Optimal Execution in Hong Kong given a Market-On-Close Benchmark

Christoph Frei[*]
University of Alberta

Nicholas Westray
Imperial College London

This version: April 27, 2017
(Quantitative Finance 18 (2018), 655–671)

Abstract

For stocks traded on the Hong Kong Exchange, the median of five prices taken over the last minute of trading is currently chosen as the closing price. We introduce a stochastic control formulation to target such a median benchmark in an empirically justified model which takes the key microstructural features into account. We solve this problem by providing an explicit and efficient algorithm which even has applications beyond this paper as it can be used for the dynamic linear approximation of any square-integrable random variable. Implementing the algorithm on the stocks of the Hang Seng Index, we find an average improvement of around 6 % in standard deviation of slippage compared to an average trader's execution. We conclude by providing a novel decomposition of the trading risk into that which is intrinsic to the median benchmark and that due to execution.

1 Introduction

Over the past decade, there has been a huge increase in algorithmic trading in the global equity markets. As a result, all large brokers now offer and market a suite of trading algorithms to clients for executing orders targeting different trading benchmarks. The standard offering includes arrival price, VWAP (volume weighted average price), TWAP (time weighted average price), POV (percentage of volume) as well as several auction strategies for MOO (market on open) and MOC (market on close) benchmarks. Such algorithms typically work by attempting to optimally schedule trades to achieve an average price close to the benchmark selected by the client. Here, close is often defined in terms of expected slippage or standard deviation of slippage (or even some combination of the two).

More concretely, if we denote the benchmark price by $\hat{P}$, the aim is to minimise some function $F(S)$ where S is the order slippage, defined (for a buy order) as the random variable $S = \bar{P}^w - \hat{P}$ where $\bar{P}^w$ is the average price obtained on the order when it is executed using strategy w. This is often formulated as a stochastic control problem in discrete or continuous time. Given the importance of such problems in equity trading, there has been a large amount of mathematical attention focussed on this area. The question of optimal execution with an arrival price benchmark is particularly well studied in the literature, going back to Almgren and Chriss [3]. For a collection of relevant literature in this topic and excellent

[*]corresponding author; Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton AB T6G 2G1, Canada; cfrei@ualberta.ca

overviews of algorithmic trading, market liquidity and microstructure, we refer to the recent books by Cartea et al. [11], Guéant [18] as well as Lehalle and Laruelle [28]. Thanks to work by Busseti and Boyd [9], Cartea and Jaimungal [10], Frei and Westray [15], as well as Guéant and Royer [19], execution problems with a VWAP benchmark are now better understood. Constraints on the execution strategy in terms of POV have also been studied in the literature, for example, by Guéant [17], and Labadie and Lehalle [24]. In contrast, the problem of trading in the auctions (specifically the closing auction) has been, with the notable exception of Bacidore et al. [4], largely overlooked in the literature to date. In many ways, this is understandable; in almost all developed markets, the closing price is derived from a standard auction and so one either enters the auction and gets the closing price or does not, and the only question is what percentage of one's order should be placed in the auction. This can be addressed by providing an accurate description of the microstructural features which traders should be aware of when participating in the closing auction (and the period immediately preceding it); see Bacidore et al. [4].

In addition to almost all developed markets, many exchanges in the emerging markets use closing auctions. There are however some exceptions, amongst others the exchanges of India, Egypt, Mexico and Shanghai, which use a period VWAP as the closing price. Given that this benchmark is of VWAP type, mathematically it is no more complex than executing a standard VWAP order and is thus covered by the approach in [15]. In this article we focus attention on the Hong Kong Exchange, unique as the only exchange which chooses the median of 5 prices taken over the last minute as the closing price. This benchmark, to the best of our knowledge, has, until now, not been studied in the literature and thus motivates the question we address, namely how should one trade in order to optimally target a median benchmark? This is a novel mathematical problem with direct applications in trading. The Hong Kong Exchange is one of the largest and most liquid exchanges worldwide, with approximately the same market capitalization and trade volume as Euronext, the largest exchange in continental Europe. In 2015, the Hong Kong Exchange was the 7th largest exchange worldwide in terms of market capitalization and the 8th largest exchange in terms of trade volume.[1] Recently, Chinese regulators approved a 'stock link' between the Hong Kong Exchange and the Shenzhen Stock Exchange (see for example Lockett [29]), further opening China's capital market and attracting more international investors. Therefore, the Hong Kong Exchange certainly merits its own specific closing auction logic and thus sets the scope of the present article.

To further highlight the relevance of this problem, we note that the history of the choice of median as the closing price is not without controversy. Indeed on May 26, 2008, the Hong Kong Exchange introduced a closing auction session to replace the closing price calculation based on the median price; see [21]. However, there were significant fluctuations in the closing prices and even suspicions of market manipulation. Notably, the stock price of HSBC, at the time the second-biggest stock listed on the benchmark Hang Seng Index, plunged by 11 % in the last few seconds of the closing auction session of March 9, 2009, resulting in a total drop of 24 % for that day. The plunge in the closing auction was mainly due to a large sell order without quoting a price, but there was no intention to manipulate the market as later ruled by the Securities and Futures Appeals Tribunal of Hong Kong. The following day, the stock price of HSBC rebounded by 14 %. As a consequence of these issues, the closing auction was suspended on March 23, 2009, and the bourse returned to calculating the closing price based on the median price.

[1] According to the statistics of the World Federation of Exchanges at http://www.world-exchanges.org

Furthermore, it is also the case that different segments of market participants have different views on the choice of median as a benchmark. Large institutional investors (whose performance is typically benchmarked to the close) are unhappy with the excess volatility created by the present choice of benchmark (which is regarded as "calculated" and not "tradeable") and advocate a more classical auction mechanism (as seen in Europe/USA). Indeed, a survey by Deutsche Bank in 2012 found that its institutional clients unanimously asked for an improved system for automatically matching trades at the end of the day, as Himaras [20] reported in Bloomberg; see also Lee [27].[2]

On the other hand, some local independent brokers oppose closing auctions because they fear that a closing auction would be less fair for small players as well as potentially allowing the price volatility as seen in HSBC; compare Himaras [20]. Finally we note that the Hong Kong Exchange is again evaluating the advantages and disadvantages of the median price calculation compared to a closing auction session; see the consultation paper [22] so that the research here is also topical. Of course, the question of the use of auctions in improving market quality is broader than the context studied here. Indeed, the empirical literature provides strong support for the use of auctions; for example, enhanced market quality by introducing auctions were reported by Pagano and Schwartz [32] for the Paris Bourse, by Comerton-Forde et al. [13] for the Singapore Exchange, and by Barclay et al. [5] for NYSE's centralized opening call auction.

The first question we address in this article is quantifying the excess volatility created by the median price benchmark. We analyse to which extent this can be reduced by choosing an appropriate optimal strategy. Even when using such a strategy, there is some intrinsic *benchmark risk* which cannot be mitigated by intelligent execution. To measure this benchmark risk, we first assume that the algorithm/trader can achieve exactly the prices used in the median calculation. This is clearly a simplification and we go on to consider a more realistic extension where we allow for uncertainty in the execution price, which is due to the fact that a trader cannot hit exactly the prices used in the median calculation.[3] This affects the volatility of slippage against the benchmark and leads to *execution risk*, in addition to the benchmark risk. Analysing the two sources of risk is interesting as it helps provide a decomposition of the different components of slippage and provides insight into the execution process during the auction in Hong Kong. Moreover, the article collects key microstructural features of the Hong Kong market which are relevant for practitioners when deciding how to trade at the close of the trading day.

The mathematical contribution of the present paper is twofold. While we specifically analyse the optimal strategy for the median price benchmark, our solution is rather generic. Indeed, we provide an optimal solution for the dynamic approximation of any L^2 random variable via an elegant recursion as well as an efficient tree based algorithm to calculate the optimal control in a problem that is fully non-Markovian. We also show empirically for the stocks in the Hang Seng Index that our model is parsimonious and captures the key aspects of the problem at hand whilst being numerically tractable.

The structure of the article is as follows. We first describe the model as well as the specifics of the closing price calculation in Hong Kong. Section 3 derives the optimal strategy as the

[2]The importance of auctions in Asian stock exchanges has also been highlighted by Gary Stone, Tom Kingsely and Gabriel Kan; compare http://issuu.com/fixglobal/docs/2015-q3/c/sc5afq2.

[3]The impact of such differences between desired and effective trade execution has also been studied in a different context of high-frequency trading: Moallemi and Sağlam [31] model and analyse the cost of latency in high-frequency trading.

explicit recursion for efficiently constructing the solution. Section 4 calibrates this model to data and analyses the results cross-sectionally across the Hang Seng Index. Section 5 concludes. Data to the optimal strategy for all the stocks in the Hang Seng Index are contained in Appendix A, and Appendix B compares our optimal stochastic strategy with deterministic strategies.

2 HK Auction Price Calculation and Model Formulation

According to the rules of the Hong Kong (HK) Exchange [23], the closing period begins at 15:59 and ends at 16:00 (HK local time). The closing price is constructed by taking the median of 5 nominal prices, with the first nominal price taken at 15:59:00 and then every 15s until the last at 16:00:00. Each nominal price is computed as

$$P_i = \min\{\max\{P_i^t, P_i^b\}, P_i^a\},$$

where P_i^t is the last traded price and (P_i^b, P_i^a) are the prevailing bid and ask prices at the snap time (indexed by i). We thus work on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_{\{1,..,5\}}, \mathbb{P})$ and consider a trader who may choose a vector w (understood as the fraction of the total order to be bought at time i) with w_i, $\mathcal{F}_{i-1}$-measurable[4] and aims to minimise

$$\mathbb{E}\left[\sum_{i=1}^{5} w_i P_i^r - \underset{j=1,\ldots,5}{\mathrm{median}}(P_j)\right] + \lambda \mathrm{Var}\left(\sum_{i=1}^{5} w_i P_i^r - \underset{j=1,\ldots,5}{\mathrm{median}}(P_j)\right), \quad \sum_{i=1}^{5} w_i = 1. \tag{1}$$

We use the notation P_i^r to indicate the realized price that the algorithm achieves at time i since we may not be able to achieve exactly P_i because of execution risk. Note that we formulate this problem from the perspective of a buyer. For a sell order, the formulation and the result are analogous. In order that all objects are well defined in the above, we make the standard assumption

$$\max_{j=1,\ldots,5} \mathbb{E}\left[P_j^2 + (P_j^r)^2\right] < \infty.$$

Given a description for the realized prices, the model will be completely specified, and we can begin to address the questions posed in the introduction. To that end, we consider the differences $\tilde{Y}_j = P_j - (P_j^b + P_j^a)/2$ between nominal prices and mid prices. In Section 4.2, we show that the nominal prices almost always (around 94 % on average) occur at the bid or at the ask. Therefore, we can assume that $\tilde{Y}_j = \delta Y_j$ where δ is half the spread and Y_j are random variables valued in $\{-1, 1\}$. In other words, we have

$$P_j = \frac{P_j^b + P_j^a}{2} + \delta Y_j.$$

Note that we do *not* assume that the Y_j are i.i.d. because their dependence structure and different values for $\mathbb{P}[Y_j = 1]$ are important features as we will see from data in Section 4.2. Denoting the changes in mid prices by Z_i, we can write the prices as

$$P_j = P_0 + \sum_{i=1}^{j} Z_i + \delta Y_j.$$

[4]Since the trader needs to choose w_i at time $i-1$, the information at time i and later cannot be used. This means that w_i can depend only on the information up to time $i-1$, hence it is $\mathcal{F}_{i-1}$-measurable.

Since the Z_i are mid price changes, we assume, in contrast to the Y_j, that they are i.i.d. Similarly to the nominal prices, the realized prices are given by

$$P_j^r = P_0 + \sum_{i=1}^{j} Z_i + \delta\theta_j,$$

where we again assume that the random variables θ_j are valued in $\{-1, 1\}$, but they do not need to be identical to Y_j. The strength of this approach is that by using different θ_j, we can explore different aspects of the problem. Specifically, there will be two cases we are interested in.

