-0.53	Larger trades have higher transaction costs, consistent with price impact and transaction cost models (Kyle, 1985). The relationship between price impact and trade size shows a concave shape, with costs growing significantly at very large sizes
θ_5	sqrt of fraction of daily volume	11.21	
θ_6	idiosyncratic volatility: the standard deviation of the residuals of one-year daily stock returns regressed against the benchmark index	0.31	Idiosyncratic Volatility measures the given stock's fluctuations as being larger or smaller than that of the benchmark
θ_7	VIX: monthly variance of the CRSP value weighted index representing market level of volatility	0.12	VIX represents the market's expectations for volatility over the coming 30 days, a measure for level of risk, fear, or stress in the market when making investment decisions. More volatile firms have higher transaction costs, consistent with models of market maker inventory risk, and more volatile market environments are also associated with larger price impact costs, consistent with market makers needing to be compensated more in more volatile markets.

Table 2 - Market Impact Model

Following a similar methodology in the original paper, we use Capital Asset Pricing Model (CAPM) to price expected returns in determining the ex-ante positions of our stocks, going long on the stocks with returns above the median, and short on the stocks below. The CAPM models excess return as a fitted multiple of the overall market risk premium, and we ended up an initial position of 311 long and 311 short stocks.

We used our available predictions for 2018 to determine our trades, and Figure 2 shows the calculated market impact of the stocks across differing trade sizes, % of one-year daily trading volume (% DTV).

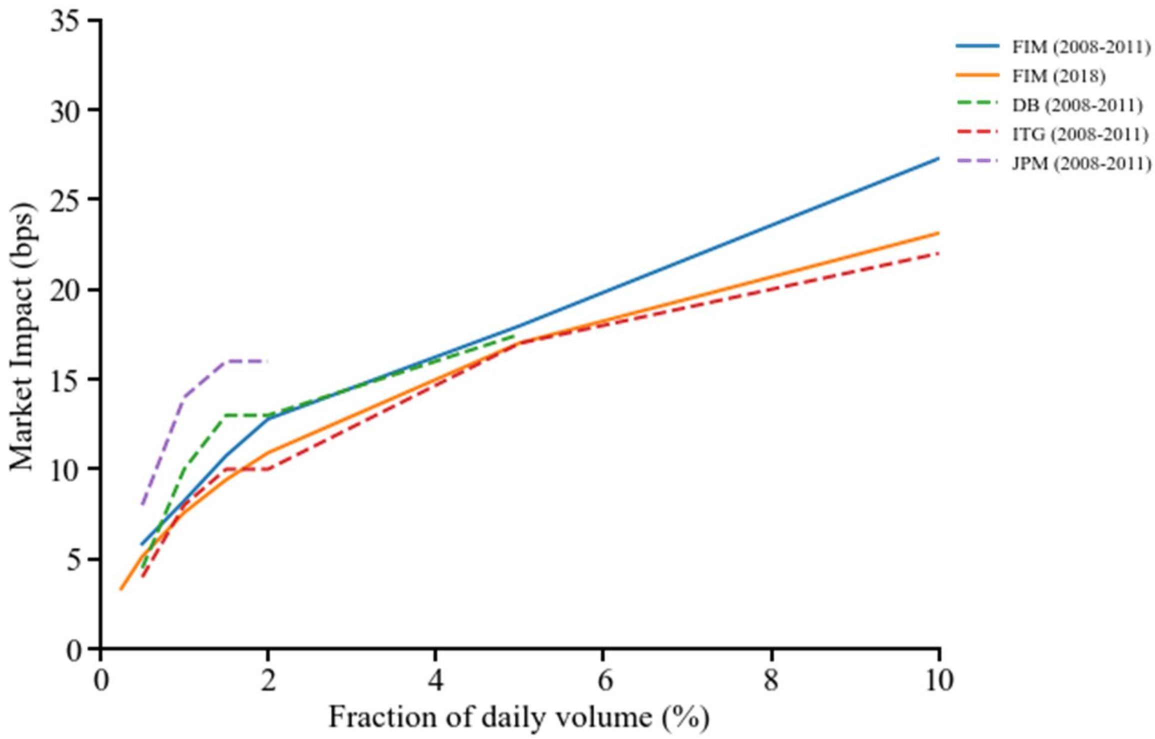


Figure 2 – Market Impact Model Estimates

Figure 2 shows average realized market impact of realized trades in dashed lines across different trade sizes (% DTV) as reported from three different brokers: ITG, Deutsche Bank (DB) and JP Morgan (JPM). Figure 1 also presents the estimated market impact from regression specification of (Frazzini, Ronen, & Moskowitz, 2018)'s Market Impact Model during the same time period in a blue solid line. Six different levels of trade sizes are reported (0.25-0.5% DTV, 0.5-1% DTV, 1-1.5% DTV, 1.5-2% DTV, 2-5% DTV, 5-10% DTV), with a common sample period from 2008-2011.

While our estimates are a fraction lower than the reported market impact, the decrease can be attributed to technological events like moving to decimalization in traded places, with no evidence of systematic reversal of quote changes of decimalization (Bessembinder, 2003), hence the decreasing linear time-trend in transaction costs of the more recent trades.

Additionally, live market impact data was also reported in two separate studies using data from ANcerno (Anand, Irvine, Puckett, & Venkataraman, 2012), (Marco, Francesco, Amir, & Carlo, 2017). ANCerno provides live trading data from 750 institutions covering 104 million trades.

	Market Impact (bps) at 2.4% DTV
ANcerno	24.5
AQR	18.2
Model Estimates	23.1

Table 3 – Reported realized market impact

The average trade size as a percentage in the two studies is 2.4% DTV, and hence the realized average market impact costs at 2.4% DTV from AQR's own live trade trades was also reported (Frazzini, Ronen, & Moskowitz, 2018). Table 2 shows the reported market impact figures, as well as estimates calculated from the Market Impact Model on our own 2018 trades.

