Contracting for Financial Execution*

Markus Baldauf[†] Christoph Frei[‡] Joshua Mollner[§]

This Version: July 10, 2018

Abstract

Financial contracts often specify reference prices whose values are undetermined at the time of contracting, which introduces the possibility that one side manipulates them to the detriment of the other. Motivated by this, we study a stylized model of financial contracting between a client, who wishes to trade a large position, and her broker. Depending on the contract offered, the broker may have an incentive to schedule the client's trades so as to manipulate the reference prices specified in the contract. We find that a simple contract based on the volume-weighted average price (VWAP) emerges as uniquely optimal solution to a principal-agent problem between the client and her broker. This result explains the popularity of guaranteed VWAP contracts in equity trading and suggests a direction for improvement upon the status quo in other asset classes.

Keywords: agency conflict, benchmark manipulation, broker-client relationship, foreign exchange fix, front-running, pre-trade hedging, volume-weighted average price, VWAP

---

*We thank Robert Almgren, Piotr Dworczak, Pete Kyle, Andreas Park (discussant), and Nicholas Westray for helpful comments. We thank Colis Cheng and Kateryna Melnykova for excellent research assistance. Christoph Frei gratefully acknowledges support by the Discovery Grant program of the Natural Sciences and Engineering Research Council of Canada.

[†]University of British Columbia, Sauder School of Business, 2053 Main Mall, Vancouver, BC, Canada V6T 1Z2: baldauf@mail.ubc.ca, https://sites.google.com/site/mbaldauf/.

[‡]University of Alberta, 621 Central Academic Building, Edmonton, AB, Canada T6G 2G1: cfrei@ualberta.ca, https://www.math.ualberta.ca/∼cfrei/.

[§]Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208: joshua.mollner@kellogg.northwestern.edu, https://sites.google.com/site/joshuamollner/.

1 Introduction

Many financial contracts reference benchmark prices whose values are yet to be determined at the time of contracting. For example, a client might agree, in advance, to trade one billion US dollars with her broker in exchange for British pounds at the 'fix', a popular end of day benchmark. Similar examples are available across a range of asset classes. A potential advantage of such arrangements (over those with predetermined prices) is that they may enable the broker to limit his exposure to price fluctuations, which could be beneficial if the broker is more risk averse or capital constrained than the client.

Nevertheless, such benchmarks are typically endogenous. Consequently, a potential disadvantage of such arrangements is that the broker may trade in a way that influences these benchmarks to his advantage, but to the detriment of the client. Financial markets abound with examples of brokers taking advantage of their clients in this way. For instance, HSBC recently agreed to pay more than 100 million dollars to enter into a deferred prosecution agreement in connection with manipulating the British pound exchange rate to the detriment of a client (and the benefit of the bank and some of its traders) on two separate occasions (DOJ, 2018). Moreover, there have been multiple incidents of brokers attempting to manipulate the World Markets/Reuters Closing Spot Rates (WM/R), which are common benchmark rates for foreign currency transactions (Bloomberg, 2013; CFTC, 2014; WSJ, 2014). The publicized cases are likely only the tip of the iceberg.

In this paper we use insights from a principal-agent problem to derive a contract that a client might offer to a broker that would perform well in spite of both aforementioned frictions: (*i*) risk aversion on the part of the broker and (*ii*) the possibility of benchmark manipulation. Analyzing a model in which a client with a trading need (the principal) hires a broker (the agent), we find a unique optimal contract. Moreover, the optimum corresponds to what is known in practice as *guaranteed VWAP:* the agreement that, at a predetermined future time, the broker and client will trade a predetermined quantity at the volume-weighted average price (VWAP) over the intervening time interval.

In the model, the client offers a contract to the broker at time 0. The contract specifies that the client and broker will conduct an over-the-counter trade at time T, in which the client will purchase one share from the broker at a price that may depend upon the history of market volumes and prices. If the broker accepts the contract, then he purchases the share on the market and may divide his trades across the intervening trading periods. Trading creates short-term price impact. The severity of price impact in a particular time period is related to the volume traded in that period, in that both price and volume are influenced by market conditions—such as the participation of other traders—about which the broker possesses superior information. Because market conditions may differ across periods, the broker's division of the client's order across time determines the expected cost of trading. Depending on the form of the contract, the broker's choice in this matter may be distorted so that he fails to schedule trades in the efficient way. In the model, prices are also affected by exogenous shocks, which might represent supply and demand imbalances among outside traders or the effect of news about economic fundamentals. The broker is risk averse, and therefore will bear risk originating from these price shocks only if the contract provides him with sufficient compensation for doing so.

We allow for a large class of possible contracts, and it is instructive to consider a few familiar cases in more detail. First, consider a fixed price contract, where the broker pays the client an exogenous price, e.g. yesterday's close. The specified price cannot be manipulated by the broker. However, as alluded to in the opening paragraph, risk aversion (or capital constraints) of the broker imply that he would need to be compensated for assuming the price risk of the position. Second, consider a contract that stipulates a transaction price equal to the prevailing price at period T. This resembles the above foreign exchange example involving HSBC. In the model, this contract is not optimal because the broker has an incentive to schedule trades in the inefficient way (in particular, he may wish to schedule trades so as to create a gap between the price in period T and the average price that he pays in acquiring the position).

In contrast, we establish in the text that the optimal contract references the VWAP benchmark. Unlike fixed price contracts, the VWAP contract insures the broker against price fluctuations, provided that he trades in the efficient way. And unlike contracts that are tied to one particular future price, the VWAP contract does not incentivize the broker to schedule trades in the inefficient way. This helps explain the popularity of such contracts in practice.[1]

The contracts described in the previous paragraphs are examples of "principal trading" arrangements, in which the broker is the counterparty of the agent. Such arrangements are our primary focus, and our main result is that the VWAP contract is uniquely optimal in this class. However, in a later section, we also consider "agency trading" arrangements, in

[1]One example from the domain of equity trading is Nomura Research Institute, Ltd. (2014).

which the broker trades on the client's behalf, acting as a matchmaker between the client and a third party. We show that such arrangements can also be optimal in the model, and we discuss how a variety of unmodeled forces might make agency trading either more or less attractive than the VWAP contract. The particular agency trading arrangement that is optimal is the one in which the client requests the broker to schedule trades in the efficient way. In certain cases of the model, this equates to using what is known in practice as a VWAP algorithm, which slices orders to match the volume profile of the market. The model therefore provides a rationale for the empirical popularity of "agency VWAP" arrangements.

The remainder of this paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the baseline model. In this baseline specification, we model the nature of price impact in a reduced form way, allowing for a large class of functional forms; however, in Section 4, we provide a micro-foundation for one particular member of this class. Section 5 states the main results on the optimality of VWAP contracts. We then turn to a discussion of the extent to which these results can be generalized. In Section 6, we provide some affirmative results of this nature, arguing that the VWAP contract can remain optimal even when the broker has access to a richer set of deviations (Section 6.1) or when agency trading is possible (Section 6.2). However, in Section 7, we provide some negative results, which highlight the conditions that are required for optimality of the VWAP contract. For instance, our baseline results apply to settings in which price impact is temporary, and we demonstrate that they do not extend to settings in which trading has a permanent impact on prices. Section 8 discusses our findings and concludes. All proofs are relegated to the Appendix.

2 Related literature

Previous literature has also studied conflicts of interest between brokers and clients. One type of conflict is created by so-called dual trading (e.g., Röell, 1990; Fishman and Longstaff, 1992; Bernhardt and Taub, 2008) in which a broker may engage in proprietary trading alongside the trades that he makes on behalf of his clients. In contrast, we analyze a setting in which suboptimal execution may occur even absent proprietary trading activity. In addition to the publicized cases mentioned in the introduction, a number of academic studies have demonstrated empirically that the trading decisions of brokers are in fact sometimes distorted in this way. Closely related to our analysis, Henderson, Pearson and Wang (2018) find evidence of distortions similar to those that we model, doing so in the context of structured equity products. In a similar spirit, Battalio, Corwin and Jennings (2016) find evidence of brokers routing client orders suboptimally in order to maximize rebates from exchanges. And Hillion and Suominen (2004) find evidence of brokers manipulating closing prices so as to enhance their reputation for execution quality.

Although less directly related to our analysis, conflicts also arise within the non-trading component of broker-client relationships. For example, analysts may have an incentive to issue biased recommendations so as to boost brokerage revenues (e.g., Michaely and Womack, 1999; Ljungqvist, Marston, Starks, Wei and Yan, 2007; Malmendier and Shanthikumar, 2014).

The choice of a benchmark price is important in our model because of its effect upon the trading incentives of the broker. For similar reasons, benchmark choice plays a major role in many other aspects of financial markets. Benchmarks for interest rates constitute one example. Banks may be incentivized to move these rates in a particular direction, and a benchmark administrator may wish to select a benchmark that is less prone to manipulation of this sort (Duffie and Dworczak, 2014; Coulter, Shapiro and Zimmerman, 2017). Despite many differences between this setting and that of our paper, these papers find the optimal benchmark to be some weighted average price, which in some cases resembles VWAP. Benchmarks for assessing the quality of fund managers constitute another example. Fund managers are incentivized to distort their trading decisions so as to perform well under the chosen metric, so that wide use of a manipulation-proof performance measure could be beneficial (Goetzmann, Ingersoll, Spiegel and Welch, 2007).

Since the contract that emerges from our framework as optimal, the VWAP contract, is a fairly simple one, this paper relates to a literature on foundations for contracts possessing simple features such as linearity. Holmström and Milgrom (1987) show that when an agent repeatedly performs the same task, the principal optimally provides the same incentives at each point in time, which makes the agent's payment a linear function of output. Carroll (2015) shows that linear contracts are optimal if the principal is uncertain about the actions available to the agent and treats this uncertainty with a worst-case criterion.