1. $\theta_j \equiv Y_j$: *Trading without execution risk.*

   In this case, the nominal prices are tradeable and the optimal solution measures the best tradeable approximation to the median benchmark. The resulting standard deviation of slippage should be thought of as the *benchmark risk* due to its construction as a median. We will compare this to a strategy based on VWAP (volume weighted average price) to quantify the improvement by using an appropriately designed optimisation. We will assign values (in basis points and spread units) to the increased risk borne by traders due to the median benchmark.

2. $\theta_j \not\equiv Y_j$: *Trading with execution risk.*

   This corresponds to a situation where the Y_j cannot be captured exactly. One might think of this intuitively via the following example. We are a buyer and we will cross the spread. Ideally, we should be the last person to do this prior to the snap time so that this is incorporated into the median calculation, but many other traders are trying to do this as well. Therefore, we will realize a price P_j^r which is close to P_j but may not be exactly the same. This means that the underlying mid price is the same, but we may be on the other side of the mid price than P_j. In other words, there is some probability that $\theta_j = -Y_j$ rather than $\theta_j = Y_j$. In the implementation in Section 4.2, we will estimate this probability based on data and conditional on the realizations of $Y_1, Y_2, \ldots, Y_j$.

The reader will note that we have not considered market impact in the above. This is because the Hong Kong market is characterized by having very wide spreads and large level 1 sizes in the limit order book, so that assuming that an order can be executed at bid or ask price is sensible. If we were to add a linear impact term, it would not change the mathematical treatment, but make the presentation more cumbersome; see Remark 3.2.[5] One may also ask whether a discrete formulation is appropriate here or whether by allowing continuous trading would improve performance. Taking aside the increased complexity of a continuous-time model, under the assumption that the drift is close to zero and observing that by trading at prices different in time from the snap prices will only increase the variance from the benchmark, it is difficult to make an intuitive argument as to why we would expect to improve our performance by increasing the strategy space in this way.

[5] There is also the question of what to do when your order is a very large multiple of that typically traded in the last minute. We omit that here, but a practical (and widely adopted) solution is to trade a piece of your order prior to the close minute and then perform the optimal solution described here.

Of course, by scaling the λ, we can indicate whether there is an effect of varying risk aversions on the behaviour in (1). In its present form the problem is not tractable as we have the standard issue related to terms of the form $\mathbb{E}^2[\cdot]$. To address this, we write

$$\operatorname{Var}\left(\sum_{i=1}^{5} w_i P_i^r - \operatorname*{median}_{j=1,\dots,5}(P_j)\right) = \mathbb{E}\left[\left(\sum_{i=1}^{5} w_i P_i^r - \operatorname*{median}_{j=1,\dots,5}(P_j)\right)^2\right] - \mathbb{E}^2\left[\sum_{i=1}^{5} w_i P_i^r - \operatorname*{median}_{j=1,\dots,5}(P_j)\right].$$

The first term is clearly amenable. To simplify the second term, we expand further and look at

$$\mathbb{E}^2\left[\operatorname*{median}_{j=1,\dots,5}(P_j)\right] - 2\mathbb{E}\left[\operatorname*{median}_{j=1,\dots,5}(P_j)\right]\mathbb{E}\left[\sum_{i=1}^{5} w_i P_i^r\right] + \mathbb{E}^2\left[\sum_{i=1}^{5} w_i P_i^r\right].$$

Observing that the median is independent of the control, we can drop the first term. For the second and third terms, note that if the w is deterministic the critical term would be of the form $\sum_{i=1}^{5} w_i \mathbb{E}\big[P_i^r\big]$. If we assume that the Z_i have approximately mean zero (which is intuitively sensible as they are mid price changes, compare Table 1), then we are left only with the means of θ_i. However, these means are approximately constant like $\mathbb{E}[Y_i]$, as displayed in Table 1. Thus we are led to an approximation of our problem as

$$\begin{aligned}
&\min_{w}\ \mathbb{E}\left[\left(\sum_{i=1}^{5} w_i P_i^r - \operatorname*{median}_{j=1,\dots,5}(P_j)\right)^2\right] + \frac{1}{\lambda}\mathbb{E}\left[\sum_{i=1}^{5} w_i P_i^r - \operatorname*{median}_{j=1,\dots,5}(P_j)\right]\\
&\text{subject to } \sum_{i=1}^{5} w_i = 1.
\end{aligned}$$

A quick note is in order here. The approximation above is critical to allow the problem to become tractable, else one ends up with a time inconsistent problem. Such an approximation is suitable for general stochastic strategies, as pointed out in Almgren [2], Cartea et al. [11] and Frei and Westray [15], who use similar approximations. In particular, the approximation is exact for deterministic strategies, and Frei and Westray [15] show that the error from the approximation is negligible even when using stochastic strategies. We should point out that the approach above does not prohibit a complicated dependency structure amongst the $(Y_i)_{i=1,\dots,5}$ or $(\theta_i)_{i=1,\dots,5}$. The approximation requires only that the *unconditional* means be constant. One could think of a 5-dimensional zero mean Gaussian as an illustration, where simply asking that the mean be 0 still allows full freedom in choosing the covariance matrix.

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$\mathbb{E}[Z_i]$ (in HK\$)	0.001	0.002	0.002	-0.002	-0.001
$\mathbb{E}[Y_i]$	0.128	0.149	0.129	0.131	0.113

Table 1: Means of Z_i and Y_i for the different periods $i = 1, \dots, 5$ based on the data set of Section 4: all stocks in the Hang Seng Index from April 1, 2014 to March 31, 2015. The means of Z_i are close to zero and the means of Y_i are approximately constant. The standard errors are 0.001 for Z_i and 0.01 for Y_i in each period. The means of Y_i are positive, corresponding to $\mathbb{P}[Y_i = 1] > 0.5$, which we will return to in the analysis of Section 4.

Remark 2.1. *A very interesting stream of recent literature deals with using and developing techniques from statistical learning to optimal trading; see for example Agarwal et al. [1] and Laruelle et al. [25] and [26]. A main idea in this area is that the trader is learning while trading by successively adjusting the strategy based on observations and results from their own earlier trades. In our problem under consideration, a trader submits orders at a time close to that of the snap prices for the median price calculation since the trader targets the median price. These prices occur at 15 seconds difference so that the trader may learn more from what is happing on the market in-between the 15 seconds, rather than from their own trade execution. Therefore, in the implementation in Section 4.2, the trader will take the market trades that are happening between the snap prices into consideration and learn from them, rather than their own trading. The market trades between the snap prices will enter the problem formulation through a suitable construction of the θ_i in Section 4.2.*

3 The Optimal Solution

Suppose we now generalise the above problem to an arbitrary time horizon N.

$$\begin{aligned} &\min_w \mathbb{E}\left[\left(\sum_{i=1}^N w_i P_i^r - \operatorname*{median}_{j=1,\ldots,N}(P_j)\right)^2\right] + \frac{1}{\lambda}\mathbb{E}\left[\sum_{i=1}^N w_i P_i^r - \operatorname*{median}_{j=1,\ldots,N}(P_j)\right] \\ &\text{subject to } \sum_{i=1}^N w_i = 1. \end{aligned} \tag{2}$$

This problem cannot be solved in closed form, but we can derive a solution algorithm, which explicitly determines the optimal weights in recursive steps. This procedure is here better suited than analysing the dynamic value function, which is an alternative solution approach, but would lead here to a numerical optimisation problem over a continuously valued parameter. Our procedure has the advantage that it provides a fast and easily implementable algorithm, which we apply to data in the next section.

We first use the constraint $\sum_{i=1}^N w_i = 1$ to replace w_N by $1-\sum_{i=1}^{N-1} w_i$ so that (2) becomes

$$\mathbb{E}\left[\left(\sum_{i=1}^{N-1} w_i(P_i^r - P_N^r) - \left(\operatorname*{median}_{j=1,\ldots,N}(P_j) - P_N^r\right)\right)^2 + \frac{1}{\lambda}\sum_{i=1}^{N-1} w_i(P_i^r - P_N^r) - \frac{1}{\lambda}\left(\operatorname*{median}_{j=1,\ldots,N}(P_j) - P_N^r\right)\right].$$

Expanding and using that all w_i are $\mathcal{F}_{N-2}$-measurable, we can write this as

$$\mathbb{E}\left[\sum_{i,j=1}^{N-1} w_i Q_{i,j}^{(N-1)} w_j - 2\sum_{i=1}^{N-1} R_i^{(N-1)} w_i + r^{(N-1)}\right], \tag{3}$$

where we define

$$Q_{i,j}^{(N-1)} = \mathbb{E}[(P_i^r - P_N^r)(P_j^r - P_N^r)|\mathcal{F}_{N-2}], \tag{4}$$

$$R_i^{(N-1)} = \mathbb{E}\left[(P_i^r - P_N^r)\left(\operatorname*{median}_{j=1,\ldots,N}(P_j) - P_N^r\right) - \frac{1}{2\lambda}(P_i^r - P_N^r)\middle|\mathcal{F}_{N-2}\right], \tag{5}$$

with $r^{(N-1)}$ not depending on w. The matrix $Q_{\cdot,\cdot}^{(N-1)}$ is symmetric and we assume that it is positive definite, which is equivalent to

$$\mathbb{E}\left[\left(\sum_{i=1}^{N-1} x_i P_i^r - P_N^r\right)^2\middle|\mathcal{F}_{N-2}\right] > 0 \tag{6}$$

for all x_i with $\sum_{i=1}^{N-1} x_i = 1$. In other words, it is not possible to replicate P_N^r as a convex combination of the earlier P_i^r on an $\mathcal{F}_{N-2}$-measurable set. This is a very weak condition and is satisfied when price changes include randomness. We now analyse (3) backward in time, starting with $N-1$. The first-order condition for w_{N-1} yields

$$w_{N-1} = \frac{R_{N-1}^{(N-1)} - \sum_{i=1}^{N-2} w_i Q_{i,N-1}^{(N-1)}}{Q_{N-1,N-1}^{(N-1)}}.$$

Plugging this into (3) and simplifying, we can obtain

$$\mathbb{E}\left[\sum_{i,j=1}^{N-2} w_i Q_{i,j}^{(N-2)} w_j - 2\sum_{i=1}^{N-2} R_i^{(N-2)} w_i + r^{(N-2)}\right], \tag{7}$$

where we define

$$Q_{i,j}^{(N-2)} = \mathbb{E}\left[Q_{i,j}^{(N-1)} - \frac{Q_{i,N-1}^{(N-1)} Q_{N-1,j}^{(N-1)}}{Q_{N-1,N-1}^{(N-1)}} \middle| \mathcal{F}_{N-3}\right], \quad R_i^{(N-2)} = \mathbb{E}\left[R_i^{(N-1)} - \frac{Q_{i,N-1}^{(N-1)} R_{N-1}^{(N-1)}}{Q_{N-1,N-1}^{(N-1)}} \middle| \mathcal{F}_{N-3}\right],$$

with $r^{(N-2)}$ not depending on w. Importantly, (7) is of the same quadratic structure as (3), but with one term less in the sums involving w_i. Therefore, we can continue in the same way by analysing (7) backward for $N-2, \ldots, 1$. This yields $Q^{(N-3)}, R^{(N-3)}, \ldots, Q^{(1)}, R^{(1)}$. Note that the $Q^{(k)}$ are symmetric $(k \times k)$-dimensional matrices, which are positive definite. The latter follows because $Q_{\cdot,\cdot}^{(k+1)} - \frac{Q_{\cdot,k+1}^{(k+1)} Q_{k+1,\cdot}^{(k+1)}}{Q_{k+1,k+1}^{(k+1)}}$ is the Schur complement of the $(k \times k)$-restriction in $Q^{(k+1)}$. The Schur complement of a positive definite matrix is again positive definite; see A.5.5 in Boyd and Vandenberghe [8]. Therefore, also its conditional expectation $Q^{(k)}$ is positive definite. Hence, we obtain the following characterisation of the optimal weights.