From the outputs from the Market Impact Model, we can estimate the realized transaction cost of our proposed trade. Market Impact can be formally defined as below:

$$MI = Q_+(P^{ex} - P^{start}) + Q_-(-P^{ex} + P^{start})$$

Q_+ and Q_- representing the quantity of shares bought and sold as a result of the trade, respectively. P^{start} and P^{ex} represents the price at the start of trading and the execution price, with the start price the first recorded price at the time the strategy begins to trade, i.e. the current day’s opening price. The Market Impact unit is in basis points (bps), expressed in terms of relative rather than absolute dollars. A basis point is one hundredth of a percent, so the Execution Price can estimated by the following:

$$P^{ex} = \left(1 + \frac{MI}{100}\right) \cdot P^{start}$$

Our Market Impact Model delivers reliably close estimates very close to actual realized market impact reported from various brokers at common trade volumes, with market impact data reported from three different brokers (ITG, Deutsche Bank and JP Morgan), as well as a consulting firm (ANcerno) which covered 750 institutions executing 104 million trades over a common time period. This addresses the issue of the ongoing issue in academia with regards to price impact research stymied due to a lack of access to proprietary institutional investor’s live trading data. Using this model, we can examine how various trading strategies can survive transaction costs.

3. Portfolio optimization

There exists a myriad of approaches to portfolio construction, both in literature and in practice, to achieve diversification for the investor in turns of risk and returns. The most ubiquitous of which being mean-variance analysis. In this section, we will describe an investor can decide on a mathematically optimal allocation in their portfolio using mean-variance optimization.

The seminal work of (Markowitz, 1952) established the Modern Portfolio Theory, formalizing a risk-reward framework, using mean-variance as a proxy in portfolio optimization. To date, the mean-variance analysis remains the most widely used optimization approach used in Finance (Kolm, Reha, & Frank, 2014). Specifically, Markowitz formulated the investor's financial problem: the investor constructs his portfolio seeking to maximize the expected return for a given level of risk:

$$max\, \mu\,(w)\; s.t.\, \sigma(w) \leq \sigma^*$$

or equivalently minimizes the variance of the portfolio given a target return level:

$$min\, \sigma\,(w)\; s.t.\, \mu(w) \geq \mu^*$$

Subjected to the following constraints:

$$\sum_{i=1}^{n} w_i = 1$$

$$-1 \leq w \leq 1$$

By considering a Quadratic Utility Function[8]:

$$\mathrm{U}(w) = w^\top \mu - \frac{\phi}{2} w^\top \Sigma w$$

this turns the problem into a Quadratic Programming (QP) Problem:

$$min\, w^\top\, \mu - \frac{\phi}{2} w^\top \Sigma w$$

Because QP problems are straightforward to solve with great efficiency, they can be implemented from an industrial point of view to manage a universe with hundreds of assets. At the same time, mean-variance suffers from problems such as its sensitivity to estimation errors,

---

[8] w: $(w_1, \dots, w_n)$ is the vector of weights in a portfolio of n assets
μ is a vector of expected returns of assets in the portfolios
Σ is the covariance matrix of returns on assets in the portfolio
ϕ is the risk adversion, where $\gamma = \phi^{-1}$ is the given risk level

and producing intractable allocations making no financial sense[9], and very rarely applied in practice in its most traditional forms.

We shall elaborate more on these problems in greater detail in the next section, followed by how we tackle practical challenges to the mean-variance optimization and our proposed solution.

3.1 Covariance matrix shrinkage

In addition to expected returns, Mean-Variance Optimization requires a risk model Σ to quantify asset risk, and in theory, the in-sample covariance matrix given by:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^{\top}$$

is a suitable choice given that it's an unbiased estimator of the covariance matrix:

$$\sum = E[(X - \mu_x)(X - \mu_x)^{\top}]$$

However, using the sample covariance matrix poses certain problems when the number of stocks *N* is an order larger than the number of historical returns per stock *T*, the sample covariance matrix is always singular. This poses a tricky problem, because the size of a tradeable universe of an institutional trader can be over a thousand, but rarely more than 10 years of monthly data. When that happens, the sample covariance matrix gets mis-specified with a lot of estimation errors, producing extreme coefficients, which then gets amplified through the mean-variance optimization, producing erroneously high values on weights based on these coefficients with high estimation errors.

On the other hand, highly structured estimators such as Sharpe's one-factor model (Sharpe, A simplified model for portfolio analysis, 1963) have little estimation errors, but may be highly biased. Recent research indicates there are much more robust statistical estimators of the covariance matrix, and one possible improvement proposed by (Ledoit & Wolf, 2003) is to shrink the sample covariance matrix by moving extreme values towards the center. The underlying idea is inspired from empirical Bayesian statistics, to compromise between the highly structured estimator *F* and an unstructured estimator *S*, formalized by:

$$\delta F + (1 - \delta)S, 0 \leq \delta \leq 1$$

---

[9] See Table 4

(Ledoit & Wolf, 2003) set all pairwise correlations to the mean of all sample correlations, this means that the shrinkage target matrix F is given by: $f_{ii} = s_{ii}, f_{ij} = \bar{r}\sqrt{s_{ii}s_{jj}}$, where s_{ij}^2 is the sample covariance between two stocks, of the sample covariance matrix S

Let y_{it} be the returns of the i^{th} equity at time t, and T be the number of price datapoints. $\hat{\pi}$ estimates the sum of asymptotic variances of S, scaled by $\sqrt{T}$

$$\hat{\pi} = \sum_{i}^{N}\sum_{j}^{N}\widehat{\pi_{ij}} \text{ with} \approx \widehat{\pi_{ij}} = \frac{1}{T}\sum_{t=1}^{T}\{(y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - s_{ij}\}^2$$

$\hat{\rho}$ estimates the sum of asymptotic covariance of entries of F with entries of S, scaled by $\sqrt{T}$

$$\hat{\rho} = \sum_{i=1}^{N}\widehat{\pi_{ij}} + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\widehat{\theta_{ii,ij}} + \sqrt{\frac{s_{ii}}{s_{jj}}}\widehat{\theta_{jj,ij}}\right)$$

Where $\widehat{\theta_{ii,ij}} = \frac{1}{T}\sum_{t=1}^{T}((y_{it} - \bar{y}_i)^2 - s_{ii})((y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - s_{ij})$

$\hat{\gamma}$ estimates misspecification of the F

$$\hat{\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(f_{ij} - s_{ij}\right)^2$$

Then the optimal shrinkage coefficient δ is then estimated by the estimator $\hat{\kappa} = \frac{\hat{\pi}-\hat{\rho}}{\hat{\gamma}}$

$$\hat{\delta} = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{T}, 1\right\}\right\}$$

The truncation is necessary because the values can be greater than 1 or less than 0. (Ledoit & Wolf, 2003) provided empirical evaluation of their Ledoit-Wolf Shrinkage method, as compared with other risk models, with promising results. To this end, we also provide our own empirical findings comparing the results from using Sample Covariance Matrix and Ledoit-Wolf Shrinkage as Risk Models.