Although our focus is on VWAP *contracts*, there is a connection to VWAP *trading strategies*, which aim to acquire a position at or near the VWAP price. Some existing literature has focused on how to devise such strategies (e.g. Humphery-Jenner, 2011; Frei and Westray, 2015; Cartea and Jaimungal, 2016). Perhaps closest to this paper is Kato (2015) who provides conditions under which VWAP emerges as the optimal trading strategy. VWAP trading is the first-best trading policy in certain cases of our model for similar reasons.

3 Model

There are two individuals in the model: a client (the principal) and a broker (the agent). There are finitely many discrete trading periods $t \in \{1, 2, \ldots, T\}$.[2] And there is a single security. The client offers the broker a contract regarding the purchase a fixed number of shares. If the broker accepts the contract offered to him, then he purchases from the market the shares that he will then sell to the client. Importantly, the price that he obtains from the client might, depending on the terms of the contract, be subject to influence by his trading activity.

Trading. The client needs to buy a number of shares of the security, which we normalize to one. To do so, she contracts to purchase the share from the broker after time T. In advance of this transaction, the broker must purchase the required share in the market. Letting x_t denote the number of shares purchased by the broker in trading period t, we therefore require

$$\sum_{t=1}^{T} x_t = 1.$$

We also require that all x_t be nonnegative, so that the broker does not sell in any period. We refer to a vector $(x_t)_{t=1}^{T}$ that satisfies both of these conditions as a *trading schedule*.

In addition, there are time-varying market conditions, denoted by n_t, which both affect total trading volume and moderate the price impact of the broker's trades. The market condition n_t is a random variable that takes positive values. Its realization is known by the broker, but the client knows only its distribution. We do not impose any assumptions on the joint distribution of $n_1, \ldots, n_T$. In particular, $n_1, \ldots, n_T$ do not need to be independent. Nor do $n_1, \ldots, n_T$ need to be identically distributed; thus, our approach permits the client to have some information about market conditions, albeit less than the broker. The ways in which n_t affects volume and price are specified below.

[2] An alternative interpretation of the model is that in which trading is done by splitting orders not across time periods but across venues (or even across both time and venues). For these interpretations, t may be taken to index venues (or time-venue pairs).

Volume. The total volume (including the broker's trades) at time t is given by $v(x_t, n_t)$ for a positive, continuous function v, whose domain is part of $\mathbb{R}_+ \times \mathbb{R}_{++}$ and contains the range of (x_t, n_t), and which satisfies

$$x''n' > x'n'' \implies x''v(x', n') > x'v(x'', n'') \tag{1}$$

for (x', n') and (x'', n'') in the domain of v. An example of such a function v is $v(x_t, n_t) = x_t + n_t$, in which case n_t represents outside volume. Another example of v will be introduced in Section 4, where n_t denotes the number of outside traders, which influences both prices and traded volumes through a market clearing condition.

Price. The price that prevails in period t is

$$p_t = h\left(\frac{x_t}{n_t}\right) + \varepsilon_t,$$

where h is an increasing function such that $yh(y)$ is strictly convex; $\varepsilon_1, \ldots, \varepsilon_T$ are random variables that are independent from $n_1, \ldots, n_T$ and have all the same mean μ. However, we impose no further assumptions on $\varepsilon_1, \ldots, \varepsilon_T$, neither requiring them to be independent of each other nor identically distributed.

For a given number of shares x_t, the impact on price is influenced by the market conditions prevailing in the period. Therefore, the price impact depends on x_t measured relative to n_t, and not on x_t itself. In the case where n_t represents outside volume, using precisely the ratio x_t/n_t makes the price impact dimensionless (Almgren, Thum, Hauptmann and Li, 2005). A class of examples of price impact functions nested by our approach is $h(x_t/n_t) = (x_t/n_t)^a$ for $a > 0$. Such specifications are supported by both theoretical and empirical results in the literature, where the predominant configurations are between $a = 0.5$ (square root price impact) and $a = 1$ (linear price impact).[3]

In addition, we also use the notation $\boldsymbol{p} = (p_t)_{t=1}^T$, $\boldsymbol{n} = (n_t)_{t=1}^T$, $\boldsymbol{\varepsilon} = (\varepsilon_t)_{t=1}^T$, and $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{n}) = (v(x_t, n_t))_{t=1}^T$. Note, throughout, that $\boldsymbol{p}$ depends on $\boldsymbol{x}$, $\boldsymbol{n}$, and $\boldsymbol{\varepsilon}$.

[3]Indeed, linear price impact is consistent with the price impact models based on adverse selection by Kyle (1985) and Kyle, Obizhaeva and Wang (2018). Using a large data set on US equity, Almgren, Thum, Hauptmann and Li (2005) estimate an exponent a of 0.6 while Mastromatteo, Tóth and Bouchaud (2014) report exponents a in the range of 0.4–0.7 across different markets (equities, futures, and foreign exchange).

Contracts. In specifying the set of feasible contracts, we imagine that the client can observe the sequence of prices $\boldsymbol{p}$ and the sequence of total volumes $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{n})$ at time T. We allow the client to offer contracts that are arbitrary functions of these market outcomes. Formally, the set of contracts that we allow for consists of measurable functions $\tau : \mathbb{R}^T \times \mathbb{R}^T_{++} \to \mathbb{R}$, specifying that the client will pay $\tau(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{n}))$ to the broker at time T in exchange for one share of the security.

This formulation permits the client to propose a wide variety of trading arrangements including many familiar ones. Many contracts observed in practice specify that the client pay the broker according to a particular benchmark price. Common benchmarks include (*i*) the closing price, which corresponds to $\tau = p_T$, (*ii*) a predetermined fixed price, which corresponds to $\tau = \tau_0$ for some constant τ_0, (*iii*) the time weighted average price (TWAP), which corresponds to $\tau = \frac{1}{T}\sum_{t=1}^{T} p_t$, and (*iv*) the volume weighted average price, which corresponds to $\tau = \frac{\sum_{t=1}^{T} p_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)}$.

Our approach focuses on what are known in practice as principal trading arrangements, wherein the broker acts as the client's counterparty. In contrast, agency trading arrangements are those in which the broker merely acts as a matchmaker in locating a counterparty. Agency trading can be thought of within our framework as a contract specifying that the client reimburses the trading costs $\boldsymbol{p} \cdot \boldsymbol{x}$ in exchange for the share. The set of feasible contracts described above does not allow for this payment rule. However, in Section 6.2, we discuss how our results would be affected if such contracts were available.

Timing. The timing of events is as follows. Prior to trading, the client offers a contract τ to the broker. The broker either accepts or rejects the contract. If he rejects the contract, then he receives an outside option of 0. If he accepts the contract, then he learns $\boldsymbol{n}$ and chooses a trading schedule $\boldsymbol{x}$. Then $\boldsymbol{\varepsilon}$ is realized, and the scheduled trades $\boldsymbol{x}$ take place at the prices $\boldsymbol{p}$. At time T, the client pays the broker as specified by τ.

Note that the choice of x_t can depend only on $\boldsymbol{n}$ and not, for example, on the contemporaneous price shock ε_t. Thus, the broker should be thought of as conducting his trades using market orders. Nor may x_t depend upon $(\varepsilon_s)_{s=1}^{t-1}$. In the interpretation of the model in which t indexes venues (*cf.* footnote 2), this restriction would be the natural assumption. But for the interpretation in which t indexes time, this restriction may artificially constrain the actions available to the broker. However, in Section 6.1, we consider a version of the model in which the broker can modify the trading schedule dynamically, so that x_t can depend not only on $\boldsymbol{n}$ but also on information observed in previous periods.

The broker's payoffs. Since the broker may schedule trades in a way that depends on $\boldsymbol{n}$, he can be thought of as choosing a *trading policy*, which is a measurable function $\boldsymbol{x}(\cdot)$ that maps each $\boldsymbol{n}$ into a trading schedule. The broker's utility function over money is some function u, which is assumed to be strictly increasing and weakly concave. From accepting a contract τ and choosing a trading policy $\boldsymbol{x}(\cdot)$, the broker receives expected utility

$$\mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\big)\big].$$

The client's payoffs. We assume that if the broker rejects the offered contract, then the client receives a payoff of negative infinity. On the other hand, if the contract is accepted, then the client is risk neutral over monetary outcomes. In particular, if the broker accepts a contract τ and chooses a trading policy $\boldsymbol{x}(\cdot)$, then the client receives expected utility

$$-\mathbb{E}\big[\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right)\big].$$

4 Micro-foundation of the model

In the baseline model presented in Section 3, prices and volumes in a period t are jointly influenced by both the broker's trades x_t and market conditions n_t. Our approach encompasses a broad class of reduced-form dependencies. In this section, we complement that previous analysis by presenting a micro-foundation for one of these functional forms.

In the micro-foundation, the broker trades using market orders x_t, as before. In addition to the broker, there exists, in every period, a continuum of outside traders who receive liquidity shocks and trade with demand schedules. The quantity n_t, previously referred to as "market conditions," now denotes the measure of outside traders present in period t. The main appeal of this micro-foundation is that prices and volumes will be derived endogenously from a market-clearing condition. Moreover, as we demonstrate below, they will depend upon x_t and n_t in a way that is nested by our previous analysis.

4.1 Setup

In this micro-foundation, the client and the broker are precisely as before. What is different is that we more precisely specify the other traders in the market and the nature of trading.

Outside traders. In each period t, there is a measure n_t of outside traders. Consistent with our previous notation, the realization of $\boldsymbol{n}$ is known by the broker but not by the client, and we do not impose any assumptions on the joint distribution of $n_1, \ldots, n_T$. For an outside trader i arriving in period t, the utility received from acquiring y shares at the per-share price p_t is $u_i(y, p_t) = \theta_i y - \frac{1}{2}y^2 - p_t y$. The quantity θ_i is a random variable that represents the liquidity shock of trader i. In particular, if trader i arrives in period t, then $\theta_i = \eta_i + \varepsilon_t$, where the idiosyncratic component η_i is an independent draw from a standard normal distribution and where the common components $\varepsilon_1, \ldots, \varepsilon_T$ are random variables independent from $n_1, \ldots, n_T$ and all with the same mean, which is also consistent with our previous assumption.