Theorem 3.1. *The optimal weights w_j are found explicitly following the procedure of Algorithm 1.*

While there is existing literature on estimating and selecting the median (see for example Dor and Zwick [14] and Manku et al. [30]), our algorithm provides an optimal dynamic and stochastic approximation to the median in a general probabilistic framework. Observe that Algorithm 1 is rather generic and describes the best dynamic linear approximation of any square-integrable random variable. We can see this algorithm as a non-Markovian generalisation of extended linear-quadratic control problems. The latter are of the form

$$\min \sum_{n=0}^{N-1} L_n(x_n, u_n) + L_N(x_N)$$

under an affine constraint of the form $x_{n+1} = A_n x_n + B_n u_n + b_n$, where the cost functions L_n have quadratic, linear and constant terms; see for example Frison and Jørgensen [16]. Those problems can be solved iteratively in a Riccati recursion using the Schur complement in each step. In our case, the affine constraint is generalised to the condition that the weights are predictable, which leads to a recursive use of conditional expectations in Algorithm 1.

Initialisation:

$$Q^{(N-1)}_{\cdot,\cdot} = \mathbb{E}\big[(P^r_\cdot - P^r_N)(P^r_\cdot - P^r_N)^\top \big| \mathcal{F}_{N-2}\big]$$
$$R^{(N-1)}_\cdot = \mathbb{E}\Big[(P^r_\cdot - P^r_N)\Big(\operatorname*{median}_{j=1,\ldots,N}(P_j) - P^r_N\Big) - \frac{1}{2\lambda}(P^r_\cdot - P^r_N)\Big|\mathcal{F}_{N-2}\Big]$$

Backward iteration:

for $k = N-2, \ldots, 1$ **do**

$$Q^{(k)}_{\cdot,\cdot} = \mathbb{E}\left[Q^{(k+1)}_{\cdot,\cdot} - \frac{Q^{(k+1)}_{\cdot,k+1} Q^{(k+1)}_{k+1,\cdot}}{Q^{(k+1)}_{k+1,k+1}}\middle|\mathcal{F}_{k-1}\right],$$
$$R^{(k)}_\cdot = \mathbb{E}\left[R^{(k+1)}_\cdot - \frac{Q^{(k+1)}_{\cdot,k+1} R^{(k+1)}_{k+1}}{Q^{(k+1)}_{k+1,k+1}}\middle|\mathcal{F}_{k-1}\right],$$

end

Forward iteration:

for $\ell = 1, \ldots, N-1$ **do**

$$w_\ell = \frac{R^{(\ell)}_\ell - \sum_{i=1}^{\ell-1} w_i Q^{(\ell)}_{i,\ell}}{Q^{(\ell)}_{\ell,\ell}}.$$

end

$$w_N = 1 - \sum_{i=1}^{N-1} w_i$$

Algorithm 1: Computation of the optimal weights

Remark 3.2. *1) We could easily incorporate a linear temporary price impact in our model, without altering the derivation of Algorithm 1. Indeed, if we assume that realized prices are given by*

$$P^{r,w}_j = P_0 + \sum_{i=1}^{j} Z_i + \delta\theta_j + \kappa w_j,$$

where κ is the coefficient of linear temporary market impact. We assume that the market impact affects only the mean, but not the variance in (1)*. This is sensible because the influence of price impact on variance is small. We can follow the same derivation as above and obtain Algorithm 1, with the difference that $Q^{(N-1)}_{\cdot,\cdot}$ and $R^{(N-1)}_\cdot$ are replaced by*

$$\tilde{Q}^{(N-1)}_{\cdot,\cdot} = Q^{(N-1)}_{\cdot,\cdot} + \frac{\kappa}{\lambda}\begin{pmatrix} 2 & 1 & \ldots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & & 1 \\ 1 & \ldots & 1 & 2 \end{pmatrix}, \qquad \tilde{R}^{(N-1)}_\cdot = R^{(N-1)}_\cdot + \frac{\kappa}{\lambda}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

2) In principle, the weights computed in Algorithm 1 could result in negative values. We analysed the number of negative weights resulting from the data set used in Section 4. The negative weights corresponded to less than 0.1 % of all weights, hence they are, for all practical purposes, negligible. Indeed, the tractability gains from having a simple recursive algorithm to compute the solution, outweighs the downside from simply taking 0 when a negative weight occurs and adjusting the other weights accordingly.

4 Numerical Results

We first explain the implementation of the algorithm and describe the data set. Section 4.2 justifies our chosen model specification and Section 4.3 addresses the questions posed in the introduction.

4.1 Implementation and Data Set

The algorithm above involves repeated evaluations of conditional expectations numerically and we thus have two methods at our disposal: trees or regressions. Due to the low dimensionality of our problem, we choose the tree method and opt for a trinomial discretisation of the random variables Z_i combined with the binomial random variables $Y_i, \theta_i \in \{-1, 1\}$. To be clear, we use the full empirical distribution of the Y_i, and then appropriately train a scheme for discretizing the Z_i. Since we have 5 steps and 12 branches at each step this leads to $12^5 \approx 250K$ nodes which is easily manageable numerically. We calculate the whole tree at the beginning, then take the first step in the tree according to the realized Y_1 and θ_1 and the realized Z_1 projected onto the nearest grid point. There are then two possibilities to calculate the control as we move further through the nodes:

1. Continue moving on the calculated tree by following the nodes according to the realized Y_i and θ_i and the realized Z_i projected onto the nearest grid point.

2. At each step i, we re-calculate the tree forward from step i using the true observed price P_{i-1} and applying the backward and forward recursions of Algorithm 1 back to and forward from i.

Mathematically, the only difference of the two methods is how the conditional expectations in the algorithm are calculated; the use of the algorithm and the constraint on the weights are the same. We opt for the second method because this gives more precise values for the conditional expectations and hence a better performance. The computational effort is still easily manageable as there are only three re-calculations of smaller and smaller trees needed (5 weights, but the first weight is determined from the original tree calculation, and the last weight is given as the remainder).

We used data from Bloomberg[6] during a one-year period from April 1, 2014 to March 31, 2015 and took all the trade and quote data for all the names in the Hang Seng Index on every day. The first step was universe construction, for which we chose those 50 names which were members of the Hang Seng Index on March 31, 2014.[7] We compute all the empirical

[6] All data is used with permission of Bloomberg L.P.

[7] During the period, there was only one change in the Hang Seng Index: COSCO Pacific was replaced by Link REIT on December 8, 2014. For consistency, we kept the same names during the whole period, but making this change would not affect our conclusions.

Z_i and Y_i from the data for each stock and day pair, using the prescription described in Section 2. We take the price 15s before the first nominal price as P_0 to keep equal time spacing between the increments. Therefore, we only require the last 1:15 minutes of trading on each day, which massively reduces the size of the data. We use the first six months (April 1 to September 30, 2014) to calibrate our model as will be described in Section 4.2 while we use the other six months (October 1, 2014 to March 31, 2015) of the data set to test the performance of the strategy from Algorithm 1. Keeping distinct training and test sets is standard in the statistical literature to minimise overfitting; the choice of equal training and test periods is for convenience only, and one could imagine a variant of the above where a rolling calibration approach would be used.

Remark 4.1. *Our method requires estimation of the points in the trinomial tree (6 parameters) which can easily be done from the approximately 125 trading days in a 6 month window for a given stock. We also require an estimate of the empirical distribution of the 5 dimensional vector* $(Y_1, Y_2, Y_3, Y_4, Y_5)$. *Since each component is binary,* $+1$ *or* -1, *we regard 125 observations of this as also being sufficient, if slightly on the low side. The calculation of the different path probabilities is highly efficient and whilst a shorter period is undesirable, it could easily be extended to a longer time horizon (either 12 or 18 months) with minimal overhead.*

4.2 Model Component Specification

Let us first consider our model for Y_i. For the stocks in the Hang Seng Index, the left panel of Figure 1 indicates the percentage of the time the variable Y_i is either ± 1. For the Hang Seng Index, it is 94 %. We conclude that this is an appropriate approximation for modelling purposes. Note that we are effectively assuming spread is constant during the auction period. This is of course a simplification, but we believe that whilst we have not captured all features, we have captured the main dynamics of interest.

Having established that the Y_i are ± 1 valued, the next important empirical observation to make is that they are not i.i.d. This is consistent with the analysis of Bouchaud et al. [7] who demonstrate strong autocorrelation of trade signs for stocks traded on EU/US exchanges. The right panel of Figure 1 shows the distribution of the different realizations of the number of up moves in the 5 periods, under two assumptions. The first is that they are i.i.d. with mean given by the empirical mean, which leads to a binomial distribution. The second is the actual empirical distribution. One sees quite significant differences and we view this as justification for the tree method we implement. Therefore, to calculate the conditional expectations in Algorithm 1, we will calculate empirical probabilities corresponding to all different possibilities of $\mathbb{P}[Y_j = \pm 1 | Y_1 = \pm 1, \ldots, Y_{j-1} = \pm 1]$ and not just $\mathbb{P}[Y_j = \pm 1]$. Interestingly, we also find that the probability of an up move is 0.56 and not 0.5. This is related to an imbalance of market orders (slightly more buy than sell orders). The order imbalance and its implications on price dynamics have been the interest of many studies; see for example Chordia et al. [12].

To determine θ_i, we distinguish the two cases from pages 5 and 6. In the first case, θ_i is identical to Y_i. In the second case, the possibility that θ_i takes the opposite sign of Y_i reflects execution risk. We model

$$\theta_i = \begin{cases} Y_i - 2U_i & \text{if } Y_i = 1, \\ Y_i + 2V_i & \text{if } Y_i = -1, \end{cases}$$

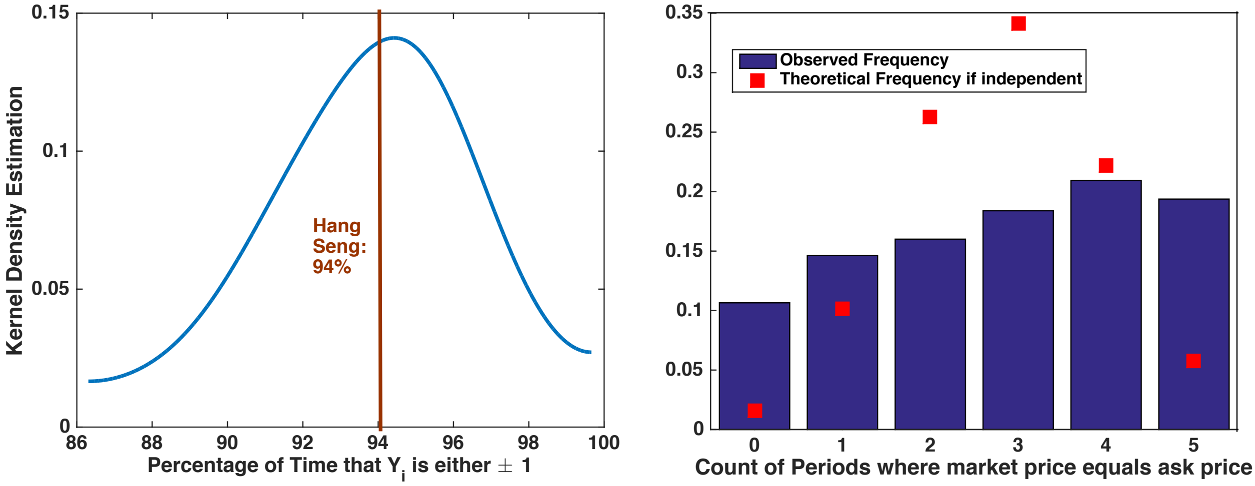

Figure 1: Left panel: The percentage that Y_i takes values ± 1 is high. The kernel density estimation is calculated across the 50 different names in the Hang Seng Index. The precise value for each name is reported in Table 3 in Appendix A.
Right panel: Empirical distribution versus theoretical distribution under independence assumption. The latter is a binomial distribution with $n = 5$ and $p = 0.56$, corresponding to the overall probability of an up move.

where U_i and V_i are Bernoulli random variables with $\mathbb{P}[U_i = 1] = u_i$ and $\mathbb{P}[V_i = 1] = v_i$. If we are a buyer, $\theta_i = -1$ is the favourable case. This occurs if either $Y_i = 1$ and we get a limit order executed ($U_i = 1$) or $Y_i = -1$ and we hit Y_i ($V_i = 0$). If all u_i and v_i were zero, we would have $\theta_i \equiv Y_i$ for all i and recover the first case. To estimate u_i and v_i, we analyse the trades in the 15s before Y_i occurs. For example, if $i = 4$, these are the trades between 15:59:30 and 15:59:45. Similarly to Y_i, they occur mostly at either bid or ask prices so that we can classify these trades by a finite sequence of ± 1. We set u_i as the ratio of the number of -1 compared to the total number of ± 1, conditional on $Y_i = 1$. We calculate v_i analogously, but conditional on $Y_i = -1$. Estimating u_i and v_i from data, we observe that their realizations also depend on the values $Y_1, \ldots, Y_{i-1}$. Therefore, we make different estimations of u_i and v_i, depending on the period i and the realizations of $Y_1, \ldots, Y_{i-1}$, but take the same estimations for all stocks in the Hang Seng Index.