Our methodology: we solved the mean-variance optimization, using the ensemble's 2017-Q2[10] predicted returns as expected returns to rebalance at 2017-Q1, using a universe of 622 equities.

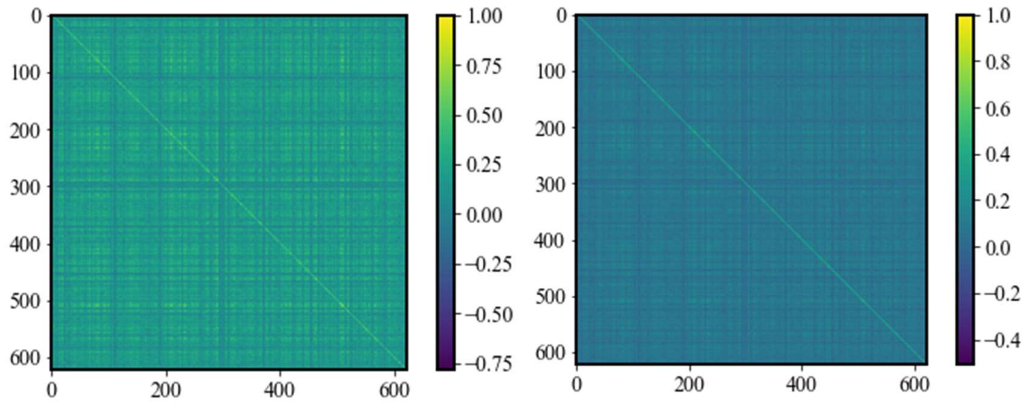


Figure 3 - Sample Covariance Matrix *Ledoit-Wolf Shrinkage*

Notice in Figure 3, the sample covariance matrix has brighter spots outside the diagonal, whereas Ledoit-Wolf Shrinkage reduced the extreme values in the covariance matrix, pulling it towards the center. The implication of this is that the mean-variance optimization using the same set of expected returns produced the following allocation:

	Sample Covariance	Ledoit-Wolf Shrinkage
A	-0.999999	-0.00363
AAN	-0.999999	0.00927
	…	…
ZBH	-0.999999	-0.00523
ZBRA	0.999997	0.00796
ZEUS	-0.999999	-0.01642

Table 4 – Weight allocations from mean-variance optimization

The resulting allocation makes mathematical sense, because the weights all sum to 1, but it proposes financially intractable positions. The mean-variance optimization allocation with the sample covariance matrix will blow any budget constraints an investor may have. Ledoit-Wolf Shrinkage, being designed to improve estimation of the covariance matrix of stock returns specifically with portfolio selection in mind, produces far more reasonable positions.

[10] Barring holidays and weekends, Q1-Q4 are 03-31, 06-30, 09-30, 12-31

4. Optimized strategies for transaction cost

The general approach for introducing liquidity management in mean-variance optimization is to assume fixed bid-ask spreads, which can then be solved using an augmented QP problem. Our findings suggest that transaction costs may be a linear function of the trade size matches existing literatures on trade friction. This implies that a model with quadratic transaction costs may be more appropriate.

In this section, we investigate a liquidity management approach to mean-variance optimization when transaction costs are no longer linear. We present the modern approaches (Pierre, 2019) in solving such quadratically constrained quadratic programs (QCQP), as well as an alternative approach to optimize portfolio for transaction costs, sidestepping solving the QCQP altogether. Recall that the mean-variance optimization takes the following form:

$$min\, w^T \mu - \frac{\phi}{2} w^T \Sigma w$$

To introduce transaction cost to mean-variance optimization, we define $C(w|\vec{w})$ as the transaction cost of rebalancing current portfolio $\vec{w}$ toward portfolio w. Effectively, this measures the difference between the results of a theoretical, frictionless portfolio and the results of a practical portfolio when subjected to actual trade prices.

$$min\, w^T \mu - \frac{\phi}{2} w^T \Sigma w - C(w|\vec{w})$$

Subjected to the following constraints:

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} w_i + C(w|\vec{w}) = 1$$

$$-1 \leq w \leq 1$$

When the selling and buying functions of the transaction cost are not equal, the budget constraint is no longer linear[11], and we obtain a QCQP problem, and hence the optimization problem is not convex. A numerical solution can be obtained by considering an interior-point algorithm specifying the gradient of the objective function, the gradient of the equality constraint and the Hessian of the Lagrangian. (Pierre, 2019)

[11] Detailed workings is in the appendix, adapted from Pierre et al (2019)

4.1 Optimized market impact portfolio

Portfolios analyzed in the literature may exhibit transaction cost inefficiencies as they are not designed to optimize for them in any way. Without taking into account after transaction cost returns, traditional mean-variance optimization may allocations that are financially intractable. Hence, we present the first strategy that optimizes for minimal transaction cost, as well as empirical examples using our approach.

We first take a look at a specific example from our backtest of an unconstrainted mean-variance during rebalancing at Q1-2017. We set a maximum risk level of 5% and maximize expected returns using alphas from our machine learning model, with no additional constraints and a starting capital of USD 100 Million.

Ticker	Price	Previous Weight	Previous Share	New Weights	New Shares	Shares Difference	1-year DAV
CAMP	16.6	-0.00839	-57863	0.0995	718807	776670	439010

Table 5 – Allocation of CAMP using unconstrainted mean-variance optimization

The resulting allocation at Q1-2017 is financial infeasible, as it asks for allocation near 200% of CAMP's one-year Daily Average Volume. The estimated market impact will be 44.27bps, with the execution price of the trade estimated to be 23.95, incurring significant transaction costs.

We can solve the QCQP to minimize transaction cost through the methodology presented in Section 4 directly, however our empirical find it incurred significant runtime to solve the QCQP for our investable universe (n=622), and an order higher than solving an unconstrainted QP. Since we are interested in implementable portfolios, we present a method of portfolio optimization that sidesteps the issue and produces an optimization in reasonable runtime.