Trading. In each period t, the broker submits a market order x_t, as before. In addition, each of the outside traders submits a demand schedule $y_i(p_t)$. The security is in zero net supply, and the price p_t is chosen to clear the market. All trades take place at that price.

4.2 Solution

Since there is a continuum of outside traders, each acts as a price-taker. Taking the first-order condition of u_i, we immediately obtain that outside trader i will submit the demand schedule

$$y_i(p_t) = \theta_i - p_t.$$

In period t, n_t such traders are active. Substituting for $\theta_i = \eta_i + \varepsilon_t$, the market-clearing condition becomes

$$x_t + n_t \int_{-\infty}^{\infty} (z + \varepsilon_t - p_t)\,\phi(z)\,dz = 0,$$

where $\phi(\cdot)$ denotes the standard normal probability density function. Likewise, $\Phi(\cdot)$ will denote the standard normal cumulative distribution function in what follows. Solving for price, the market-clearing condition becomes

$$p_t = \frac{x_t}{n_t} + \varepsilon_t.$$

We therefore obtain linear price impact for the broker's trades, which is indeed nested by our previous analysis (with $h(y) = y$).

Having submitted the demand schedule $y_i(p_t)$, the number of shares purchased by trader i at this market-clearing price will be

$$y_i\left(\frac{x_t}{n_t} + \varepsilon_t\right) = \theta_i - \frac{x_t}{n_t} - \varepsilon_t = \eta_i - \frac{x_t}{n_t}.$$

Thus, trader i will be a buyer only if $\eta_i > \frac{x_t}{n_t}$. In consequence, the number of shares bought in period t (and therefore also the total volume traded) will be

$$\begin{aligned} v(x_t, n_t) &= x_t + n_t \int_{\frac{x_t}{n_t}}^{\infty} \left(z - \frac{x_t}{n_t}\right) \phi(z)\, dz \\ &= x_t \Phi\left(\frac{x_t}{n_t}\right) + n_t \phi\left(\frac{x_t}{n_t}\right). \end{aligned}$$

The following proposition establishes that this function satisfies (1), so that the expression for volume is also nested by our previous analysis.

Proposition 1. *Letting $v(x, n) = x\Phi\left(\frac{x}{n}\right) + n\phi\left(\frac{x}{n}\right)$, it is the case that*

$$x''n' > x'n'' \implies x''v(x', n') > x'v(x'', n'')$$

for all $x', x'' \geq 0$ and all $n', n'' > 0$.

5 Main results

Our first main result, Theorem 4, states the VWAP contract is optimal. Our second main result, Theorem 5, states a sense in which the VWAP form is necessary for optimality under certain conditions. Before coming to these main results, we characterize the first-best trading policy in Lemma 2. Proofs are relegated to the Appendix.

5.1 Characterization of the first-best trading policy

We begin by defining and characterizing the first-best trading policy. Given that the client is risk neutral, a trading policy $\boldsymbol{x}(\cdot)$ is *first best* if, for all $\boldsymbol{n}$,

$$\boldsymbol{x}(\boldsymbol{n}) \in \arg\min_{\boldsymbol{x}} \mathbb{E}[\boldsymbol{p} \cdot \boldsymbol{x} | \boldsymbol{n}].$$

Thus, a first-best trading policy minimizes expected trading cost conditional on all realizations of the market conditions $\boldsymbol{n}$. Consequently, such a policy also minimizes the (unconditional) expected trading cost $\mathbb{E}[\boldsymbol{p} \cdot \boldsymbol{x}]$.

Given the structure of the model, there is a unique first-best trading policy, which Lemma 2 characterizes as the policy under which the broker's trades x_t are proportional to market conditions n_t. This trading policy is efficient because it equates the marginal cost of trading an extra unit across time periods. In the case where n_t represents outside volume, this is a VWAP trading policy (or what is known in practice as a "VWAP algorithm"), which slices the parent order so as to mimic the volume profile of the market. This is the first-best trading policy for the same reason that a VWAP policy is optimal in Kato (2015).

Lemma 2. *The first-best trading policy is*

$$\boldsymbol{x}^{FB}(\boldsymbol{n}) = \left(\frac{n_t}{\sum_{s=1}^{T} n_s} \right)_{t=1}^{T}.$$

The expected trading cost incurred by this policy is $\mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t} \right) \right]$.

5.2 The client's problem

Having derived the first-best trading policy, we now turn our attention to the second best. The client chooses a contract τ, as well as a "recommended trading policy" $\boldsymbol{x}(\cdot)$ to maximize her expected utility (equivalently, to minimize her expected payment) subject to individual rationality and incentive compatibility constraints of the broker:

$$\min_{\tau, \boldsymbol{x}(\cdot)} \mathbb{E}\big[\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right)\big] \quad \text{subject to}$$

$$\mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\big)\big] \geq u(0) \tag{IR}$$

$$\forall \hat{\boldsymbol{x}}(\cdot) : \mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\big)\big] \geq \mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\hat{\boldsymbol{x}}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \hat{\boldsymbol{x}}(\boldsymbol{n})\big)\big] \tag{IC}$$

As is the standard approach in contract theory, we presume that the principal can choose the agent's action so long as (IR) and (IC) are satisfied. In particular, if the broker is indifferent among several policies—that is, if (IC) holds with equality for a certain alternative trading policy—then this approach presumes that the broker will resolve his indifference in favor of the policy recommended by the client.

A contract τ is *optimal* if it is part of a solution to this problem. Following from Lemma 2, we can derive useful intermediate results about properties of optimal contracts, which are stated in Lemma 3.

Lemma 3. *The following conditions are together sufficient for a contract τ to be optimal:*

(i) for all $\hat{\boldsymbol{x}}(\cdot)$:

$$\mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n})\big)\big] \geq \mathbb{E}\big[u\big(\tau\left(\boldsymbol{p}, \boldsymbol{v}(\hat{\boldsymbol{x}}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \hat{\boldsymbol{x}}(\boldsymbol{n})\big)\big]$$

(ii) $\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right) = \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n})$ almost surely.

Moreover, if some such contract exists, and if u is strictly concave, then the conditions are also necessary.

For the first part of the result, notice that condition (i) is equivalent to $(\tau, \boldsymbol{x}^{FB}(\cdot))$ satisfying (IC). And condition (ii) implies both that the broker is fully insured and that (IR) is satisfied with equality. Thus, $(\tau, \boldsymbol{x}^{FB}(\cdot))$ implements the efficient outcome—both the efficient trading policy and efficient risk sharing—and leaves the broker with zero surplus. Clearly, no contract can do better than that. The second part of the result observes that if some contract satisfies those conditions, then *all* optimal contracts must implement the efficient outcome and leave the broker with zero surplus. To implement the efficient trading policy, condition (i) must hold. And, if the broker is risk averse, then to implement efficient risk sharing and leave the broker with zero surplus, condition (ii) must hold.

5.3 Optimality of VWAP

Building on Lemma 3, we now proceed to state our main results, both of which pertain to the VWAP contract

$$\tau^{VWAP} = \frac{\sum_{t=1}^{T} p_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)}.$$

The following two theorems concern the optimality of this contract. Theorem 4 states that the VWAP contract is optimal. Theorem 5 says that if the broker is risk averse and a certain full support condition is satisfied, then the VWAP contract is also the *unique* contract that is optimal.

Theorem 4. *The contract τ^{VWAP} is optimal.*

Theorem 5. *If u is strictly concave and the distributions of $\boldsymbol{\varepsilon}$ and $\boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})$ have full support over $\mathbb{R}^T$ and $\mathbb{R}^T_{++}$, respectively, then a contract τ is optimal only if $\tau = \tau^{VWAP}$ almost everywhere on its domain.*

To prove Theorem 4, we establish that τ^{VWAP} satisfies the conditions of Lemma 3. The result is then immediately implied by that lemma. For proving Theorem 5, we build on the conclusion that τ^{VWAP} satisfies the conditions of Lemma 3 to deduce that all optimal contracts must satisfy those same conditions. These conditions are demanding and severely restrict the possibilities for τ. With the full support assumptions, they in fact pin down τ to equal τ^{VWAP} almost everywhere.[4]

To see that uniqueness requires risk aversion, note that an appropriately specified fixed price (i.e. "sell the firm") contract would also be optimal under risk neutrality. Similarly, uniqueness also requires the full support assumptions, for otherwise there would be certain "irrelevant" regions of the domain of τ in which the contract could be altered without affecting optimality.

6 Extensions

In this section, we consider two modifications of the model. In Section 6.1, we enrich the set of deviations available to the broker by allowing him to condition his trading decisions on

[4] The full-support assumption could be abandoned if the uniqueness statement were weakened to $\tau(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})) = \tau^{VWAP}(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n}))$ almost surely, where $\boldsymbol{p}$ are the prices corresponding to the first-best trading policy $\boldsymbol{x}^{FB}(\boldsymbol{n})$.

the realizations of previous prices. Under an additional assumption on the volume function $v(x, n)$, we show that the VWAP contract remains optimal. In Section 6.2, we enrich the set of contracts available to the client to include the agency contract $\tau^A = \boldsymbol{p} \cdot \boldsymbol{x}$. The VWAP contract again remains optimal; however, it is no longer uniquely so: the agency contract is also optimal within the model when it is available.

6.1 Dynamic trading policies

In the baseline model, the trading policies available to the broker were functions that mapped market conditions into trading schedules. In this section, we permit a more general class of trading policies in which the broker may also condition his trading decisions on previous prices. For this section only, a *trading policy* is a measurable vector $(x_t(\cdot))_{t=1}^T$, such that (*i*) for all t, x_t depends on $\boldsymbol{n}$ and $(p_s)_{s=1}^{t-1}$, (*ii*) for all t, $x_t\big(\boldsymbol{n}, (p_s)_{s=1}^{t-1}\big) \geq 0$ almost surely, and (*iii*) $\sum_{t=1}^T x_t\big(\boldsymbol{n}, (p_s)_{s=1}^{t-1}\big) = 1$ almost surely. Note that, given the broker's knowledge of market conditions $\boldsymbol{n}$, dependence on the previous prices $(p_s)_{s=1}^{t-1}$ is equivalent to dependence on $(\varepsilon_s)_{s=1}^{t-1}$.