In summary, the probabilities u_i and v_i are the fractions of the trades executed at ask and bid prices during a 15s interval conditional on the value at the end of the interval. Intuitively, calibrating them in this way is effectively a method of moments approximation to extracting the probability of getting a limit order executed given the value at the end of the interval and assuming that on average our execution price is equivalent to market VWAP over the period.[8]

In passing we highlight another consequence of the choice of median as the closing price. The left panel of Figure 2 shows the trading density in the last minute of the trading day. We see huge spikes around the snap prices at 15:59:00, 15:59:15, ..., 16:00:00. Traders are clearly (and understandably) attempting to have their trades enter the median calculation. This has implications for the exchange, as it needs to have infrastructure that can support

[8] We remark that if a sell order is considered, one needs to alter the definitions of u_i and v_i and the result will change due to the asymmetry of bid and ask executions during the last 1:15 minutes of trading.

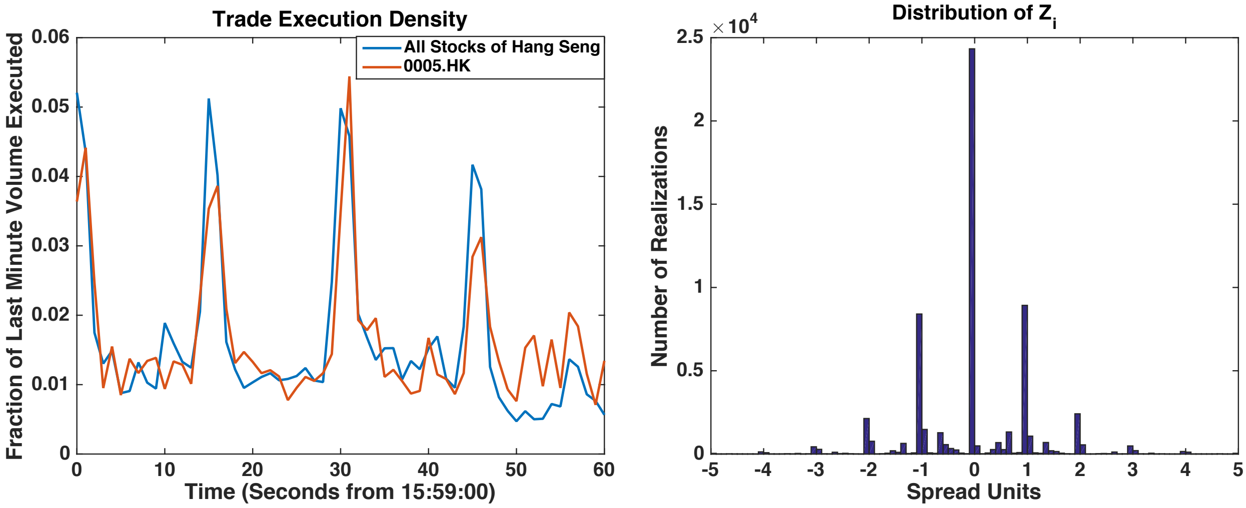

Figure 2: Left panel: The average daily trading density shows a huge rush to trade at the snap prices for the Hang Seng Index and, as example, HK.0005 (HSBC).
Right panel: Histogram of Z_i for all stocks in the Hang Seng Index. To make Z_i comparable across different stocks, its value is divided by the prevailing spread.

these huge bursts of messages, which will increase infrastructure costs.

Turning finally to the Z_i, we note that their empirical distribution is heavily centred and discrete, which is shown in the right panel of Figure 2. Given the dominance of just three values (representing around 73 % of all realizations) and since the increasing complexity does not justify a high-order approximation, we approximate the range of Z_i by the three most frequently occurring values for each stock. This is justified by the following auxiliary result.

Lemma 4.2. *Let Z be a random variable with cumulative distribution function F. Assume that there exist N values $z_1 < z_2 < \cdots < z_N$ such that*

$$F(z_i) - F(z_i-) > F(z_i-) - F(z_{i-1}), \qquad F(z_i) - F(z_i-) > F(z_{i+1}-) - F(z_i) \tag{8}$$

for all $i = 1, \ldots, N$, where we set $z_0 = -\infty$ and $z_{N+1} = \infty$. Let $\hat{Z}$ be a discrete random variable taking values $\hat{z}_1 < \hat{z}_2 < \cdots < \hat{z}_N$ with cumulative distribution function $\hat{F}$. Then, choosing $\hat{z}_i$ such that $\int_{-\infty}^{\infty} \left|F(z) - \hat{F}(z)\right|^p \mathrm{d}F(z)$ for some fixed $p > 0$ is minimised leads to $\hat{z}_i = z_i$ for all $i = 1, \ldots, N$.

The expression $F(z_i-)$ in (8) means the left limit $F(z_i-) = \lim_{z \nearrow z_i} F(z)$. Condition (8) says that Z takes the values z_i with high probability. Indeed, (8) implies that the probability of $Z = z_i$ is greater than the probabilities that Z takes values in (z_{i-1}, z_i) or (z_i, z_{i+1}). Lemma 4.2 then says that the best N-point approximation for Z is based on choosing exactly these values z_i. Lemma 4.2 can be proven by writing

$$\begin{aligned}\int_{-\infty}^{\infty} \left|F(z) - \hat{F}(z)\right|^p \mathrm{d}F(z) = \cdots &+ \int_{z_{i-1}}^{z_i-} \left|F(z) - \hat{F}(z)\right|^p \mathrm{d}F(z) \\ &+ \left|F(z_i) - \hat{F}(z_i)\right|^p \big(F(z_i) - F(z_i-)\big) \\ &+ \int_{z_i}^{z_{i+1}-} \left|F(z) - \hat{F}(z)\right|^p \mathrm{d}F(z) + \ldots \end{aligned} \tag{9}$$

Assume the minimising $\hat{z}_i$ satisfy $\hat{z}_i = z_i$ for all $i = 1, \ldots, j-1$ with $\hat{F}(\hat{z}_i) \in \big[F(z_i), F(z_{i+1}-)\big]$. If we now choose $\hat{z}_j = z_j$ for some j with $\hat{F}(\hat{z}_j) \in \big[F(z_j), F(z_{j+1}-)\big]$, then it can be checked that condition (8) implies that a deviation to either $\hat{z}_j < z_j$ or $\hat{z}_j > z_j$ leads to a higher value of (9). By induction, we obtain that the optimal choice is $\hat{z}_i = z_i$ for all i. The corresponding optimal values for $\hat{F}$ are given such that $\hat{F}(\hat{z}_i) \in \big[F(z_i), F(z_{i+1}-)\big]$ minimises

$$\big|F(z_i) - \hat{F}(\hat{z}_i)\big|^p \big(F(z_i) - F(z_i-)\big) + \int_{z_i}^{z_{i+1}-} \big|F(z) - \hat{F}(\hat{z}_i)\big|^p \, \mathrm{d}F(z).$$

Lemma 4.2 gives a theoretical justification why we use an approximation for Z_i based on the three most frequently occurring values of Z_i. Indeed, the three most frequent values of Z_i account for approximately 41 %, 16 % and again 16 % of all realizations (compare the right panel of Figure 2) while all the probabilities of values in-between the three most frequent realizations are significantly smaller so that we can apply Lemma 4.2 with $N = 3$. In fact, testing with out-of-sample realizations, we obtain relatively small approximation errors as reported in Table 4 in the Appendix A.

4.3 Main Findings

Recall from the introduction that the median benchmark has been heavily criticised due to it not being tradeable. To measure this and get an appropriate scale for our results, we look at the large λ limit in (2), i.e., where we are only focussed on having the smallest standard deviation (variance) vs. the benchmark. The motivation for this is twofold. Firstly, when institutional investors complain about the untradeability of the median, they are referring to the tracking error, best described via standard deviation (variance). Secondly, the mean portion of the problem can be thought of as slightly degenerate. To be more precise, if one is targeting best performance in terms of mean, due to $\mathbb{E}[Z_i] \approx 0$ (see Table 1), it is effectively optimal to place a limit order for the full quantity and then cross the spread at the end if one is not executed. As such this affords little insight into decomposing the different components of the slippage. To help interpret our results, we compare all the performances of the algorithm suggested here to that of a simple VWAP (volume weighted average price) strategy. VWAP is well known for being “fair” in the sense that it can be achieved by a random nonstrategic trader; see Berkowitz et al. [6].[9] Therefore, in our context, VWAP represents a simple benchmark that is easily accessible to all market participants, and in the sense that median is also an average (as opposed to arrival/last price or something similar), it is an appropriate data point for comparison and evaluation of our results.

For each stock in the Hang Seng Index, we calculate the standard deviation between VWAP and median price benchmark over the trading days between October 1, 2014 and March 31, 2015. The bars in Figure 3 show these values for five selected companies of the Hang Seng Index in terms of basis points of the initial prices P_0 at each day and the prevailing spread. Since the error is related to the spread, measuring it in terms of the prevailing spread allows for a more homogeneous presentation across different stocks; compare left and right panels of Figure 3. The five companies are Cheung Kong, HSBC, Cathay Pacific, Bank of China, and Petro China, which are chosen in order to have a good mixture of market

[9] We calculate the VWAP for each trading day and stock over the last 1:15 minutes of trading to achieve consistency: traded prices are taken at any 15 seconds window before the prices used in the median calculation, in line with how we estimated θ_i in Section 4.2.

capitalizations and sectors. The complete results for all stocks of the Hang Seng Index are contained in Table 5 in the Appendix A.

We now compare the results of the strategy from Algorithm 1 with those corresponding to VWAP. We start with the first case where all $\theta_i \equiv Y_i$, which means that the trader is able to hit exactly the prices entering the median calculation. Figure 3 demonstrates that the strategy from Algorithm 1 performs significantly better than VWAP among all stocks, with an average improvement of about 40 %. All index level averages are taken using market capitalization weights to better account for differences in traded notional.