The optimization problem of the Market Impact Portfolio is as follows:

$$\min_{w} \text{Transaction cost (w)}$$

Subjected to:

1. Tracking Error Constraint: $\sqrt{(w - w_b)'\Sigma(w - w_b)} \leq 1\%$
2. Fraction of Daily Average Volume $\leq$ 10%

The first constraint is to allow us to produce a portfolio optimized for transaction costs, but at the same time retaining the style of the original model. We want to preserve the multi-factor signal from our alpha model, and hence the first constraint minimizes the style drift we allow our Market Impact Portfolio to have, by allowing a 1% tracking error on the unconstrainted portfolio. w_b represents the weight of the unconstrainted portfolio, which we calculate ex-ante, and then we seek to find a vector of weights w of the chosen portfolio weights that minimizes transaction cost. The transaction cost function is based on our market impact model in Table 2.

The second constraint is commonly used as a heuristic to limit market impact and ensuring that the portfolio does not overly dominate a market for very small stocks, and it is also consistent with reports of an institutional investor's own trades (Frazzini, Ronen, & Moskowitz, 2018), where most of their reported trades do not exceed this threshold, either as a direct result of the investor's asset allocation strategies, or as an indirect result of their optimization strategies.

Following the same methodology as the unconstrainted portfolio, our market impact optimized portfolio produced the following allocation:

Ticker	Price	Previous Weight	Previous Share	New Weights	New Shares	Shares Difference	1-year DAV
CAMP	16.6	-0.015	-15329	0.015	16760	32089	439010

Table 6 – Allocation of CAMP using market impact optimized strategy

The Market Impact Portfolio produces a comparatively more reasonable allocation, with an estimated execution price of the trade being 18.26 after market impact. The subsequent transaction cost would much less significant than the unconstrainted portfolio in Table 5.

4.2 Optimized mean-variance portfolio

Literature (Borghi & Giuliano, 2020) also proposes an additional set of these constraints that seek to minimize transaction cost through heuristic arguments, and commonly used in quantitative strategies in practice.

- Net sector exposure is limited within -10% and 10% of net portfolio value using Level 1 GICS sectors
- Individual stock weights are capped at 1.5%, and a threshold of 0.5% is imposed to avoid very small positions.
- Limit the amount that can be traded at rebalancing in an equity to 10% of its 1-year Daily Average Volume

To test the robustness of our proposed methodologies, we apply the strategies to an investable universe and validate its performance using historical data. The baseline strategy is a Long/Short, fully collateralized[12] portfolio, and we feed the predictions from (Rayakar, 2020)'s multi-factor model in as an input to maximize the expected returns using Mean-Variance Optimization, subjected to a maximum risk level of 5%.

We assumed a starting cost of USD 100 Million in Q3-2016, and allocated our initial portfolios[13] using expected returns from CAPM. The start of the backtesting period is Q4-2016, where the portfolio was rebalanced using Q1-2017 predictions from the multi-factor model. To

12 Fully collateralized portfolio means that i.e. if the amount of collateral is 100, we open long and short positions worth a total of 100 on each side

13 Our methodologies only seek to augment an existing portfolio, not initialize a portfolio from scratch

gauge the impact of transaction costs on performance, we estimated the market impact of the trades with our market impact model, then added an additional one-way cost of 10bps[14], in line with transaction costs of an institutional investor. (Frazzini, Ronen, & Moskowitz, 2018)

5. Backtest results

Figure 4 reports the cumulative returns of the strategy using different portfolio rebalancing strategies without factoring in transaction costs. To better assess the stock picking skills of our multi-factor model, we compared them with the returns of using a long-short CAPM, fitted on the overall market premium. Unsurprisingly, the time pattern of returns is very similar across strategies, as the final building blocks for all strategies is a Linear Model. The biggest gap in returns between the CAPM and our multi-factor model is during the 2018 period, where the multi-factor model was able to provide returns from their ARP when the stock market had an extreme volatile year.

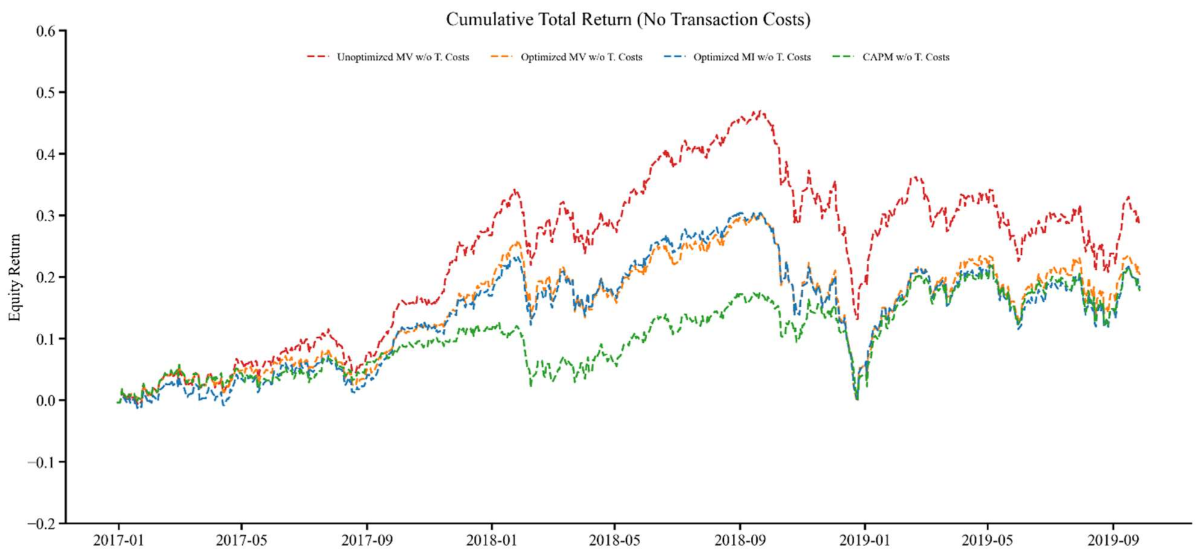


Figure 4 – Cumulative total returns before transaction costs

The dashed lines[15] are the cumulative returns from strategies before accounting for transaction costs. However, we have built up the case that transaction cost is significant, and that our market impact model gives us a reasonable estimate of the actual transaction costs. So, using our market impact model to estimate the transaction costs in our backtests, you will see in Figure 5 that there is a significant decrease in returns.