For our results to be robust to widening the class of trading policies in this way, we must narrow the class of volume functions that we consider. For this section only, we assume that $v(x, n) = x + n$. In Section 7.1, we discuss why, under other choices for v, the VWAP contract can fail to be optimal if the broker may adjust his strategy in response to previously observed prices.

For our previous results to extend, we must also impose an additional restriction on the distribution of $\boldsymbol{\varepsilon}$. For this section only, we assume that $\boldsymbol{\varepsilon}$ satisfies

$$\mathbb{E}[\varepsilon_{t+1} - \varepsilon_t | \varepsilon_1, \ldots, \varepsilon_{t-1}] = 0 \tag{2}$$

for all $t = 1, \ldots, T-1$. Condition (2) means that the predictions of ε_{t+1} and ε_t result in the same value when using only knowledge of the process up to time $t-1$. Examples of such processes include sets of independent random variables with constant mean as well as random walks of the form $\varepsilon_{t+1} = \varepsilon_t + w_{t+1}$ for $w_1, \ldots, w_T$ independent zero-mean random variables and with a constant ε_0. Under condition (2), all ε_t have the same mean, which we still denote by μ.

All results stated in Section 5 extend to this version of the model. Most proofs remain unchanged. The only difference lies in one part of the proof of Theorem 4. Nevertheless, that theorem remains true.

Theorem 4′. *In the alternate version of the model described in this section, the contract* τ^{VWAP} *is optimal.*

6.2 Agency trading

In the baseline model, the feasible contracts were all functions $\tau : \mathbb{R}^T \times \mathbb{R}^T_{++} \to \mathbb{R}$ of prices $\boldsymbol{p}$ and total volumes $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{n})$. This would be the appropriate set of contracts to consider if what the client could observe at time T were the sequence of prices $\boldsymbol{p}$ and the sequence of total volumes $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{n})$.

However, agency trading constitutes another commonly-observed set of trading arrangements in practice. Under such arrangements, the client reimburses the broker for all trading costs that he incurs. In the language of our model, this corresponds to reimbursement in the amount of $\tau^A = \boldsymbol{p} \cdot \boldsymbol{x}$. This contract is not included in the set of feasible contracts considered in the baseline model, for the reason that $\boldsymbol{x}$ is typically not observed by the client in practice. However, in many asset classes, there exist regulatory bodies who can observe $\boldsymbol{x}$ and do enforce contracts of the form τ^A (though typically they do not also enforce contracts based on arbitrary functions involving $\boldsymbol{x}$).[5]

In such asset classes, it may therefore be natural to extend the class of feasible contracts to include τ^A. Because the VWAP contract τ^{VWAP} satisfies the conditions of Lemma 3, it remains optimal regardless of which other contracts are feasible. In other words, Theorem 4 extends to the version of the model in which the agency contract is available. However, Theorem 5, which states conditions under which the VWAP contract is uniquely optimal, does not extend to this version of the model. Indeed, the agency contract also solves the client's problem described in Section 5.2, with $\boldsymbol{x}^{FB}(\cdot)$ as the corresponding recommended trading policy. Consequently, we have the following result, which is an immediate corollary of Lemma 3.

Corollary 6. *If the agency contract* τ^A *is feasible, then it is optimal.*

[5]In the case of equities, there are many regulations detailing a broker's obligations for agency trades. For example, for NMS securities, SEC Rule 34-43590 stipulates certain rights of customers that pertain to information about order routing by their broker. Moreover, the existence of a consolidated tape as well as a consolidated audit trail place enables regulatory bodies to enforce these regulations. But in the case of foreign exchange trading no such regulation is currently in place.

Thus, two contracts perform particularly well in the model: the VWAP contract and the agency contract. However, the model highlights one potentially important difference between these two optima. On one hand, the agency contract makes the broker indifferent between all trading policies, and its optimality relies upon this indifference being broken in favor of the first-best policy. On the other hand, the VWAP contract provides the broker with a strict incentive to pursue the first-best policy and is therefore robust to how the broker breaks his indifference. Mathematically, we show in the proof of Theorem 4 that for all trading policies $\hat{\boldsymbol{x}}(\cdot)$ not equal to $\boldsymbol{x}^{FB}(\cdot)$ almost surely, (IC) holds with strict inequality:

$$\mathbb{E}\big[u\big(\tau^{VWAP}\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n})\big)\big] > \mathbb{E}\big[u\big(\tau^{VWAP}\left(\boldsymbol{p}, \boldsymbol{v}(\hat{\boldsymbol{x}}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \hat{\boldsymbol{x}}(\boldsymbol{n})\big)\big].$$

In addition to the above, these two contracts also differ in terms of their robustness to several unmodeled elements. For example, if a risk-averse broker might have imperfect knowledge of $\boldsymbol{n}$, then the VWAP contract would require the broker to bear risk. As a result it would no longer be optimal; in fact, it would not even be individually rational. (Nevertheless, the addition of a suitable risk premium in the form of a commission would restore individual rationality.) On the other hand, the agency contract completely insulates the broker from risk, even under imperfect knowledge of $\boldsymbol{n}$. And mainly for that reason, the agency contract would continue to be optimal. As another example, if there are many brokers who have the same level of risk aversion, but are heterogenous in terms of their knowledge of $\boldsymbol{n}$—in other words, if there is uncertainty about the skill level of different brokers—then the VWAP contract has the added advantage that it may serve as a screening device: high-ability brokers would accept the contract at a lower premium than low-ability brokers would. By shopping around, a client could use the VWAP contract to locate and employ a high-ability broker. The agency contract, on the other hand, would be equally acceptable to all types of brokers, which could prove expensive if accepted by a low-ability broker.

7 When and why VWAP optimality may fail

The baseline model described in Section 3 allows for a great deal of generality in (*i*) the price impact function h, (*ii*) the volume function v, (*iii*) the distribution of market conditions $\boldsymbol{n}$, and (*iv*) the distribution of price shocks $\boldsymbol{\varepsilon}$. Furthermore, the previous section describes two modifications of the model, both of which preserve optimality of the VWAP contract. Nevertheless, there are some limits to this generality, and this section identifies two. This analysis highlights the conditions under which the VWAP contract is optimal and also illustrates some of the economic forces at play.

7.1 Dynamic trading policies with general volume functions

The baseline model allows for a very general class of of volume functions $v(x, n)$, but restricts the set of trading policies to functions of $\boldsymbol{n}$. In contrast, the version of the model considered in Section 6.1 allows for a broader class of trading policies, allowing the broker to adjust his strategy based on prices he observes in the course of trading, but restricts to $v(x, n) = x + n$. We next explain why the VWAP contract may no longer be optimal if both v and the set of trading policies are general.

Assume that after trading according to $\boldsymbol{x}^{FB}$ in the first trading period, the broker observes a realization of ε_1 that is much greater than the expected value μ. If the broker will be compensated according to a VWAP contract, he may then want to deviate from the first-best trading policy if it is possible that by doing so he can distort $\sum_{t=1}^{T} v(x_t, n_t)$ downward. To see this, note that this distortion would increase the weight placed on p_1 in the payment specified by τ^{VWAP} above the weight placed on p_1 in the broker's trading costs $\boldsymbol{p} \cdot \boldsymbol{x}$. If p_1 is sufficiently high, this deviation would be profitable.

Such a deviation is not possible in the baseline model. There, the broker must commit to a trading policy prior to observing any information about $\boldsymbol{\varepsilon}$, and so he cannot condition his trading decisions on ε_1. Such a deviation is also not possible if $v(x, n) = x + n$, as in Section 6.1. Although trading decisions can depend on ε_1 in that version of the model, they cannot create the requisite distortion, since the total volume $\sum_{t=1}^{T} v(x_t, n_t) = \sum_{t=1}^{T} x_t + \sum_{t=1}^{T} n_t = 1 + \sum_{t=1}^{T} n_t$ does not depend on the trading policy. However, if neither of these conditions holds, then a profitable deviation of this form might exist, and therefore the VWAP contract might not achieve first best.

7.2 Permanent price impact

Although our model allows price impact to be determined by a wide class of functional forms—including both commonly-used specifications and specifications with empirical support—our approach does require that price impact is temporary. To illustrate, we show in this section that our results on the optimality of VWAP contracts do not extend to settings in which there is a permanent component of price impact. An intuition is that in such settings, the order of events matters: trades in earlier periods influence prices in later periods. In consequence, an optimal contract must account for this, so that early periods and later periods would be handled differently in determining the broker's compensation. However, the VWAP contract does not possess this property, due to the commutative property of the weighted average. Therefore, when offered a VWAP contract, it is profitable for the broker to deviate from the first-best trading policy

Turning to a specific counterexample, consider the setting of the baseline model but with the following modifications: (*i*) there are only two periods (i.e. $T = 2$); (*ii*) total volume is determined by $v(x, n) = x + n$; (*iii*) both the broker and the client are risk neutral (i.e. $u(w) = w$); and (*iv*) the price impact function allows for temporary and permanent price impact of the form

$$p_t = c \sum_{s<t} x_s + \frac{x_t}{n_t} + \varepsilon_t.$$

When $c = 0$, there is no permanent price impact, and the setting is nested by our previous analysis (for the case in which $h(y) = y$). In contrast, when $c > 0$, there is permanent price impact, since the first period volume x_1 affects the second period price p_2. We assume that c and the distribution of n_1, n_2 are such that $c < \min\{2/n_2, 1/n_1 + 1/(n_1 + 2n_2)\}$ almost surely. It can be shown that this is a necessary and sufficient condition to ensure that the constraint $x_t \geq 0$ does not bind, either in computing the first-best trading policy or in computing the broker's best response to a VWAP contract. The formulae below could be easily adjusted to accommodate violations of this condition, but at the price of a less simple presentation.