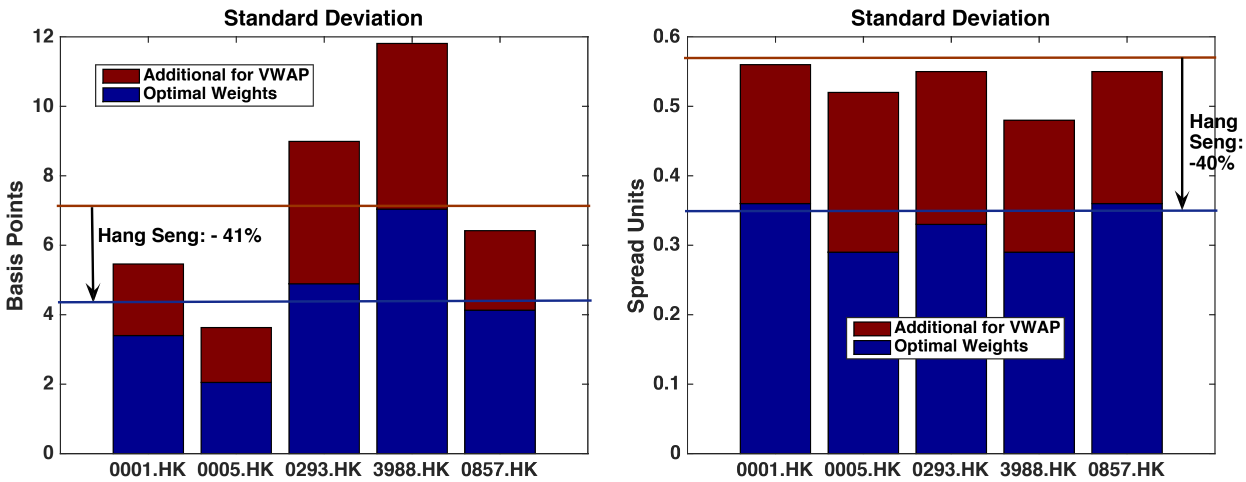

Figure 3: Standard deviations between median and strategies based on an average trader's execution (VWAP) and Algorithm 1 without execution risk. The numbers are in basis points of the initial prices (left panel) and in terms of prevailing spread (right panel). Displayed are Cheung Kong (0001.HK), HSBC (0005.HK), Cathay Pacific (0293.HK), Bank of China (3988.HK), and Petro China (0857.HK). The horizontal lines show the overall performance, based on the Hang Seng Index.

In reality, however, a trader cannot hit exactly the snap prices, and we now turn to the second case where the θ_i are based on the estimations described in Section 4.2 and where execution risk is taken into consideration. We added the corresponding results to the plots as green squares and lines, shown in Figure 4. There is now no more a uniform improvement across all names, for example, for Petro China, the performance of Algorithm 1 is worse than that of VWAP. The average improvement is about 6 %, compared to 40 % without execution risk.

In summary for the Hang Seng Index, our algorithm gives an average improvement from a VWAP strategy of around 6 % in standard deviation when execution risk is taken into account. The risk of such an optimal strategy can be understood as follows: 40 % of the variance is due to the difficulty in linearly approximating the median. The other 60 % of the variance is related to execution risk, due to not being able to execute at the prices used in the median calculation. The conclusion is therefore nuanced. The challenge in targeting a median benchmark is not exclusively that the median is inherently difficult to approximate; additionally, the wide spreads give traders a high margin for error in terms of the execution risk. Expressing this differently, the median is robust against outliers, but jumpy in small samples, which is what is present in this problem.

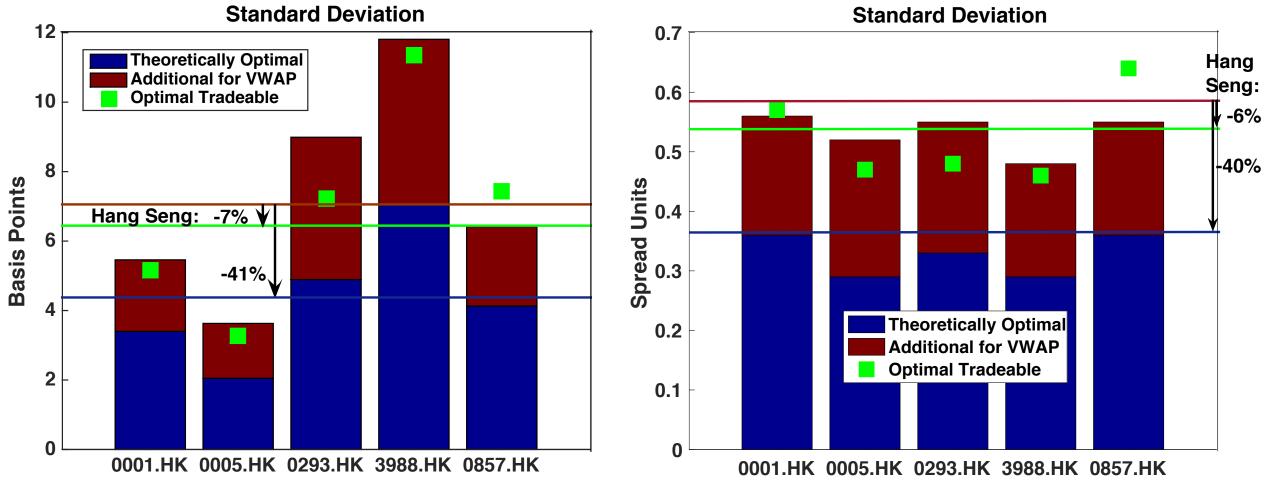

Figure 4: Additionally to the data from Figure 3, standard deviations between median and the strategy from Algorithm 1 with execution risk are displayed in green. For the Hang Seng Index, the improvement compared to VWAP is around 6 % when execution risk is taken into account.

The decomposition of risk in execution and benchmark risk components is fairly orthogonal. Indeed, let P^M, $P^{\text{incl.}}$ and $P^{\text{excl.}}$ be the median price, the realized price with execution risk, and the achieved price without execution risk, respectively. Then we can write

$$\underbrace{\mathrm{Var}\left(P^M - P^{\text{incl.}}\right)}_{(\text{total risk})^2} = \underbrace{\mathrm{Var}\left(P^M - P^{\text{excl.}}\right)}_{(\text{benchmark risk})^2} + \underbrace{\mathrm{Var}\left(P^{\text{excl.}} - P^{\text{incl.}}\right)}_{(\text{execution risk})^2} + \underbrace{2\,\mathrm{Cov}\left(P^M - P^{\text{excl.}}, P^{\text{excl.}} - P^{\text{incl.}}\right)}_{\approx 0}.$$

Figure 5 shows that the covariance term is very small compared to the two variance terms on the right-hand side of the equation. This means that the total risk can be decomposed orthogonally in the intrinsic benchmark risk and the execution risk. Therefore, we can consider benchmark and execution risk as two uncorrelated sources of risk.

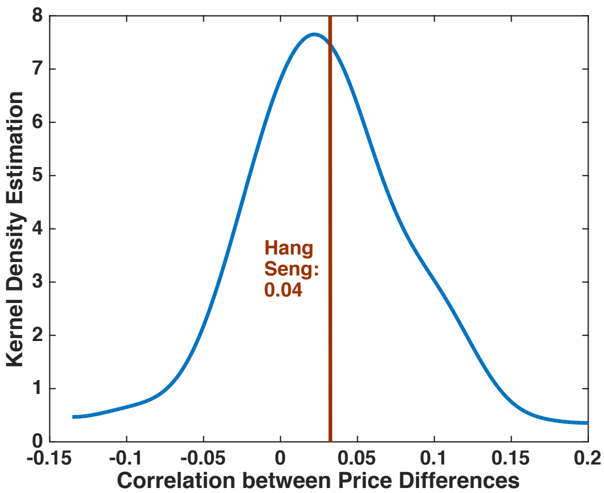

Figure 5: The empirical correlation between $P^M - P^{\text{excl.}}$ and $P^{\text{excl.}} - P^{\text{incl.}}$ is small, where P^M, $P^{\text{incl.}}$ and $P^{\text{excl.}}$ are the median price, and the realized prices with and without execution risk, respectively.

Name	*Optimal with execution risk*				*Optimal without execution risk*			
	bps		*spread*		*bps*		*spread*	
	det.	*stoch.*	*det.*	*stoch.*	*det.*	*stoch.*	*det.*	*stoch.*
HSBC Hldgs	5.52	3.27	0.77	0.47 (−39%)	2.15	2.05	0.30	0.29 (−5%)
Bank of China	18.87	11.36	0.77	0.46 (−40%)	7.49	7.04	0.30	0.29 (−6%)
Petro China	10.18	7.42	0.86	0.64 (−26%)	4.63	4.13	0.40	0.36 (−10%)
Cheung Kong	8.41	5.17	0.80	0.57 (−29%)	3.40	3.40	0.34	0.36 (+6%)
Cathay Pacific	13.29	7.24	0.82	0.48 (−41%)	5.46	4.89	0.36	0.33 (−9%)
Hang Seng	**10.83**	**6.56**	**0.84**	**0.55 (−35%)**	**4.51**	**4.19**	**0.38**	**0.35 (−8%)**

Table 2: Standard deviations between median and results from optimal deterministic (*det.*) and stochastic (*stoch.*) strategies. The numbers are in basis points (*bps*) of the starting prices at each day and in terms of prevailing *spread*. The numbers in parentheses show the reduction in standard deviation of the stochastic strategy to the corresponding values of the deterministic strategy.

4.4 Improvement thanks to Stochastic Learning

Our stochastic control approach is one way of solving this problem. To give some further context for the solution and investigate its appropriateness from a fitting perspective, we compare the numerical results from Section 4.3 with a deterministic solution to the problem (2). For a full derivation of the solution, please see Appendix B. In brief, the optimal weights of the deterministic solution are given by

$$(w_1, \dots, w_{N-1})^\top = A^{-1} b \quad \text{and} \quad w_N = 1 - \sum_{i=1}^{N-1} w_i,$$

where we define the matrix $A = (A_{k,i})_{1 \le k,i \le N-1}$ and the vector $b = (b_1, \dots, b_{N-1})^\top$ by

$$\begin{aligned} A_{k,i} &= \mathbb{E}[(P_k^r - P_N^r)(P_i^r - P_N^r)], \\ b_k &= \mathbb{E}\Big[(P_k^r - P_N^r)\Big(\operatorname*{median}_{j=1,\dots,N}(P_j) - P_N^r\Big) - \frac{1}{2\lambda}(P_k^r - P_N^r)\Big]. \end{aligned}$$

Using the same dataset and method as in Section 4.1, we analyse the performance of this optimal deterministic strategy and compare it with that of our optimal stochastic strategy from Algorithm 1. Table 2 shows the results of the comparison between these two strategies for the Hang Seng Index and a selection of five of its stocks (the same stocks as chosen in the analysis of Section 4.3). If execution risk is not considered, the stochastic strategy outperforms the deterministic one by around 8 % on average. In the more realistic case with execution risk, the average outperformance of the stochastic strategy is even around 35 %. This indicates that with our multi-period non-Markovian approach, we are able to better fit the data and offer clear improvements over a deterministic strategy, lending extra credence to the findings from Section 4.3.

Remark 4.3. *Under further assumptions, we can even derive an explicit formula for the improvement of the performance of the optimal stochastic weights compared to that of the optimal deterministic weights. Indeed, if we assume that θ_i are constant and $\mathbb{E}[Z_i] = 0$, the* *difference in squared errors between the two models compared to the media price benchmark is given by*

$$\frac{1}{Var(Z_1)} \sum_{i=1}^{N} Var\Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big).$$

In this formula, the improvement from using stochastic strategies is reflected in approximating $Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)$ by successive conditional expectations, rather than taking directly an expectation. While this formula holds under specific model assumptions, it shows the general idea behind the improvement: the stochastic strategy allows for using successive predictions of functionals of the median via conditional expectations. When we ran this on a numerical simulation, we found the relative improvement in standard deviation to be approximately 23 %. For further details, please see again Appendix B.