Profitability from machine learning-driven strategies is quickly eroded after transaction costs. Evident by the steep drops in return at every quarter, large amounts of trading will incur a lot of transaction cost whenever you turnover the portfolio. One possible explanation as to how

14 To liquidate a long position and substitute with short positions would cost 20bps of the portfolio value

15 HQ Images for figures are attached at Appendix III

the multi-factor model produced such strong out-performance from its exposure to fast-moving signals from momentum factors.

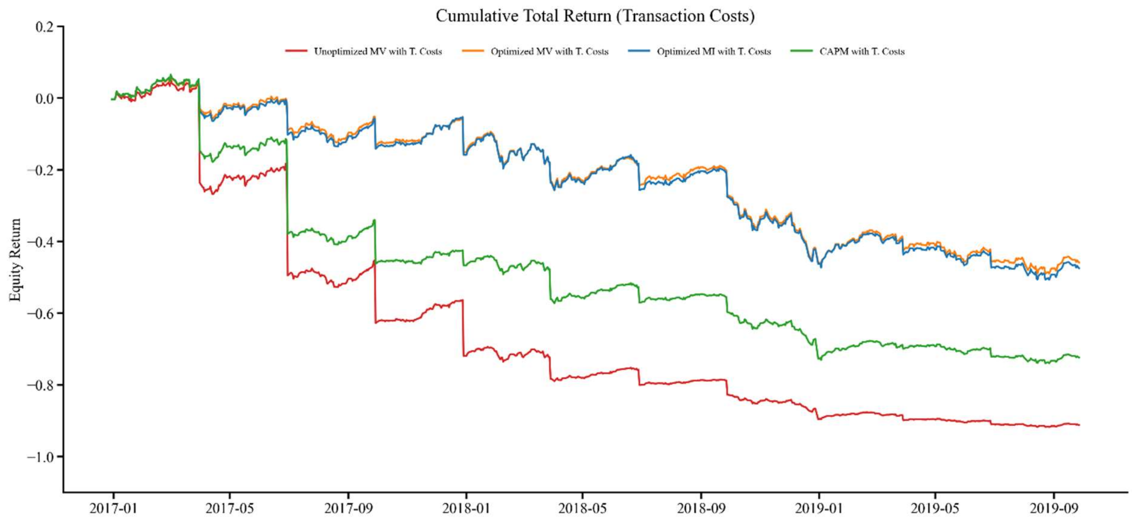


Figure 5 – Cumulative total returns after transaction costs

We can see significantly less steep drops accredited to our optimized strategies bringing down transaction costs, but ultimately the strategies remain unprofitable after applying transaction costs to bring about a realistic testing environment. The poor returns can be attributed to the fact that once you subtract the transaction cost, you have even less capital at the next available rebalance due to the fully collateralized portfolio. So, while a machine learning model had exceptional out-of-sample predictive ability, they do not necessarily translate into stable positions in a portfolio.

This means that without the constraints from the proposed optimized strategies, you incur a lot of transaction costs in live trading. However, the gap between transaction costs can be closed by our proposed optimized strategies, where an investor can minimize the transaction cost of any portfolio. Figure 6 below decomposed the reduction in transaction cost for both strategies.

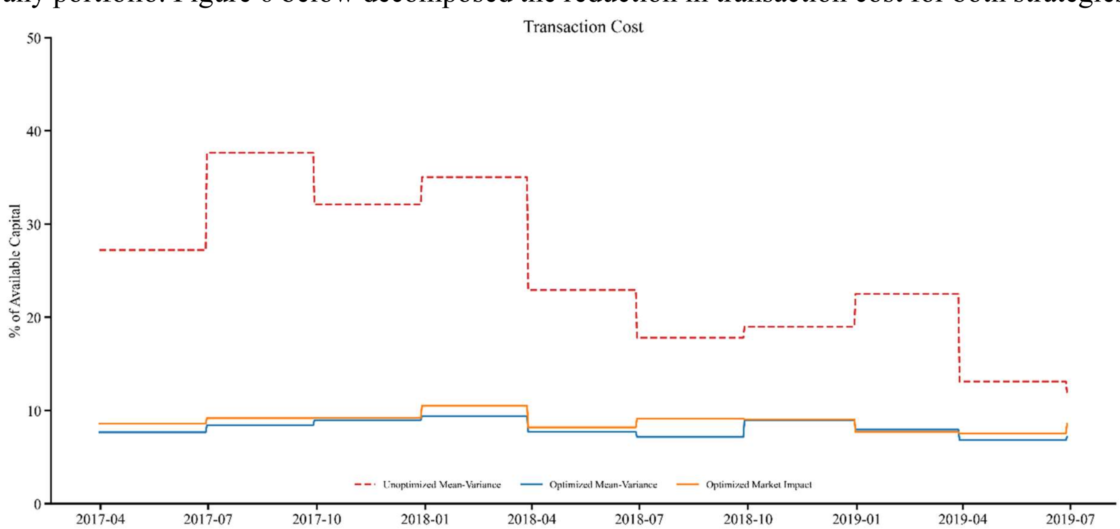


Figure 6 – Transaction costs before and after optimized strategies

In both cases, the constraints help lower the transaction cost of a trade significantly, and they reduce the amount of portfolio turnover in Table 7.

Furthermore, we also find that even though not explicitly targeting to minimizing transaction cost in the optimization, the second optimized strategy was able to achieve better levels of transaction cost-adjusted performance than the first optimized strategy. We can see the appeal on the set of constraints from the second optimized strategy and why it is often used in practice, as the set of constraints serves as a heuristic into minimizing transaction costs.

5.1 Metrics

The following Table 7 in the next page contains the calculated evaluation metrics from our backtests, with an additional column showing the reported results from (Borghi & Giuliano, 2020), who also ran backtests using predictions from a proprietary Quant Alpha model. The Alpha model is a challenging benchmark, being a large institutional investor's global stock picking model, in use since 2008 with solid out-of-sample performance, but it provides a much more realistic test of our multi-factor model's stock picking abilities.