For the purposes of this analysis, we use the notation $x_1 = x$ and $x_2 = 1 - x$. We begin by deriving the first-best trading policy:

$$\min_{x \in [0,1]} \mathbb{E}\big[p_1 x + p_2(1 - x) \big| n_1, n_2\big],$$

which yields the optimal number of shares in the first period

$$x^{FB} = \frac{1/n_2 - c/2}{1/n_1 + 1/n_2 - c}. \tag{3}$$

Next, consider the broker's optimal trading strategy under a VWAP contract. Under this contract, the payment made to the broker by the client is $\tau^{\text{VWAP}} = \frac{(x+n_1)p_1 + (1-x+n_2)p_2}{n_1+n_2+1}$, so that the broker optimizes

$$\max_{x\in[0,1]} \mathbb{E}\big[\tau^{\mathrm{VWAP}} - p_1 x - p_2(1-x)\big|n_1, n_2\big].$$

Consequently, the number of shares traded by the broker in the first period is

$$\hat{x} = x^{FB} + \frac{cn_2}{2(1/n_1 + 1/n_2 - c)(n_1+n_2)},$$

where x^{FB} is the first best from (3). If $c > 0$ then $\hat{x} > x^{FB}$, which shows that the VWAP contract does not achieve first best when there is permanent price impact.

As in our baseline analysis, if the broker were to trade $\frac{n_1}{n_1+n_2}$ in the first period, then his compensation would exactly offset his costs, leaving him with zero profits. Without permanent price impact, that is the best he can do. But with permanent price impact he can do better. To obtain intuition into that, we observe that $\hat{x}$ can be alternatively expressed as

$$\hat{x} = \frac{n_1}{n_1+n_2} + \frac{cn_1}{2(1/n_1 + 1/n_2 - c)(n_1+n_2)}.$$

From this, we can see that the broker trades slightly more than $\frac{n_1}{n_1+n_2}$ in the first period. This "manipulation" leads to a slightly higher price in the second period (compared to the case in which $x_1 = \frac{n_1}{n_1+n_2}$). The higher price in the second period affects both the market VWAP and the broker's costs. However, because the broker traded slightly more in the first period, he is less affected on a volume-weighted basis than what is reflected in the market VWAP, hence can beat the market VWAP overall.

Note that this counterexample is presented in a risk-neutral setting. When the broker is risk averse as in the model of Section 3, how much he can benefit depends on his risk preferences and the relation between permanent price impact and price risk.

8 Conclusion

Institutional investors often delegate the execution of their trades to brokers. The guaranteed VWAP contract is an agreement that the broker will trade with the client a predetermined quantity of a security at some future time at the volume-weighted average price (VWAP) over the intervening time interval. This contract falls in the class of principal trading arrangements and is used across many asset classes.

In this paper we show that contracts based on the VWAP benchmark uniquely solve a principal-agent problem between a client and her broker. The optimality of this contract stems from the fact that it not only incentivizes the broker to schedule trades in the efficient way but also mitigates the need for the client to compensate the broker for bearing risk induced by price fluctuations.

These insights might be applied to foreign exchange markets, which have been riddled with scandals of manipulation, a fact that has long been recognized by the members of the Foreign Exchange Joint Standing Committee (FXJSC, 2008). Foreign exchange trading is an area with considerable data limitations that affect the degree to which prices and volumes are visible, especially to smaller clients. In the language of our model, not every arbitrary function of prices and volumes is a feasible contract. Nevertheless, there is scope for authorities—either a platform or a regulatory body—with access to data to increase the set of feasible contracts by publishing certain computations.

In particular, there may be scope for authorities to publish a benchmark price, which can be used as the reference price in the optimal client-broker contract. And to the extent that the FX market resembles the setting of our model, this would be the VWAP computed over a certain interval. By so enabling clients to propose the optimal contract, or even a near-optimal one, transactions costs would be reduced, with the likely effect of allowing more gains from trade to be realized. Platforms might also benefit, since traded volume would likely increase as well. This suggests that the definition of the fix should be amended to more closely resemble a volume-weighted average price. Consistent with this, the Financial Stability Board as well as the FXSCJ have in fact recommended modifying prevailing definitions to include the use of volumes (FSB, 2014; FXJSC, 2008).

A Auxiliary lemmas

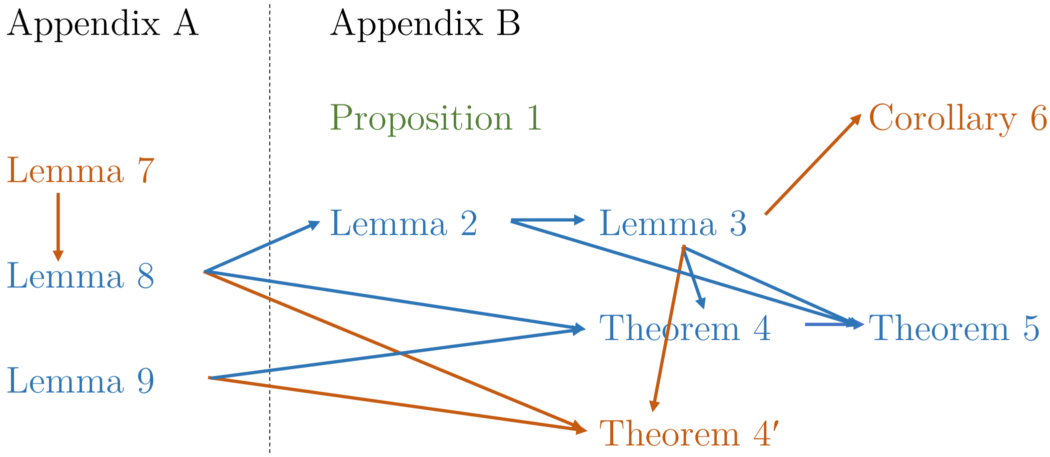

Figure 1: Roadmap of the Appendix: green indicates the result is used for the micro-foundation of the model, blue relates to a main result, and red is about model extensions.

Lemma 7. *For random variables $\varepsilon_1, \dots, \varepsilon_T$, the following are equivalent:*

(a) $\mathbb{E}[\varepsilon_t - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}] = 0$ *for all* $t = 1, \dots, T-1$,

(b) $\mathbb{E}[\varepsilon_{t+1} - \varepsilon_t | \varepsilon_1, \dots, \varepsilon_{t-1}] = 0$ *for all* $t = 1, \dots, T-1$.

Proof of Lemma 7. To show that (a) implies (b), we compute

$$\begin{aligned}
\mathbb{E}[\varepsilon_{t+1} - \varepsilon_t | \varepsilon_1, \dots, \varepsilon_{t-1}] &= \mathbb{E}[\varepsilon_{t+1} - \varepsilon_T + \varepsilon_T - \varepsilon_t | \varepsilon_1, \dots, \varepsilon_{t-1}] \\
&= \mathbb{E}[\varepsilon_{t+1} - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}] - \underbrace{\mathbb{E}[\varepsilon_t - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}]}_{=0 \text{ by (a)}} \\
&= \mathbb{E}\Big[\underbrace{\mathbb{E}[\varepsilon_{t+1} - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}, \varepsilon_t]}_{=0 \text{ by (a)}} \Big| \varepsilon_1, \dots, \varepsilon_{t-1} \Big] \\
&= 0.
\end{aligned}$$

Conversely, assume that (b) holds so that

$$\begin{aligned}
\mathbb{E}[\varepsilon_t - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}] &= \mathbb{E}\Big[\varepsilon_t - \underbrace{\mathbb{E}[\varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}, \dots, \varepsilon_{T-2}]}_{=\mathbb{E}[\varepsilon_{T-1} | \varepsilon_1, \dots, \varepsilon_{t-1}, \dots, \varepsilon_{T-2}] \text{ by (b)}} \Big| \varepsilon_1, \dots, \varepsilon_{t-1} \Big] \\
&= \mathbb{E}[\varepsilon_t - \varepsilon_{T-1} | \varepsilon_1, \dots, \varepsilon_{t-1}].
\end{aligned}$$

From this, we deduce

$$\mathbb{E}[\varepsilon_t - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}] = \mathbb{E}[\varepsilon_t - \varepsilon_{T-1} | \varepsilon_1, \dots, \varepsilon_{t-1}] = \cdots = \mathbb{E}[\varepsilon_t - \varepsilon_{t+1} | \varepsilon_1, \dots, \varepsilon_{t-1}] = 0$$

by repeatedly applying (b). □

Lemma 8. *Consider random variables* $\boldsymbol{\varepsilon}$ *that are independent of* $\boldsymbol{n}$*. Assume that either*

1. *there exist functions* $\boldsymbol{x}(\cdot)$ *depending on* $\boldsymbol{n}$ *such that it holds* $\sum_{t=1}^{T} x_t(\boldsymbol{n}) = 1$ *almost surely, and* $\mathbb{E}[\varepsilon_t] = \mu$ *for all* t*, or*

2. *there exist functions* $\boldsymbol{x}(\cdot)$ *with each* x_t *depending on* $\boldsymbol{n}$ *and* $(\varepsilon_s)_{s=1}^{t-1}$ *such that it holds* $\sum_{t=1}^{T} x_t\big(\boldsymbol{n}, (\varepsilon_s)_{s=1}^{t-1}\big) = 1$ *almost surely, and* $\boldsymbol{\varepsilon}$ *satisfies* (2) *for all* t*.*

Then

$$\sum_{t=1}^{T} \mathbb{E}[\varepsilon_t x_t] = \mu,$$

suppressing in x_t *the arguments* $\boldsymbol{n}$ *and* $\big(\boldsymbol{n}, (\varepsilon_s)_{s=1}^{t-1}\big)$ *in cases 1 and 2, respectively.*