5 Conclusion

The algorithm derived in this paper not only gives an efficient way to optimally target a median benchmark, but also allows us to draw several relevant conclusions on trading in Hong Kong. The first is that even optimal stochastic weights only offer a small improvement compared to the execution of an average trader targeting the median benchmark because of the difficulty in targeting this median benchmark. Secondly, this difficulty is due to both risk inherent to the benchmark and execution risk. The latter is mainly a consequence of the trader's inability to trade at the exact price used in the benchmark calculation, combined with the occurrence of wide spreads. These risks for a trader should be borne in mind when defining a mechanism for the calculation of closing prices.

A Data Tables

Name	*RIC*	%	*Name*	*RIC*	%
HSBC Hldgs	0005.HK	96.8	China Unicom	0762.HK	95.8
Tencent	0700.HK	89.8	China Shenua	1088.HK	96.0
China Mobile	0941.HK	90.3	Want Want China	0151.HK	95.0
AIA	1299.HK	92.1	Lenovo Group	0992.HK	92.4
CCB	0939.HK	97.3	Hengan Int'l	1044.HK	87.4
ICBC	1398.HK	97.9	Swire Pacific 'A'	0019.HK	89.7
Bank of China	3988.HK	98.5	Hang Lung Prop	0101.HK	95.3
CNOOC	0883.HK	94.5	Li & Fung	0494.HK	94.7
Petro China	0857.HK	95.2	Henderson Land	0012.HK	92.6
Hutchison	0013.HK	92.4	New World Dev	0017.HK	91.9
Sinopec Corp	0386.HK	94.5	Bankcomm	3328.HK	96.8
HKEx	0388.HK	91.3	Mengniu Daily	2319.HK	93.9
Cheung Kong	0001.HK	90.3	Belle Int'l	1880.HK	95.3
China Life	2628.HK	95.3	Bank of E Asia	0023.HK	94.2
SHK Prop	0016.HK	93.9	China Res Power	0836.HK	93.7
Galaxy Ent	0027.HK	92.7	MTR Corporation	0066.HK	95.5
Ping An	2318.HK	92.3	Tingyi	0322.HK	94.7
CLP Hldgs	0002.HK	91.5	Kunlun Energy	0135.HK	95.3
Sands China	1928.HK	91.5	Sino Land	0083.HK	93.9
HK & China Gas	0003.HK	94.0	China Res Land	1109.HK	92.6
Hang Seng Bank	0011.HK	92.6	China Mer Hldgs	0144.HK	94.0
Power Assets	0006.HK	88.1	China Resources	0291.HK	96.0
BOC Hong Kong	2388.HK	95.5	CITIC Pacific	0267.HK	94.2
Wharf (Hldgs)	0004.HK	91.3	COSCO Pacific	1199.HK	94.4
China Overseas	0688.HK	96.0	Cathay Pacific	0293.HK	96.3

Table 3: The first and second columns are the names and RIC (Reuters Instrument Code) of the 50 companies of the Hang Seng Index, ordered by decreasing market capitalisation. The third column shows the percentages of the time that Y_i is either ± 1 (after diving by half the prevailing spread), with a value of 93.9 % for the Hang Seng Index overall.

Name	% error
HSBC Hldgs	6.0
Tencent	16.5
China Mobile	17.2
AIA	9.6
CCB	7.3
ICBC	4.3
Bank of China	2.6
CNOOC	7.2
Petro China	7.8
Hutchison	13.6
Sinopec Corp	7.5
HKEx	9.7
Cheung Kong	13.1
China Life	9.4
SHK Prop	17.4
Galaxy Ent	11.6
Ping An	14.5
CLP Hldgs	12.6
Sands China	11.9
HK & China Gas	10.0
Hang Seng Bank	11.7
Power Assets	19.8
BOC Hong Kong	9.2
Wharf (Hldgs)	12.5
China Overseas	8.2

Name	% error
China Unicom	11.4
China Shenua	8.0
Want Want China	12.1
Lenovo Group	8.7
Hengan Int'l	18.5
Swire Pacific 'A'	12.9
Hang Lung Prop	6.5
Li & Fung	16.7
Henderson Land	13.0
New World Dev	8.2
Bankcomm	7.9
Mengniu Daily	9.0
Belle Int'l	6.6
Bank of E Asia	8.5
China Res Power	7.1
MTR Corporation	7.2
Tingyi	12.1
Kunlun Energy	9.6
Sino Land	9.1
China Res Land	16.1
China Mer Hldgs	11.1
China Resources	14.2
CITIC Pacific	7.3
COSCO Pacific	6.3
Cathay Pacific	10.6

Table 4: Approximation error (9) of the distribution of Z_i with $p = 1$ in percentage of $\int_{-\infty}^{\infty} F(z)\,\mathrm{d}F(z) = 1/2$. Estimations are based on data for Apr.–Sept. 2014 while the realizations are for Oct. 2014–Mar. 2015. For the Hang Seng Index, this relative error is 10.3.

Name	*VWAP*		*Optimal with exec. risk*		*Optimal without exec. risk*	
	bps	*spr.*	*bps*	*spread*	*bps*	*spread*
HSBC Hldgs	3.63	0.52	3.27 (−10%)	0.47 (−9%)	2.05 (−44%)	0.29 (−44%)
Tencent	6.19	0.75	5.66 (−9%)	0.71 (−5%)	3.81 (−38%)	0.47 (−38%)
China Mobile	6.14	0.88	4.83 (−21%)	0.70 (−20%)	3.99 (−35%)	0.56 (−36%)
AIA	6.78	0.57	5.63 (−17%)	0.47 (−17%)	3.86 (−43%)	0.32 (−43%)
CCB	8.55	0.52	8.08 (−5%)	0.49 (−6%)	5.05 (−41%)	0.31 (−41%)
ICBC	9.37	0.50	8.90 (−5%)	0.47 (−6%)	5.40 (−42%)	0.28 (−44%)
Bank of China	11.81	0.48	11.36 (−4%)	0.46 (−4%)	7.04 (−40%)	0.29 (−40%)
CNOOC	9.52	0.52	9.06 (−5%)	0.50 (−4%)	5.63 (−41%)	0.31 (−41%)
Petro China	6.42	0.55	7.42 (+16%)	0.64 (+16%)	4.13 (−36%)	0.36 (−35%)
Hutchison	6.18	0.71	5.31 (−14%)	0.68 (−3%)	3.46 (−44%)	0.44 (−37%)
Sinopec Corp	7.79	0.46	7.63 (−2%)	0.47 (+2%)	4.79 (−38%)	0.29 (−36%)

. . . over

Name	*VWAP* *bps*	*spr.*	*Optimal with exec. risk* *bps*	*spread*	*Optimal without exec. risk* *bps*	*spread*
HKEx	3.65	0.60	2.93 (−20%)	0.50 (−18%)	1.65 (−55%)	0.26 (−56%)
Cheung Kong	5.46	0.56	5.17 (−5%)	0.57 (+1%)	3.40 (−38%)	0.36 (−36%)
China Life	9.42	0.50	8.84 (−6%)	0.47 (−6%)	5.52 (−41%)	0.29 (−41%)
SHK Prop	5.88	0.55	5.35 (−9%)	0.53 (−3%)	3.32 (−44%)	0.34 (−38%)
Galaxy Ent	6.87	0.54	6.30 (−8%)	0.51 (−5%)	4.01 (−42%)	0.33 (−39%)
Ping An	5.14	0.72	4.50 (−12%)	0.63 (−12%)	3.14 (−39%)	0.46 (−35%)
CLP Hldgs	5.48	0.54	5.15 (−6%)	0.56 (+4%)	3.31 (−40%)	0.33 (−38%)
Sands China	8.23	0.58	7.34 (−11%)	0.53 (−8%)	4.38 (−47%)	0.32 (−44%)
HK & China Gas	6.53	0.50	6.01 (−8%)	0.49 (−1%)	4.03 (−38%)	0.33 (−34%)
Hang Seng Bank	5.64	0.57	5.31 (−6%)	0.60 (+5%)	2.95 (−48%)	0.32 (−44%)
Power Assets	5.63	0.57	6.12 (+9%)	0.62 (+9%)	4.55 (−19%)	0.39 (−31%)
BOC Hong Kong	9.41	0.48	9.12 (−3%)	0.47 (−3%)	5.53 (−41%)	0.29 (−41%)
Wharf (Hldgs)	7.06	0.53	6.89 (−2%)	0.61 (+14%)	4.43 (−37%)	0.35 (−35%)
China Overseas	10.68	0.46	9.61 (−10%)	0.42 (−8%)	6.80 (−36%)	0.29 (−36%)
China Unicom	9.69	0.53	9.29 (−4%)	0.52 (−2%)	6.15 (−37%)	0.34 (−36%)
China Shenua	11.40	0.53	11.02 (−3%)	0.53 (+1%)	6.51 (−43%)	0.28 (−47%)
Want Want	9.36	0.58	9.36 (−0%)	0.67 (+16%)	4.42 (−53%)	0.29 (−51%)
Lenovo Group	11.78	0.61	9.29 (−21%)	0.48 (−22%)	6.45 (−45%)	0.32 (−47%)
Hengan Int'l	7.94	0.72	6.50 (−18%)	0.67 (−7%)	4.68 (−41%)	0.45 (−38%)
Swire Pacific 'A'	7.43	0.52	7.06 (−5%)	0.60 (+17%)	4.05 (−45%)	0.32 (−37%)
Hang Lung Prop	12.37	0.45	12.98 (+5%)	0.50 (+11%)	7.33 (−41%)	0.27 (−41%)
Li & Fung	7.90	0.52	7.79 (−1%)	0.55 (+5%)	4.36 (−45%)	0.29 (−45%)
Henderson Land	7.22	0.56	6.35 (−12%)	0.55 (−2%)	4.21 (−42%)	0.33 (−40%)
New World Dev	7.96	0.54	6.49 (−19%)	0.54 (−0%)	4.10 (−49%)	0.34 (−37%)
Bankcomm	8.14	0.45	7.55 (−7%)	0.45 (+1%)	4.48 (−45%)	0.25 (−43%)
Mengniu Daily	10.30	0.51	8.82 (−14%)	0.48 (−5%)	5.83 (−43%)	0.31 (−40%)
Belle Int'l	8.64	0.69	7.52 (−13%)	0.61 (−11%)	5.61 (−35%)	0.45 (−34%)
Bank of E Asia	8.46	0.44	9.11 (+8%)	0.52 (+18%)	5.55 (−34%)	0.30 (−34%)
China Res Power	11.87	0.46	13.35 (+12%)	0.67 (+46%)	7.06 (−41%)	0.29 (−37%)
MTR Corp.	8.88	0.50	8.96 (+1%)	0.52 (+4%)	6.19 (−30%)	0.34 (−32%)
Tingyi	9.75	0.46	10.82 (+11%)	0.74 (+59%)	5.59 (−43%)	0.31 (−34%)
Kunlun Energy	8.45	0.57	9.59 (+13%)	0.67 (+18%)	4.55 (−46%)	0.31 (−46%)
Sino Land	9.39	0.50	9.25 (−1%)	0.50 (−0%)	5.89 (−37%)	0.31 (−39%)
China Res Land	10.47	0.55	8.29 (−21%)	0.49 (−10%)	6.02 (−43%)	0.33 (−39%)
China Mer Hldgs	13.48	0.49	12.99 (−4%)	0.52 (+7%)	8.07 (−40%)	0.27 (−44%)
China Resources	9.74	0.52	11.01 (+13%)	0.75 (+45%)	5.59 (−43%)	0.32 (−37%)
CITIC Pacific	8.45	0.51	8.16 (−3%)	0.50 (−2%)	5.19 (−39%)	0.32 (−39%)
COSCO Pacific	11.13	0.46	9.66 (−13%)	0.46 (+1%)	5.53 (−50%)	0.26 (−43%)
Cathay Pacific	8.99	0.55	7.24 (−19%)	0.48 (−13%)	4.89 (−46%)	0.33 (−41%)
Hang Seng	**7.09**	**0.58**	**6.56 (−7%)**	**0.55 (−6%)**	**4.19 (−41%)**	**0.35 (−40%)**

Table 5: Standard deviations between median and strategies based on *VWAP* and *Optimal weights with and without execution risk*. The numbers are in basis points (*bps*) of the starting prices at each day and in terms of prevailing *spread*. The numbers in parentheses show the changes compared to the corresponding values of the VWAP strategy.