Our methodology is largely similar, with identical starting capital, maximum risk level, constraints, and benchmark index (MSCI-US). The differences are that (Borghi & Giuliano, 2020) have a much longer backtesting period from January 2011 to December 2018, with their portfolios rebalanced monthly, whereas in ours we rebalance quarterly.

Our backtesting period is comparatively shorter, from January 2017 to September 2019, as our multi-factor model requires us to train on daily frequency data, which we only have spanning 2005-2019. The differences are because the data our multi-factor model was trained on had predictors that were only updated quarterly and/or yearly, we only have end-quarter predictions from Q1-2017 to Q3-2019.

Another important distinction is that (Borghi & Giuliano, 2020) only applied a one-way transaction cost of 10bps, while in our case we applied the 10bps transaction cost on top of the estimates from our market impact model, the implication of this is that the transaction cost will impact our multi-factor model significantly more, our net returns are 30% lower. It was a conscious decision to apply 10bps transaction cost prior to estimating transaction costs with our market impact model estimates, so that our model's Gross Returns comparable to the Gross Returns of the Alpha Model.

	Mean-Variance	Market Impact	Alpha Model
Gross Return	14.3%	13.9%	6.2%
Net Return	-18.6%	-18.9%	
Gross Information Ratio	0.797	0.802	0.870
Net Information Ratio	0.794	0.793	

	Mean-Variance	Market Impact	Alpha Model
Avg. Turnover[16]	2.47	2.47	
Avg. Turnover – Optimized	-1.35	-1.37	1.39
Max Drawdown	42.2%	42.9%	11.9%
No. of Observations	11	11	95

Table 7 – Backtest results

This is also suggested by our very respectable Information Ratio[17], where the Alpha model reported an Information Ratio of 0.870, and our multi-factor model in the different optimized strategies having Information Ratio of 0.802 and 0.797 respectively. We can also see how our optimized strategies work in reducing portfolio turnover, given that our unoptimized strategies had an average turnover of 247% prior to applying them, which gave the optimized strategies an average turnover of 136% similar to Alpha Model's 139%

However, our multi-factor model suffers from much more pronounced 42.2% and 42.9% Max Drawdowns[18], compared to 11.9% of the Alpha model. Downside risk is a key concern for most investors who wishes to assess the relative riskiness of one strategy over another. So, while the resultant portfolio using our multi-factor model do have attractive net returns and portfolio turnover rates, it can be perceived as less stable given its Max Drawdown metrics.

While we can see that under similar methodologies, machine learning driven investment strategies do provide attractive returns, the impact of transaction cost on returns is critical. All the strategies were clearly all profitable before applying transaction costs, but the very good backtesting results, and hence profitability, under realistic conditions, gets eroded away by transaction costs.

16 Turnover is calculated as a total of long and short sides and expressed as a percentage of the collateral value. Hence, the range of turnover that can be generated when rebalancing the portfolio is [-400%, 400%]

17 The Information Ratio measures the outperformance of a model relative to a benchmark index, in this case the MSCI (US)

18 Max Drawdown measures the maximum observed loss from a peak to a trough of a portfolio and a measure of downside risk over a time period

6. Conclusion and further work

Our empirical findings can be summarized as follows. First, we confirm the strong potential of machine learning techniques in building a multi-factor model, having significant predictive power in their predictions, when translated into positions in a portfolio, generate attractive returns, even when compared against a proprietary model with a successful track-record. Second, we found transaction cost to be significant enough to erode the profitability of a strategy. In a realistic optimization set-up, we proposed ways to minimize the transaction costs of any arbitrary setup in reasonable computational runtimes. To conclude, under our proposed optimized strategies, the savings in transaction costs is good, but overall, not great because we remain ultimately unprofitable.

Further work can be developed in several directions. First, we recognize that while our machine learning techniques results in a multi-factor model with strong predictive ability, one can possibly stabilize the resultant portfolio by training initially on a much more diverse set of ARP factors, before we feed the resulting alphas into our portfolio optimization set-up. Second, there exists a myriad of risk models, such as Conditional Value at Risk (Rockafellar & Stanislav, 2000), which can be tried in our portfolio optimization for more effective risk-management of the downside risk. This can be developed further to provide an attractive level of risk-adjusted returns for our portfolio. To that end, we provide our documented code we used for portfolio optimization and our backtesting results for additional work.

References

Anand, A., Irvine, P., Puckett, A., & Venkataraman, K. (2012). Performance of instutitional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies 25.2*, 557-598.

Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis*, 747-777.

Borghi, R., & Giuliano, D. (2020). The Artificial Intelligence Approach to Picking Stocks. In *Machine Learning for Asset Management: New Developments and Financial Applications* (pp. 115-166). Wiley.

Boyd, S., & et al. (2017). Multi-period trading via convex optimization. *arXiv preprint*, arXiv:1705.00109.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52.1*, 57-82.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance 47.2*, 427-465.

Frazzini, A., Ronen, I., & Moskowitz, T. J. (2018). Trading Costs. *SSRN*, SSRN 3229719.

Gu, S., Bryan, K., & Dacheng, X. (2018). Empirical asset pricing via machine learning. *National bureau of economic research*(No. w25398).

Hamdan, R., Fabien, P., Thierry, R., & Ban, Z. (2016). A primer on alternative risk premia. *SSRN 2766850*.

Hasbrouck, J. (2009). Trading costs and returns for US equities: Estimating effective costs from daily data. *The Journal of Finance*, 64.3: 1445-1477.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag. Retrieved 2 24, 2021, from http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Hu, G., Koren, J. M., Yi, A. W., & Jing, X. (2018). Institutional trading and Abel Noser data. *Journal of Corporate Finance 52*, 143-167.

Jurczenko, E. (2015). *Risk-based and factor investing.* Elsevier.

Kolm, P. N., Reha, T., & Frank, F. J. (2014). 60 Years of portfolio optimization: Practical challenges and current trends. *European Journal of Operation Research*, 356-371.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315-1335.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance 10.5*, 603-621.

Lobo, M., Maryam, F., & Stephen, B. (2007). Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 341-365.

Marco, D., Francesco, F., Amir, K., & Carlo, S. (2017). The Relevance of Broker Networks for Information Diffusion in the Stock Market. *Working Paper*.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 77-91.

Nakagawa, K., & et al. (2019). Deep recurrent factor model: interpretable non-linear and time-varying multi-factor model. *arXiv:1901.11493*.