Proof of Lemma 8. We also suppress the argument in x_t in this proof. In the first case, we compute

$$\sum_{t=1}^{T} \mathbb{E}[\varepsilon_t x_t] = \sum_{t=1}^{T} \mathbb{E}\big[x_t \mathbb{E}[\varepsilon_t | \boldsymbol{n}]\big] = \mathbb{E}\left[\sum_{t=1}^{T} x_t \mu\right] = \mu,$$

using that $\mathbb{E}[\varepsilon_t | \boldsymbol{n}] = \mathbb{E}[\varepsilon_t] = \mu$ almost surely since $\boldsymbol{\varepsilon}$ and $\boldsymbol{n}$ are independent. In the second case, we argue as follows:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[\varepsilon_t x_t] &= \mathbb{E}[\varepsilon_T] + \sum_{t=1}^{T-1} \mathbb{E}[(\varepsilon_t - \varepsilon_T) x_t] \\
&= \mu + \sum_{t=1}^{T-1} \mathbb{E}\big[\mathbb{E}[\varepsilon_t - \varepsilon_T | \boldsymbol{n}, \varepsilon_1, \dots, \varepsilon_{t-1}] x_t\big] \\
&= \mu + \sum_{t=1}^{T-1} \mathbb{E}\big[\mathbb{E}[\varepsilon_t - \varepsilon_T | \varepsilon_1, \dots, \varepsilon_{t-1}] x_t\big] \\
&= \mu,
\end{aligned}$$

where the penultimate step is thanks to independence of $\boldsymbol{\varepsilon}$ from $\boldsymbol{n}$, and the final step is thanks to (2) and Lemma 7. □

Lemma 9. *If a positive function v with domain $dom(v) \subseteq \mathbb{R}_+ \times \mathbb{R}_{++}$ satisfies* (1)*, then*

$$\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} h\left(\frac{x_t}{n_t}\right)v(x_t,n_t)\right) \leq \left(\sum_{t=1}^{T} v(x_t,n_t)\right)\left(\sum_{t=1}^{T} h\left(\frac{x_t}{n_t}\right)x_t\right) \tag{4}$$

for all $T \in \mathbb{N}$ and $(x_1,n_1),\ldots,(x_T,n_T) \in dom(v)$. If v satisfies (1) *and $\sum_{t=1}^{T} x_t = 1$, then equality in* (4) *holds if and only if $x_t = \frac{n_t}{\sum_{s=1}^{T} n_s}$ for all $t = 1,\ldots,T$. Moreover, these conditions also imply that $x_t = \frac{v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)}$ for all $t = 1,\ldots,T$.*

Proof of Lemma 9. We prove (4) by induction over T.
Induction base: For $T = 1$, (4) becomes

$$x_1 h\left(\frac{x_1}{n_1}\right)v(x_1,n_1) \leq v(x_1,n_1)h\left(\frac{x_1}{n_1}\right)x_1,$$

which holds with equality.
Induction step: We can write (4) as

$$\begin{aligned}
&\left(\sum_{t=1}^{T-1} x_t\right)\left(\sum_{t=1}^{T-1} h\left(\frac{x_t}{n_t}\right)v(x_t,n_t)\right) + h\left(\frac{x_T}{n_T}\right)v(x_T,n_T)\sum_{t=1}^{T-1} x_t + x_T\sum_{t=1}^{T-1} h\left(\frac{x_t}{n_t}\right)v(x_t,n_t) \\
&\leq \left(\sum_{t=1}^{T-1} v(x_t,n_t)\right)\left(\sum_{t=1}^{T-1} h\left(\frac{x_t}{n_t}\right)x_t\right) + h\left(\frac{x_T}{n_T}\right)x_T\sum_{t=1}^{T-1} v(x_t,n_t) + v(x_T,n_T)\sum_{t=1}^{T-1} h\left(\frac{x_t}{n_t}\right)x_t.
\end{aligned}$$

Using the induction hypothesis, it is enough to show

$$h\left(\frac{x_T}{n_T}\right)v(x_T,n_T)x_t + x_T h\left(\frac{x_t}{n_t}\right)v(x_t,n_t) \leq h\left(\frac{x_T}{n_T}\right)x_T v(x_t,n_t) + v(x_T,n_T)h\left(\frac{x_t}{n_t}\right)x_t \tag{5}$$

for every $t = 1,2,\ldots,T-1$. Rearranging terms, (5) is equivalent to

$$\big(x_t v(x_T,n_T) - x_T v(x_t,n_t)\big)\left(h\left(\frac{x_T}{n_T}\right) - h\left(\frac{x_t}{n_t}\right)\right) \leq 0. \tag{6}$$

By assumption (1) and because h is increasing, the first factor is positive (negative) whenever the second factor is negative (positive), which concludes the proof of (4).

It is straightforward to check that if $x_t = \frac{n_t}{\sum_{s=1}^{T} n_s}$ for all $t = 1,\ldots,T$, then (4) holds with equality. To establish the other direction, suppose that $\sum_{t=1}^{T} x_t = 1$ and that (4) holds with equality. By the induction hypothesis, this can be the case only if (6) holds with equality for all $t = 1, \ldots, T-1$. Note that (6) holds with equality if and only if $x_t v(x_T, n_T) = x_T v(x_t, n_t)$ or $x_T n_t = x_t n_T$. However, $x_t v(x_T, n_T) = x_T v(x_t, n_t)$ implies $x_T n_t = x_t n_T$. To see this, note that $x_T n_t > x_t n_T$ implies $x_T v(x_t, n_t) > x_t v(x_T, n_T)$ by (1), and similarly, $x_T n_t < x_t n_T$ implies $x_T v(x_t, n_t) < x_t v(x_T, n_T)$ so that the equality $x_t v(x_T, n_T) = x_T v(x_t, n_t)$ can hold only if $x_T n_t = x_t n_T$. Hence, we can have equality in (4) only if $x_T n_t = x_t n_T$ for all t, which means $x_t = \frac{n_t}{\sum_{s=1}^T n_s}$ since $\sum_{t=1}^T x_t = 1$.

Finally, we show that in this case, we also have $x_t = \frac{v(x_t,n_t)}{\sum_{s=1}^T v(x_s,n_s)}$. By way of contradiction, suppose that there exists t with $x_t > \frac{v(x_t,n_t)}{\sum_{s=1}^T v(x_s,n_s)}$. Then there must exist r with $x_r < \frac{v(x_r,n_r)}{\sum_{s=1}^T v(x_s,n_s)}$ because $\sum_{t=1}^T x_t = 1$. But then $x_r v(x_t, n_t) < x_t v(x_r, n_r)$. Next, note that in general, (1) and continuity of v imply

$$x'' n' \geq x' n'' \implies x'' v(x', n') \geq x' v(x'', n''),$$

or equivalently,

$$x'' v(x', n') < x' v(x'', n'') \implies x'' n' < x' n''.$$

Therefore, $x_r v(x_t, n_t) < x_t v(x_r, n_r)$ implies $x_r n_t < x_t n_r$, which, as shown above, contradicts equality in (4). In conclusion, when there is equality in (4), we must have $x_t = \frac{v(x_t,n_t)}{\sum_{s=1}^T v(x_s,n_s)}$ for all $t = 1, \ldots, T$. □

B Proofs

Throughout this Appendix, we will typically write a trading policy as $\boldsymbol{x}(\boldsymbol{n})$ for notational convenience, despite the potential dependence on previous prices (*viz.* for the version of the model considered in Section 6.1).

Proof of Proposition 1. We first note that the function

$$\Gamma(x) = \Phi(x) + \frac{1}{x}\phi(x)$$

is decreasing on the positive reals because

$$\Gamma'(x) = \phi(x) + \frac{1}{x}\underbrace{\phi'(x)}_{=-x\phi(x)} - \frac{1}{x^2}\phi(x) = -\frac{1}{x^2}\phi(x) < 0.$$

Next, we verify that the claim holds for cases in which $x' > 0$. Suppose we have $x''n' > x'n'' > 0$. Then $\frac{x''}{n''} > \frac{x'}{n'} > 0$. From the above, we therefore have that $\Gamma\left(\frac{x'}{n'}\right) > \Gamma\left(\frac{x''}{n''}\right)$, which is equivalent to

$$\Phi\left(\frac{x'}{n'}\right) + \frac{n'}{x'}\phi\left(\frac{x'}{n'}\right) > \Phi\left(\frac{x''}{n''}\right) + \frac{n''}{x''}\phi\left(\frac{x''}{n''}\right).$$

Multiplying both sides by $x'x''$ and using $v(x,n) = x\Phi\left(\frac{x}{n}\right) + n\phi\left(\frac{x}{n}\right)$, we obtain $x''v(x',n'') > x'v(x'',n'')$, as desired.