B Comparison with Deterministic Strategies

Algorithm 1 provides an optimal stochastic strategy, which depends on taking successively conditional expectations. In this appendix, we derive the optimal deterministic strategy as comparison and show how much our optimal stochastic strategy improves the performance compared to the deterministic strategy.

We still consider the minimisation problem (2), but now restrict to deterministic weights $w_1, \dots, w_N \in \mathbb{R}$ while the weights in the previous section were stochastic. Using $w_N = 1 - w_1 - \cdots - w_{N-1}$, we can rewrite the objective function in (2) as

$$\begin{aligned} f(w_1, \dots, w_{N-1}) = \mathbb{E}\Bigg[\bigg(\sum_{i=1}^{N-1} w_i(P_i^r - P_N^r) + P_N^r - \operatorname*{median}_{j=1,\dots,N}(P_j)\bigg)^2\Bigg] \\ + \frac{1}{\lambda}\mathbb{E}\Bigg[\sum_{i=1}^{N} w_i(P_i^r - P_N^r) + P_N^r - \operatorname*{median}_{j=1,\dots,N}(P_j)\Bigg]. \end{aligned}$$

The first-order condition for the optimal weights is

$$\frac{\partial f}{\partial w_k} = 2\,\mathbb{E}\Bigg[(P_k^r - P_N^r)\bigg(\sum_{i=1}^{N-1} w_i(P_i^r - P_N^r) + P_N^r - \operatorname*{median}_{j=1,\dots,N}(P_j) + \frac{1}{2\lambda}\bigg)\Bigg] = 0$$

for $k = 1, \dots, N-1$. This can be written as $Aw = b$, where we define a matrix $A = (A_{k,i})_{1 \le k,i \le N-1}$ and a vector $b = (b_1, \dots, b_{N-1})^\top$ by

$$\begin{aligned} A_{k,i} &= \mathbb{E}[(P_k^r - P_N^r)(P_i^r - P_N^r)], \\ b_k &= \mathbb{E}\Big[(P_k^r - P_N^r)\Big(\operatorname*{median}_{j=1,\dots,N}(P_j) - P_N^r\Big) - \frac{1}{2\lambda}(P_k^r - P_N^r)\Big]. \end{aligned}$$

A and b correspond to static analogues to Q and R, which were dynamically defined in Algorithm 1. We assume that

$$\begin{aligned} &\text{there does } \textit{not} \text{ exist } k \in \{1, \dots, N\} \text{ and } (\lambda_i)_{i \ne k} \in \mathbb{R}^{N-1} \\ &\text{with } \sum_{i \ne k} \lambda_i = 1 \text{ and } P_k^r = \sum_{i \ne k} \lambda_i P_i^r \text{ a.s.,} \end{aligned} \tag{10}$$

which is the static analogue to (6). This assumption is very natural because otherwise, we would know some of the prices in advance, which is not realistic. Under assumption (10), the matrix A is invertible as we will show next so that the optimal weights are given by

$$(w_1, \dots, w_{N-1})^\top = A^{-1}b \quad \text{and} \quad w_N = 1 - \sum_{i=1}^{N-1} w_i. \tag{11}$$

To show that A is invertible, we first note that A is symmetric and positive semidefinite because for every $x \in \mathbb{R}^{N-1}$, we have

$$x^\top A x = \mathbb{E}\Bigg[\sum_{k,i=1}^{N-1} x_k(P_k^r - P_N^r)(P_i^r - P_N^r)x_i\Bigg] = \mathbb{E}\Bigg[\bigg(\sum_{i=1}^{N-1} x_i(P_i^r - P_N^r)\bigg)^2\Bigg] \ge 0.$$

Therefore, A is invertible if it is positive definite. This holds if there does *not* exist $x \in \mathbb{R}^{N-1} \setminus \{0\}$ such that

$$\sum_{i=1}^{N-1} x_i(P_i^r - P_N^r) = 0 \quad \text{a.s.} \tag{12}$$

If $\sum_{i=1}^{N-1} x_i \neq 0$, (12) can be written as $P_N^r = \sum_{i=1}^{N-1} \lambda_i P_i^r$ a.s. with $\lambda_i = x_i / \sum_{k=1}^{N-1} x_k$, which is in contraction to (10). If $\sum_{i=1}^{N-1} x_i = 0$, (12) simplifies to $\sum_{i=1}^{N-1} x_i P_i^r = 0$ a.s. Since $x \neq 0$, there exists $x_k \neq 0$, and we set $\lambda_i = -x_i/x_k = x_i / \sum_{j \neq k} x_j$ for $i \neq N$ and $\lambda_N = 0$ so that we have $P_k^r = \sum_{i \neq k} \lambda_i P_i^r$ a.s., again in contraction to (10).

Comparing the optimal static solution given by (11) with the optimal dynamic solution from Algorithm 1, we see that the optimal static solution corresponds to a one-step expectation while the optimal dynamic solution is based on taking successively conditional expectations. To analyse in more detail this difference of the optimal strategies and the different values of the objective function, we now restrict our attention to the case where θ_i are constant and $\mathbb{E}[Z_i] = 0$. Then, we can simplify A and b to

$$A_{k,i} = \sigma^2(N - \max\{k,i\}), \quad b_k = \sigma^2(N-k) - \sum_{i=k+1}^{N} \mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\ldots,N}(P_j)\Big].$$

where we set $\sigma = \sqrt{\mathbb{E}[Z_i^2]}$. We can explicitly calculate the inverse of A, which is given by

$$A_{k,i}^{-1} = \begin{cases} 1/\sigma^2 & \text{if } k = i = 1 \\ 2/\sigma^2 & \text{if } k = i > 1 \\ -1/\sigma^2 & \text{if } k = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

so that the optimal weights equal

$$w = A^{-1}b = \begin{pmatrix} 1 - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_2 \operatorname*{median}_{j=1,\ldots,N}(P_j)\Big] \\ \frac{1}{\sigma^2}\mathbb{E}\Big[(Z_2 - Z_3)\operatorname*{median}_{j=1,\ldots,N}(P_j)\Big] \\ \cdots \\ \frac{1}{\sigma^2}\mathbb{E}\Big[(Z_{N-1} - Z_N)\operatorname*{median}_{j=1,\ldots,N}(P_j)\Big] \end{pmatrix},$$

$$w_N = 1 - w_1 - \cdots - w_{N-1} = \frac{1}{\sigma^2}\mathbb{E}\Big[Z_N \operatorname*{median}_{j=1,\ldots,N}(P_j)\Big].$$

Straightforward calculations yield

$$\mathbb{E}\left[\sum_{i=1}^{N} w_i P_i^r\right] = 0, \tag{13}$$

$$\mathbb{E}\left[\left(\sum_{i=1}^{N} w_i P_i^r - \operatorname*{median}_{j=1,\ldots,N}(P_j)\right)^2\right] = \mathbb{E}\left[\left(\operatorname*{median}_{j=1,\ldots,N}(P_j)\right)^2\right] - \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\ldots,N}(P_j)\Big]\right)^2.$$

As comparison, we prove the following result for the optimal stochastic weights.

Proposition B.1. *Assume* $\mathbb{E}[Z_i] = 0$ *and that* $\theta_1 = \cdots = \theta_N$ *are constant, and set* $\sigma = \sqrt{\mathbb{E}[Z_i^2]}$. *Then the optimal stochastic weights from Algorithm 1 are explicitly given by*

$$
\begin{aligned}
w_1 &= 1 - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_2 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big],\\
w_2 &= \frac{1}{\sigma^2}\mathbb{E}\Big[Z_2 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_3 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_1\Big],\\
&\dots\\
w_{N-1} &= \frac{1}{\sigma^2}\mathbb{E}\Big[Z_{N-1} \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-3}\Big] - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_N \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-2}\Big],\\
w_N &= \frac{1}{\sigma^2}\mathbb{E}\Big[Z_N \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-2}\Big]. \qquad (14)
\end{aligned}
$$

For the optimal stochastic weights w_j, *we have*

$$
\mathbb{E}\Big[\sum_{i=1}^N w_i P_i^r\Big] = 0, \qquad (15)
$$

$$
\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{i=1}^N w_i P_i^r - \operatorname*{median}_{j=1,\dots,N}(P_j)\Big)^2\Big] &= \mathbb{E}\Big[\Big(\operatorname*{median}_{j=1,\dots,N}(P_j)\Big)^2\Big] \qquad (16)\\
&\quad - \frac{1}{\sigma^2}\sum_{i=1}^N \mathbb{E}\Big[\Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big)^2\Big],
\end{aligned}
$$

where we set $\mathcal{F}_{-1} = \mathcal{F}_0$.

Comparing (13) with (15), we see that the optimal deterministic and stochastic strategies lead to the same mean, but the squared error to the median of the stochastic strategy compared to that of the deterministic strategy is reduced by

$$
\begin{aligned}
&\frac{1}{\sigma^2}\sum_{i=1}^N \mathbb{E}\Big[\Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big)^2\Big] - \frac{1}{\sigma^2}\sum_{i=1}^N \Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big]\Big)^2\\
&= \frac{1}{\sigma^2}\sum_{i=1}^N \operatorname{Var}\Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big).
\end{aligned}
$$

Note that for normally distributed Z_i, this depends only on N and not on σ. A numerical calculation shows that for $N = 5$, the relative improvement in standard deviation is approximately 23 % from using the stochastic strategy compared to the deterministic strategy.

Proof of Proposition B.1. We first show that $Q_{k,j}$ and $A_{k,j}$ in Algorithm 1 are given by

$$
Q_{k,j} = P_j^r - P_{k+1}^r \text{ and } A_{k,j} = 1 \text{ for } k = 1, \dots, N-1 \text{ and } j \le k. \qquad (17)
$$

We prove this claim by backward induction over $k = N-1, \dots, 1$. For $k = N-1$, we have $Q_{N-1,j} = Q_j - Q_N$ by definition and

$$
\begin{aligned}
A_{N-1,j} &= \frac{\mathbb{E}[Q_{N-1,N-1}Q_{N-1,j}|\mathcal{F}_{N-2}]}{\mathbb{E}[Q_{N-1,N-1}^2|\mathcal{F}_{N-2}]} = \frac{\mathbb{E}[(P_{N-1}^r - P_N^r)(P_j^r - P_N^r)|\mathcal{F}_{N-2}]}{\mathbb{E}[(P_{N-1}^r - P_N^r)^2|\mathcal{F}_{N-2}]}\\
&= \frac{\mathbb{E}[Z_N \sum_{i=j+1}^N Z_i|\mathcal{F}_{N-2}]}{\mathbb{E}[Z_N^2|\mathcal{F}_{N-2}]} = \frac{\sigma^2}{\sigma^2} = 1 \quad \text{for } j \le N-1.
\end{aligned}
$$

Assume now that $Q_{k+1,j} = P_j^r - P_{k+2}^r$ and $A_{k+1,j} = 1$. Then we have

$$\begin{aligned}
Q_{k,j} &= Q_{k+1,j} - Q_{k+1,k+1}A_{k+1,j} = P_j^r - P_{k+1}^r, \\
A_{k,j} &= \frac{\mathbb{E}[Q_{k,k}Q_{k,j}|\mathcal{F}_{k-1}]}{\mathbb{E}[Q_{k,k}^2|\mathcal{F}_{k-1}]} = \frac{\mathbb{E}[(P_k^r - P_{k+1}^r)(P_j^r - P_{k+1}^r)|\mathcal{F}_{k-1}]}{\mathbb{E}[(P_k^r - P_{k+1}^r)^2|\mathcal{F}_{k-1}]} \\
&= \frac{\mathbb{E}[Z_{k+1}\sum_{i=j+1}^{k+1} Z_i|\mathcal{F}_{N-2}]}{\mathbb{E}[Z_{k+1}^2|\mathcal{F}_{k-1}]} = \frac{\sigma^2}{\sigma^2} = 1 \quad \text{for } j \le k.
\end{aligned}$$