Pierre, C. (2019). A Note on Portfolio Optimization with Quadratic Transaction Costs. *SSRN 3683466*.

Rayakar, A. (2020). *Machine Learning for Non-Linear Factor Modelling.* NUS.

Rockafellar, T., & Stanislav, U. (2000). Optimization of conditional value-at-risk. *Journal of risk 2*, 21-42.

Scherer, B. (2010). *Portfolio construction and risk budgeting.* Risk Books.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science 9.2*, 277-293.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19.3*, 425-442.

Appendix

Appendix I – Alternative Risk Premia Strategies for Trading Equities

1. Carry

Well known in currency markets, investors borrow currency with a low interest rate to buy a currency with a high interest rate, with the difference in rates the investor returns.

Koijen et al (2013) abstracted the principles behind this to identify Carry strategies in equities: Go long on stocks with net benefit (positive carry), short on stocks with net cost (negative carry)

Costs: Risk-free rate (to finance its purchase)
Benefits: Trailing dividend yield (to predict future dividend yield).

2. Size

Assets with smaller market cap generate higher returns, so gain additional exposure through the illiquidity of the asset: (Smaller size, lower trading volume, higher bid-ask spread, slippage, etc.)

3. Momentum

Momentum strategies essentially go long on past winners, and short on past losers. The differences between them are those of implementation, particular in terms of how momentum of a given stock is defined. Jegadeesh and Titman (1993) ranked stocks based on returns over the past J months, then go long on past winners (top deciles), short on past lowers (bottom deciles)

4. Reversal

Construct autocorrelation functions to establish if a mean-reverting pattern exists, buy the stock, because on average a negative return is followed by a positive return. Otherwise known as Mean-Reverting Strategies, capitalize based on investor's short-term overreaction to information and fads

5. Value

Value strategies tilt asset allocation towards under-valuated stocks over over-valuated ones, in expectation greater rates of return for the former as the market acknowledges its value

6. Profitability

Profitability, as measured by a ratio of the firm's gross profits to its asset, holds roughly the same power as book-to-market factor predicting the cross-section of average returns. The idea is to profitable stocks provide an excellent hedge for value, as they are very much in common philosophically, but highly dissimilar in characteristics

7. Liquidity

Liquidity trading strategies assume the existence of an illiquidity premium in returns of illiquid stocks over otherwise equivalent liquid stocks

8. Volatility

Equity-specific volatility risk factor refers to building a portfolio with exposure to low volatility stocks, stemming from the success of minimum variance portfolios

Appendix II – Quadratic Transaction Costs

If we consider the unit transaction cost as a linear function of the rebalancing weight, it follows:

$$C(w|\vec{w}) = \sum_{i=1}^{n} \Delta w_i^- (c_i^- + \delta_i^- \Delta w_i^-) + \sum_{i=1}^{n} \Delta w_i^+ (c_i^+ + \delta_i^+ \Delta w_i^+)$$

Where δ_i^- and δ_i^+represents linear function of transaction cost of sale and purchase of Asset I $\Delta w_i^- = max(\vec{w_i} - w_i, 0)$ and $\Delta w_i^+ = max\left(w - \vec{w_i}_i, 0\right)$ represents the sale and purchase of Asset I.

and by definition, we have $\Delta w_i^- \cdot \Delta w_i^+ = 0$ and $w_i = \vec{w_i} + \Delta w_i^+ - \Delta w_i^-$

$$= \Delta {w^-}^\top c^- + \Delta {w^-}^\top \Delta^- \Delta w^- + \Delta {w^+}^\top c^+ + \Delta {w^+}^\top \Delta^+ \Delta w^+$$

Where $\Delta^- = diag(\delta_1^-, \dots, \delta_n^-)$ and $\Delta^+ = diag(\delta_1^+, \dots, \delta_n^+)$

It follows that the objective function remains quadratic:

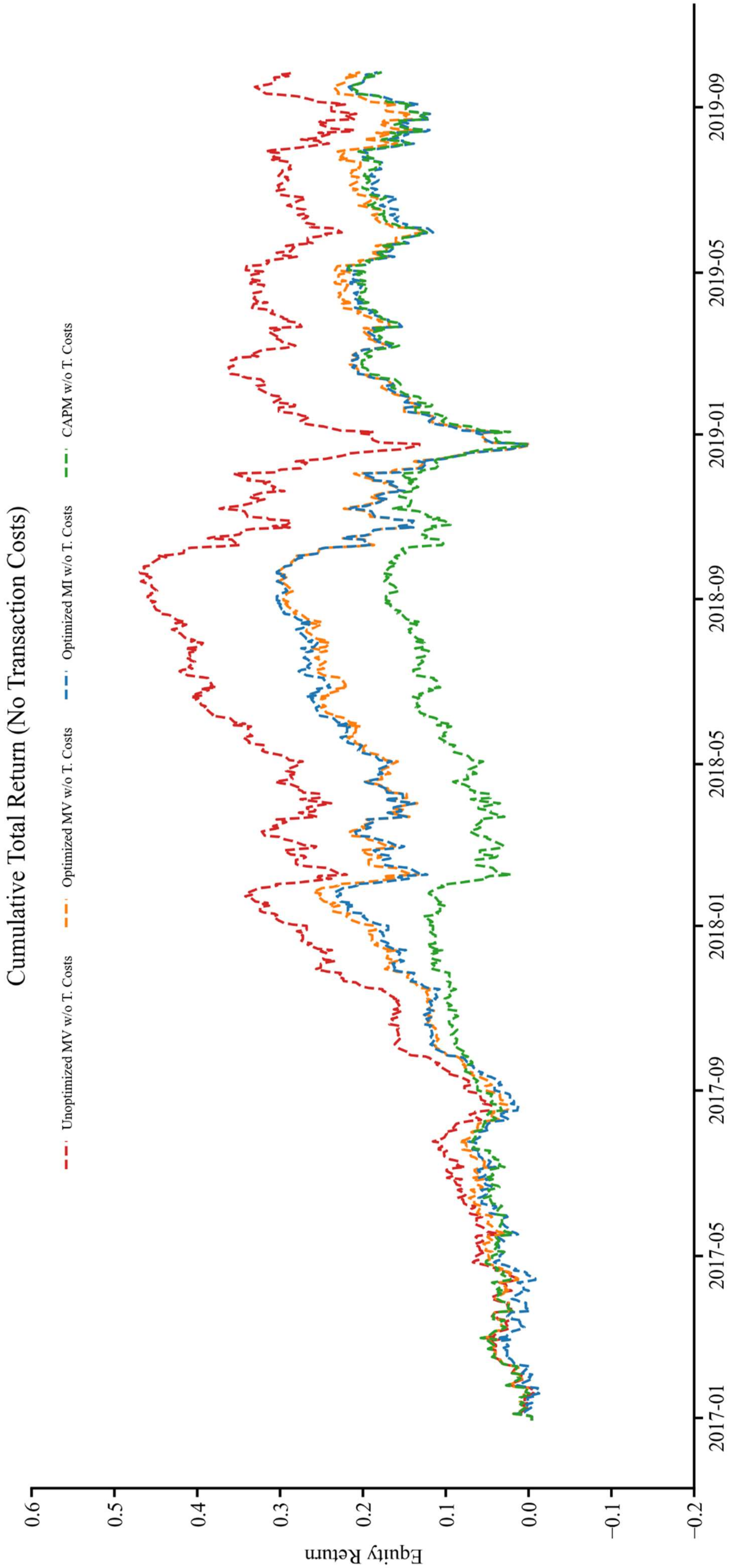
$$f(w, \Delta w^-, \Delta w^+) = \frac{\phi}{2} w^\top \Sigma w - \phi w^\top \mu + C(w|\vec{w})$$