Finally, observe that the claim holds for cases in which $x' = 0$. To see this, note that in that case, the right-hand sides of both the antecedent and consequent in the claim are equal to zero. Thus, if the antecedent holds, then its left-hand side must be positive, which implies $x'' > 0$, and which implies that the left-hand side of the consequent must be positive so that the consequent holds as well. $\square$

Proof of Lemma 2. Plugging in $\boldsymbol{p}$, a trading policy $\boldsymbol{x}(\cdot)$ is first best if for all $\boldsymbol{n}$, $\boldsymbol{x}(\boldsymbol{n})$ minimizes the following objective subject to the constraint $\sum_{t=1}^{T} x_t = 1$:

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(h\left(\frac{x_t}{n_t}\right)x_t + \varepsilon_t x_t\right)\middle|\boldsymbol{n}\right] = \sum_{t=1}^{T}\mathbb{E}\left[h\left(\frac{x_t}{n_t}\right)x_t\middle|\boldsymbol{n}\right] + \sum_{t=1}^{T}\mathbb{E}[\varepsilon_t x_t|\boldsymbol{n}].$$

By Lemma 8, the last term equates to μ almost surely. We therefore find that this objective, the expected trading cost conditional on $\boldsymbol{n}$, is

$$\begin{aligned}
&\mu + \mathbb{E}\left[\sum_{t=1}^{T} h\left(\frac{x_t}{n_t}\right)x_t\middle|\boldsymbol{n}\right] \\
&= \mu + \mathbb{E}\left[\left(\sum_{s=1}^{T} n_s\right)\left(\frac{1}{\sum_{s=1}^{T} n_s}\sum_{t=1}^{T}\frac{x_t}{n_t}h\left(\frac{x_t}{n_t}\right)n_t\right)\middle|\boldsymbol{n}\right] \\
&\geq \mu + \mathbb{E}\left[\left(\sum_{s=1}^{T} n_s\right)\left(\frac{1}{\sum_{s=1}^{T} n_s}\sum_{t=1}^{T}\frac{x_t}{n_t}n_t\right)h\left(\frac{1}{\sum_{s=1}^{T} n_s}\sum_{t=1}^{T}\frac{x_t}{n_t}n_t\right)\middle|\boldsymbol{n}\right] \\
&= \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{s=1}^{T} n_s}\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} x_t\right)\middle|\boldsymbol{n}\right] \\
&= \mu + h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)
\end{aligned}$$

almost surely, where the second step in the above uses Jensen's inequality applied to the convex function $yh(y)$, and the final step uses $\sum_{t=1}^{T} x_t = 1$. Equality in the above holds if and only if $x_1/n_1 = x_2/n_2 = \cdots = x_T/n_T$. Since we must have $\sum_{t=1}^{T} x_t = 1$, the expected trading cost conditional on $\boldsymbol{n}$ is minimized if and only if the trading schedule is $\boldsymbol{x} = \left(\frac{n_t}{\sum_{s=1}^{T} n_s}\right)_{t=1}^{T}$. Thus, $\boldsymbol{x}^{FB}(\cdot)$ is the first-best trading policy, and it results in the unconditional expected trading cost

$$\mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right]. \qquad \square$$

Proof of Lemma 3. Sufficiency. If τ satisfies condition (i), then $(\tau, \boldsymbol{x}^{FB}(\cdot))$ satisfies (IC). Similarly, if τ satisfies condition (ii), then $(\tau, \boldsymbol{x}^{FB}(\cdot))$ satisfies (IR). Furthermore, condition (ii) also implies

$$\mathbb{E}\big[\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right)\big] = \mathbb{E}\big[\boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n})\big] = \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right]$$

by Lemma 2. Moreover, no pair $(\tau', \boldsymbol{x}(\cdot))$ satisfying (IR) can better this objective. To see this, first note that (IR) requires

$$\mathbb{E}\Big[u\big(\tau'\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})\right) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\big)\Big] \geq u(0).$$

Since u is concave and strictly increasing, this requires

$$\mathbb{E}\left[\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n}))\right] \geq \mathbb{E}\left[\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right] \geq \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right],$$

where the last step follows from Lemma 2.

Necessity. Now assume that there exists τ satisfying the two conditions and let τ' also be an optimal contract. Then there must exist some $\boldsymbol{x}(\cdot)$ such that $(\tau', \boldsymbol{x}(\cdot))$ satisfies (IR) and (IC) and where

$$\mathbb{E}\left[\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n}))\right] = \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right]. \tag{7}$$

First, we claim that $\boldsymbol{x}(\cdot) = \boldsymbol{x}^{FB}(\cdot)$ almost surely. Suppose by way of contradiction that this is not the case. Then Lemma 2 implies

$$\mathbb{E}\left[\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right] > \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right]. \tag{8}$$

Combining (7) and (8),

$$\mathbb{E}\left[\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n}))\right] < \mathbb{E}\left[\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right].$$

Because u is concave and strictly increasing, this implies that

$$\mathbb{E}\left[u\left(\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right)\right] < u(0),$$

which violates (IR). Next, observe that because $\boldsymbol{x}(\cdot) = \boldsymbol{x}^{FB}(\cdot)$ almost surely, $(\tau', \boldsymbol{x}(\cdot))$ satisfying (IC) implies that $(\tau', \boldsymbol{x}^{FB}(\cdot))$ satisfies it as well, which implies condition (i).

Finally, suppose by way of contradiction that condition (ii) is violated. Because $\boldsymbol{x}(\cdot) = \boldsymbol{x}^{FB}(\cdot)$ almost surely, this implies it is not the case that $\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})) = \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})$ almost surely. If u is strictly concave, then Jensen's inequality implies that

$$\mathbb{E}\left[u\left(\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right)\right] < u\left(\mathbb{E}\left[\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right]\right).$$

Because $(\tau', \boldsymbol{x}(\cdot))$ satisfies (IR), the lefthand side is bounded below by $u(0)$. Because u is strictly increasing, this implies

$$\mathbb{E}\left[\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}), \boldsymbol{n})) - \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right] > 0. \tag{9}$$

Combining (7) and (9), we obtain

$$\mathbb{E}\left[\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{n})\right] < \mu + \mathbb{E}\left[h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\right],$$

which is impossible by Lemma 2. $\square$

Proof of Theorem 4. To show that τ^{VWAP} is an optimal contract, it suffices to establish that it satisfies the two conditions of Lemma 3. We begin by observing that

$$\mathbb{E}\left[\frac{\sum_{t=1}^{T}\varepsilon_t v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{t=1}^{T}\varepsilon_t v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)}\middle|\boldsymbol{n}\right]\right] = \mathbb{E}\left[\frac{\sum_{t=1}^{T}\mathbb{E}[\varepsilon_t|\boldsymbol{n}] v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)}\right] = \mu, \quad (10)$$

using $\mathbb{E}[\varepsilon_t|\boldsymbol{n}] = \mathbb{E}[\varepsilon_t] = \mu$ almost surely.

Applying Jensen's inequality to the concave function u, we obtain that the broker's expected utility from pursuing a trading schedule $\boldsymbol{x}$ is

$$\begin{aligned}
\mathbb{E}\big[u\big(\tau^{VWAP} - \boldsymbol{p}\cdot\boldsymbol{x}\big)\big] &= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\big(h\big(\frac{x_t}{n_t}\big)+\varepsilon_t\big)v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)} - \sum_{t=1}^{T}\left(h\left(\frac{x_t}{n_t}\right)+\varepsilon_t\right)x_t\right)\right] \\
&\le \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\varepsilon_t v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)} - \sum_{t=1}^{T}\varepsilon_t x_t\right)\right] \quad (11) \\
&\le u\left(\mathbb{E}\left[\sum_{t=1}^{T}\frac{\varepsilon_t v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)} - \sum_{t=1}^{T}\varepsilon_t x_t\right]\right) \quad (12) \\
&= u(0),
\end{aligned}$$

where (11) follows from the first part of Lemma 9; and the last equality is implied by Lemma 8 and equation (10). Equality in (11) holds if and only if (4) holds almost surely, hence if and only if $x_t = \frac{n_t}{\sum_{s=1}^{T} n_s}$ almost surely, by the second part of Lemma 9. Note that in this case, we also have $x_t = \frac{v(x_t,n_t)}{\sum_{s=1}^{T} v(x_s,n_s)}$ almost surely by the third part of Lemma 9, so that there is equality in (12) as well. Thus, a trading policy $\boldsymbol{x}(\cdot)$ maximizes $\mathbb{E}\big[u\big(\tau^{VWAP}(\boldsymbol{p},\boldsymbol{v}(\boldsymbol{x}(\boldsymbol{n}),\boldsymbol{n})) - \boldsymbol{p}\cdot\boldsymbol{x}(\boldsymbol{n})\big)\big]$ if and only if it implies that $x_t = \frac{n_t}{\sum_{s=1}^{T} n_s}$ almost surely, or equivalently, if and only if it equals $\boldsymbol{x}^{FB}(\cdot)$ almost surely.

We therefore conclude that $(\tau^{VWAP}, \boldsymbol{x}^{FB}(\cdot))$ satisfies (IC), implying condition (i) of Lemma 3. But in fact, we also obtain the stronger conclusion that for all trading policies $\hat{\boldsymbol{x}}(\cdot)$ not equal to $\boldsymbol{x}^{FB}(\cdot)$ almost surely, (IC) holds with strict inequality:

$$\mathbb{E}\big[u\big(\tau^{VWAP}\big(\boldsymbol{p},\boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}),\boldsymbol{n})\big) - \boldsymbol{p}\cdot\boldsymbol{x}^{FB}(\boldsymbol{n})\big)\big] > \mathbb{E}\big[u\big(\tau^{VWAP}(\boldsymbol{p},\boldsymbol{v}(\hat{\boldsymbol{x}}(\boldsymbol{n}),\boldsymbol{n})) - \boldsymbol{p}\cdot\hat{\boldsymbol{x}}(\boldsymbol{n})\big)\big].$$

The same computation reveals that $\tau^{VWAP}\big(\boldsymbol{p},\boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}),\boldsymbol{n})\big) - \boldsymbol{p}\cdot\boldsymbol{x}^{FB}(\boldsymbol{n}) = 0$. We therefore obtain condition (ii) of Lemma 3. □

Proof of Theorem 5. Suppose that u is strictly concave and that the distributions of $\boldsymbol{\varepsilon}$ and $\boldsymbol{n}$ have full support over $\mathbb{R}^T$ and $\mathbb{R}^T_{++}$, respectively. Suppose that τ is an optimal contract. In proving Theorem 4, we established that τ^{VWAP} satisfies the conditions of Lemma 3. Therefore, the second half of that lemma requires that τ does the same. Condition (ii) of that lemma requires

$$\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right) = \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n}) \tag{13}$$

almost surely. We next rewrite the righthand side of (13):

$$\begin{aligned}
\boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{n}) &= \sum_{t=1}^{T} \left(h\left(\frac{x_t^{FB}(\boldsymbol{n})}{n_t}\right) + \varepsilon_t \right) x_t^{FB}(\boldsymbol{n}) \\
&= \sum_{t=1}^{T} \left(h\left(\frac{1}{\sum_{s=1}^{T} n_s}\right) + \varepsilon_t \right) \frac{n_t}{\sum_{s=1}^{T} n_s} \\
&= h\left(\frac{1}{\sum_{s=1}^{T} n_s}\right) + \frac{\sum_{t=1}^{T} n_t \varepsilon_t}{\sum_{s=1}^{T} n_s}.
\end{aligned}$$