This shows (17). To analyse $v_{N-1}, \dots, v_1$, we use $m_{N-1} = \operatorname*{median}_{j=1,\dots,N}(P_j) - Q_N$ and find

$$v_{N-1} = \frac{\mathbb{E}\big[Q_{N-1,N-1}\big(m_{N-1} - \frac{1}{2\lambda}\big)\big|\mathcal{F}_{N-2}\big]}{\mathbb{E}[Q_{N-1,N-1}^2|\mathcal{F}_{N-2}]} = 1 - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_N \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-2}\Big]$$

and then

$$\begin{aligned}
m_{N-2} &= m_{N-1} - Q_{N-1,N-1}v_{N-1} \\
&= \operatorname*{median}_{j=1,\dots,N}(P_j) - P_{N-1}^r - \frac{Z_N}{\sigma^2}\mathbb{E}\Big[Z_N \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-2}\Big], \\
v_{N-2} &= \frac{\mathbb{E}\big[Q_{N-2,N-2}\big(m_{N-2} - \frac{1}{2\lambda}\big)\big|\mathcal{F}_{N-3}\big]}{\mathbb{E}[Q_{N-2,N-2}^2|\mathcal{F}_{N-3}]} \\
&= 1 - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_{N-1} \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{N-3}\Big],
\end{aligned}$$

using the independence of Z_N, Z_{N-1} and $\mathcal{F}_{N-3}$. Continuing like this, we find

$$v_i = 1 - \frac{1}{\sigma^2}\mathbb{E}\Big[Z_{i+1} \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]$$

for all $i = N-1, \dots, 1$. We use $w_1 = v_1$,

$$w_j = v_j - \sum_{\ell=1}^{j-1} A_{j,\ell}w_\ell = v_j - \sum_{\ell=1}^{j-1} w_\ell = v_j - v_{j-1}$$

for $j = 2, \dots, N-1$ and $w_N = 1 - \sum_{\ell=1}^{N-1} w_\ell = 1 - v_{N-1}$ to deduce (14).

We derive (15) by the tower property of conditional expectation, which yields

$$\begin{aligned}
\mathbb{E}\bigg[\sum_{i=1}^N w_iQ_i\bigg] &= \mathbb{E}\bigg[\sum_{j=1}^N Q_{j,j}v_j\bigg] = \mathbb{E}\bigg[\sum_{j=1}^N \mathbb{E}[-Z_{j+1}v_j|\mathcal{F}_j]\bigg] \\
&= \mathbb{E}\bigg[\sum_{j=1}^N \mathbb{E}[-Z_{j+1}|\mathcal{F}_j]v_j\bigg] = 0.
\end{aligned}$$

For (16), we first deduce

$$\mathbb{E}\bigg[\bigg(\sum_{i=1}^N w_iQ_i - \operatorname*{median}_{j=1,\dots,N}(P_j)\bigg)^2\bigg] = \mathbb{E}\big[(m_1 - Q_{1,1}v_1)^2\big]$$

and then simplify $m_1 - Q_{1,1}v_1$ to

$$
\begin{aligned}
m_1 - Q_{1,1}v_1 &= m_2 - Q_{2,2}v_2 + Z_2 - \frac{Z_2}{\sigma^2}\mathbb{E}\Big[Z_2 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] \\
&= m_3 - Q_{3,3}v_3 + Z_3 + Z_2 - \frac{Z_3}{\sigma^2}\mathbb{E}\Big[Z_3 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_1\Big] - \frac{Z_2}{\sigma^2}\mathbb{E}\Big[Z_2 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] \\
&= \dots \\
&= m_N - Q_{N,N}v_N + \sum_{i=2}^{N} Z_i - \frac{1}{\sigma^2}\sum_{i=2}^{N} Z_i \mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big] \\
&= \operatorname*{median}_{j=1,\dots,N}(P_j) - Z_1 - \frac{1}{\sigma^2}\sum_{i=2}^{N} Z_i \mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big].
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
\mathbb{E}\big[(m_1 - Q_{1,1}v_1)^2\big] &= \mathbb{E}\Big[\Big(\operatorname*{median}_{j=1,\dots,N}(P_j)\Big)^2\Big] - 2\mathbb{E}\Big[Z_1 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] + \sigma^2 \\
&\quad - \frac{2}{\sigma^2}\mathbb{E}\Big[\sum_{i=2}^{N} Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big] \\
&\quad + \frac{2}{\sigma^2}\mathbb{E}\Big[\sum_{i=2}^{N} Z_i Z_1 \mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big] \\
&\quad + \frac{1}{\sigma^2}\mathbb{E}\Big[\sum_{i=2}^{N} \Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\operatorname*{median}_{j=1,\dots,N}(P_j)\Big)^2\Big] - 2\mathbb{E}\Big[Z_1 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] + \sigma^2 \\
&\quad - \frac{1}{\sigma^2}\mathbb{E}\Big[\sum_{i=2}^{N} \Big(\mathbb{E}\Big[Z_i \operatorname*{median}_{j=1,\dots,N}(P_j)\Big|\mathcal{F}_{i-1}\Big]\Big)^2\Big],
\end{aligned}
$$

using the tower property of conditional expectation, the independence of Z_i from $\mathcal{F}_{i-1}$, and $\mathbb{E}[Z_i] = 0$. This implies (16) using

$$
\begin{aligned}
\mathbb{E}\Big[Z_1 \operatorname*{median}_{j=1,\dots,N}(P_j)\Big] &= \sigma^2 + \mathbb{E}\Big[Z_1 \operatorname*{median}_{j=1,\dots,N}(P_j - Z_1)\Big] \\
&= \sigma^2 + \mathbb{E}[Z_1]\mathbb{E}\Big[Z_1 \operatorname*{median}_{j=1,\dots,N}(P_j - Z_1)\Big] = \sigma^2. \qquad \square
\end{aligned}
$$

Acknowledgements

We would like to thank Yaokun Lin for assisting us in collecting data from Bloomberg and two anonymous referees for valuable comments, which enabled us to significantly improve the paper. Christoph Frei gratefully acknowledges financial support by the Natural Sciences and Engineering Research Council of Canada under grant RGPIN/402585-2011.

References

[1] A. Agarwal, P.L. Bartlett and M. Dama (2010): Optimal allocation strategies for the dark pool problem. In: Y.W. Teh and M. Titterington (Eds.): Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10), 9–16

[2] R. Almgren (2012): Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics* **3**, 163–181

[3] R. Almgren and N. Chriss (2000): Optimal execution of portfolio transactions. *Journal of Risk* **3**, 5–39

[4] J. Bacidore, B. Polidore, W. Xu and C. Yang (2013): Trading around the close. *Journal of Trading* **8**, 48–57

[5] M. Barclay, T. Hendershott and C. Jones (2008): Order consolidation, price efficiency, and extreme liquidity shocks. *Journal of Financial and Quantitative Analysis* **43**, 93–122

[6] S. Berkowitz, D. Logue and E. Noser (1988): The total cost of transactions on the NYSE, *Journal of Finance* **43**, 97–112

[7] J.-P. Bouchaud, Y. Gefen, M. Protters and M. Wyart (2004): Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance* **4**, 176–190

[8] S. Boyd and L. Vandenberghe (2004): *Convex Optimization*, Cambridge University Press

[9] E. Busseti and S. Boyd (2015): Volume weighted average price optimal execution. Preprint available at http://arxiv.org/abs/1509.08503

[10] Á. Cartea and S. Jaimungal (2016): A closed-form execution strategy to target volume weighted average price. *SIAM Journal on Financial Mathematics* **7**, 760–785

[11] Á. Cartea, S. Jaimungal and J. Penalva (2015): Algorithmic and High-Frequency Trading, Cambridge University Press

[12] T. Chordia, R. Roll and A. Subrahmanyam (2002): Order imbalance, liquidity, and market returns. *Journal of Financial Economics* **65**, 111–130

[13] C. Comerton-Forde, S.T. Lau and T. McInish (2007): Opening and closing behavior following the introduction of call auctions in Singapore. *Pacific-Basin Finance Journal* **15**, 18–35

[14] D. Dor and U. Zwick (1995): Selecting the median, *SIAM Journal on Computing* **28**, 1722–1758

[15] C. Frei and N. Westray (2015): Optimal execution of a VWAP order: a stochastic control approach. *Mathematical Finance* **25**, 612–639

[16] G. Frison and J. Jørgensen (2013): Efficient implementation of the Riccati recursion for solving linear-quadratic control problems. Proceedings of the IEEE Multiconference on Systems and Control (MSC), Hyderabad, India, August 28–30, 2013

[17] O. Guéant (2014): Execution and block trade pricing with optimal constant rate of participation. *Journal of Mathematical Finance* **4**, 255–264

[18] O. Guéant (2016): The Financial Mathematics of Market Liquidity: From Optimal Execution to Market Making, Chapman and Hall/CRC

[19] O. Guéant and G. Royer (2014): VWAP execution and guaranteed VWAP. *SIAM Journal on Financial Mathematics* **5**, 445–471

[20] E. Himaras (2012): HK bourse closing auction needed, Deutsche Bank poll says. *Bloomberg*. Available at www.bloomberg.com/news/ articles/2012-04-22/closing-auction-needed-on-hong-kong-exchange-deutsche-bank-says

[21] Hong Kong Exchanges and Clearing Limited (2007): Consultation paper: the introduction of a closing auction session. Available at www.hkex.com.hk/consul/paper/c_auction_e.pdf

[22] Hong Kong Exchanges and Clearing Limited (2015): Consultation paper: proposal for introduction of volatility control mechanism in the securities and derivatives markets and closing auction session in the securities market. Available at www.hkex.com.hk/eng/newsconsul/mktconsul/Documents/cp201501.pdf

[23] Hong Kong Exchanges and Clearing Limited (2015): Rules of the exchange. Available at www.hkex.com.hk/eng/rulesreg/traderules/sehk/exrule.htm

[24] M. Labadie and C.-A. Lehalle (2014): Optimal starting times, stopping times and risk measures for algorithmic trading. *Journal of Investment Strategies* **3**, Issue 2, 3–33

[25] S. Laruelle, C.-A. Lehalle, G. Pagès (2011): Optimal split of orders across liquidity pools: a stochastic algorithm approach. *SIAM Journal on Financial Mathematics* **2**, 1042–1076

[26] S. Laruelle, C.-A. Lehalle, G. Pagès (2013): Optimal posting price of limit orders: learning by trading. *Mathematics and Financial Economics* **7**, 359–403

[27] G. Lee (2012): Closing price uncertainty hits traders on Hong Kong Stock Exchange. *Asia Risk.* Available at www.risk.net/asia-risk/news/2152340/closing-price-uncertainty-hits-traders-hong-kong-stock-exchange

[28] C.-A. Lehalle and S. Laruelle (2013): Market Microstructure in Practice, World Scientific Publishing

[29] H. Lockett (2016): China's State Council approves Shenzhen-HK market linkup. *Financial Times.* Available at www.ft.com/content/f24ed64c-639b-11e6-8310-ecf0bddad227

[30] G.S. Manku, S. Rajagopalan and B.G. Lindsay (1998): Approximate medians and other quantiles in one pass and with limited memory. In: A. Tiwary and M. Franklin (Eds.): SIGMOD '98 Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, 426–435

[31] C. Moallemi and M. Sağlam (2013): The cost of latency in high-frequency trading. *Operations Research* **61**, 1070–1086

[32] M. Pagano and R. Schwartz (2003): A closing call's impact on market quality at Euronext Paris. *Journal of Financial Economics* **68**, 439–484