$$= 1/2 \left(w^\top \Sigma w + \Delta {w^-}^\top (2\phi\Delta^-) \Delta w^- + \Delta {w^+}^\top (2\phi\Delta\text{^} + \Delta \text{w^\{}\top\text{\})} - \phi(\mathrm{w}^\top \mu - \Delta {\mathrm{w}^-}^\top \mathrm{c}^- - \Delta {\mathrm{w}^+}^\top \mathrm{c}^\top\right)$$

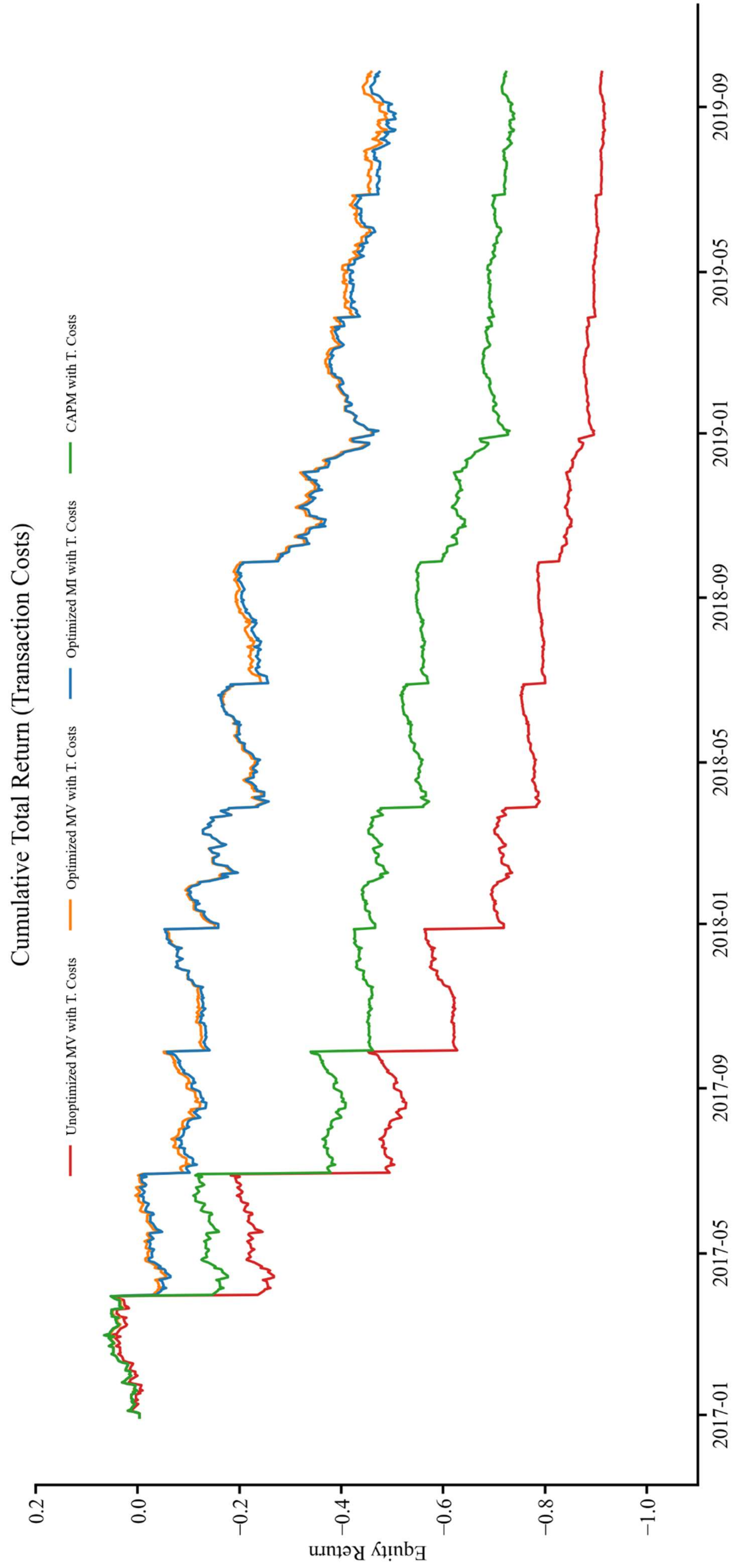
But the budget constraint is no longer linear, composed of a linear term and a quadratic term:

$$1_n^\top + \Delta {w^-}^\top c^- + \Delta {w^+}^\top c^+ + \Delta {w^-}^\top \Delta^- \Delta w^- + \Delta w +^\top \Delta + \Delta w^+ = 1'$$

Appendix III – Backtest Results (No Transaction Costs)



Appendix III – Backtest Results (Transaction Costs)



Appendix III – Backtest Results (Breakdown by Transaction Costs)