Using $\boldsymbol{\iota}$ to denote a vector of ones, we also rewrite the lefthand side of (13):

$$\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})\right) = \tau\left(h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\boldsymbol{\iota} + \boldsymbol{\varepsilon}, v\left(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n}\right)\right).$$

As argued, the two expressions above coincide almost surely. We also write

$$\begin{aligned}
\tau^{VWAP}\left(h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\boldsymbol{\iota} + \boldsymbol{\varepsilon}, \boldsymbol{v}\left(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n}\right)\right) &= \frac{\sum_{t=1}^{T} \left(h\left(\frac{1}{\sum_{s=1}^{T} n_s}\right) + \varepsilon_t \right) v\left(x_t^{FB}(\boldsymbol{n}), n_t\right)}{\sum_{s=1}^{T} v\left(x_s^{FB}(\boldsymbol{n}), n_s\right)} \\
&= h\left(\frac{1}{\sum_{s=1}^{T} n_s}\right) + \frac{\sum_{t=1}^{T} v\left(x_t^{FB}(\boldsymbol{n}), n_t\right) \varepsilon_t}{\sum_{s=1}^{T} v\left(x_s^{FB}(\boldsymbol{n}), n_s\right)} \\
&= h\left(\frac{1}{\sum_{s=1}^{T} n_s}\right) + \frac{\sum_{t=1}^{T} n_t \varepsilon_t}{\sum_{s=1}^{T} n_s},
\end{aligned}$$

where the last step uses that $\frac{v(x_t^{FB}(\boldsymbol{n}), n_t)}{\sum_{s=1}^{T} v(x_s^{FB}(\boldsymbol{n}), n_s)} = \frac{n_t}{\sum_{s=1}^{T} n_s}$ by the last part of Lemma 9. Combining the above, we therefore conclude that

$$\tau\left(h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\boldsymbol{\iota}+\boldsymbol{\varepsilon}, \boldsymbol{v}\left(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n}\right)\right)=\tau^{VWAP}\left(h\left(\frac{1}{\sum_{t=1}^{T} n_t}\right)\boldsymbol{\iota}+\boldsymbol{\varepsilon}, \boldsymbol{v}\left(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n}\right)\right)$$

almost surely. By the full support assumptions on $\boldsymbol{\varepsilon}$ and $\boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{n}), \boldsymbol{n})$, this requires that $\tau = \tau^{VWAP}$ almost everywhere on its domain. □

Proof of Theorem 4′. Since $v(x, n) = x + n$, the total volume

$$\sum_{s=1}^{T} v(x_s, n_s) = \sum_{s=1}^{T} x_s + \sum_{s=1}^{T} n_s = 1 + \sum_{s=1}^{T} n_s$$

depends only on $\boldsymbol{n}$ so that

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{t=1}^{T}\varepsilon_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{t=1}^{T}\varepsilon_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)}\middle|\boldsymbol{n}\right]\right] \\
&= \mathbb{E}\left[\frac{\sum_{t=1}^{T}\mathbb{E}[\varepsilon_t x_t|\boldsymbol{n}] + \sum_{t=1}^{T}\mathbb{E}[\varepsilon_t n_t|\boldsymbol{n}]}{1+\sum_{s=1}^{T} n_s}\right] \\
&= \mathbb{E}\left[\frac{\mu + \mu\sum_{t=1}^{T} n_t}{1+\sum_{s=1}^{T} n_s}\right] = \mu, \qquad (14)
\end{aligned}$$

using Lemma 8 and $\mathbb{E}[\varepsilon_t|\boldsymbol{n}] = \mathbb{E}[\varepsilon_t] = \mu$ almost surely.

Applying Jensen's inequality to the concave function u, we obtain that the broker's expected utility from pursuing a trading schedule $\boldsymbol{x}$ is

$$\begin{aligned}
&\mathbb{E}\left[u\left(\tau^{VWAP} - \boldsymbol{p}\cdot\boldsymbol{x}\right)\right] \\
&= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\left(h\left(\frac{x_t}{n_t}\right)+\varepsilon_t\right)v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)} - \sum_{t=1}^{T}\left(h\left(\frac{x_t}{n_t}\right)+\varepsilon_t\right)x_t\right)\right] \\
&\leq \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\varepsilon_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)} - \sum_{t=1}^{T}\varepsilon_t x_t\right)\right] \qquad (15) \\
&\leq u\left(\mathbb{E}\left[\sum_{t=1}^{T}\frac{\varepsilon_t v(x_t, n_t)}{\sum_{s=1}^{T} v(x_s, n_s)} - \sum_{t=1}^{T}\varepsilon_t x_t\right]\right) \qquad (16) \\
&= u(0),
\end{aligned}$$

where (15) follows from the first part of Lemma 9; and the last equality is implied by Lemma 8 and equation (14). The remainder of the argument proceeds as in the proof of Theorem 4. □

References

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li, "Direct Estimation of Equity Market Impact," *Risk*, 2005, *18* (7), 58–62.

Battalio, Robert, Shane A. Corwin, and Robert Jennings, "Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality," *The Journal of Finance*, 2016, *71* (5), 2193–2238.

Bernhardt, Dan and Bart Taub, "Front-Running Dynamics," *Journal of Economic Theory*, 2008, *138* (1), 288–296.

Bloomberg, "Traders Said to Rig Currency Rates to Profit Off Clients," June 2013. `https://www.bloomberg.com/news/articles/2013-06-11/traders-said-to-rig-currency-rates-to-profit-off-clients` Accessed: 2018-04-25.

Carroll, Gabriel, "Robustness and Linear Contracts," *The American Economic Review*, 2015, *105* (2), 536–563.

Cartea, Álvaro and Sebastian Jaimungal, "A Closed-Form Execution Strategy to Target Volume Weighted Average Price," *SIAM Journal on Financial Mathematics*, 2016, *7* (1), 760–785.

Coulter, Brian, Joel Shapiro, and Peter Zimmerman, "A Mechanism for Libor," *Review of Finance*, 2017, *22* (2), 491–520.

Duffie, Darrell and Piotr Dworczak, "Robust Benchmark Design," *Working Paper*, 2014. `http://ssrn.com/abstract=2505846`.

Financial Stability Board, "Foreign Exchange Benchmarks," *Final Report*, September 2014.

Fishman, Michael J. and Francis A. Longstaff, "Dual Trading in Futures Markets," *The Journal of Finance*, 1992, *47* (2), 643–671.

Foreign Exchange Joint Standing Committee Chief Dealers Sub-Group, "Draft Minutes of the 13th Meeting," July 2008.

Frei, Christoph and Nicholas Westray, "Optimal Execution of a VWAP Order: A Stochastic Control Approach," *Mathematical Finance*, 2015, *25* (3), 612–639.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch, "Portfolio Performance Manipulation and Manipulation-Proof Performance Measures," *The Review of Financial Studies*, 2007, *20* (5), 1503–1546.

Henderson, Brian J., Neil D. Pearson, and Li Wang, "Pre-Trade Hedging," *Working Paper*, 2018. http://ssrn.com/abstract=3068903.

Hillion, Pierre and Matti Suominen, "The Manipulation of Closing Prices," *Journal of Financial Markets*, 2004, *7* (4), 351–375.

Holmström, Bengt and Paul Milgrom, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 1987, *55* (2), 303–328.

Humphery-Jenner, Mark, "Optimal VWAP Trading under Noisy Conditions," *Journal of Banking & Finance*, 2011, *35* (9), 2319–2329.

Kato, Takashi, "VWAP Execution as an Optimal Strategy," *JSIAM Letters*, 2015, *7*, 33–36.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

_ , Anna A. Obizhaeva, and Yajun Wang, "Smooth Trading with Overconfidence and Market Power," *Review of Economic Studies*, 2018, *85* (1), 611–662.

Ljungqvist, Alexander, Felicia Marston, Laura T Starks, Kelsey D Wei, and Hong Yan, "Conflicts of Interest in Sell-Side Research and the Moderating Role of Institutional Investors," *Journal of Financial Economics*, 2007, *85* (2), 420–456.

Malmendier, Ulrike and Devin Shanthikumar, "Do Security Analysts Speak in Two Tongues?," *The Review of Financial Studies*, 2014, *27* (5), 1287–1322.

Mastromatteo, Iacopo, Bence Tóth, and Jean-Philippe Bouchaud, "Agent-Based Models for Latent Liquidity and Concave Price Impact," *Physical Review E*, 2014, *89*, 042805.

Michaely, Roni and Kent L Womack, "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations," *The Review of Financial Studies*, 1999, *12* (4), 653–686.

Michelberger, Patrick Steffen and Jan Hendrik Witte, "Foreign Exchange Market Microstructure and the WM/Reuters 4 PM Fix," *The Journal of Finance and Data Science*, 2016, *2* (1), 26–41.

Nomura Research Institute, Ltd., "Asset Management Companies' Evaluation of Brokers – State of Asset Management Companies' Trading in 2014," April 2014. https://www.nri.com/~/media/PDF/global/opinion/lakyara/2014/lkr2014192.pdf Accessed: 2018-04-25.

Röell, Ailsa, "Dual-Capacity Trading and the Quality of the Market," *Journal of Financial Intermediation*, 1990, *1* (2), 105–124.

The United States Commodity Futures Trading Commission, "CFTC Orders Five Banks to Pay over $1.4 Billion in Penalties for Attempted Manipulation of Foreign Exchange Benchmark Rates," November 2014. https://www.cftc.gov/PressRoom/PressReleases/pr7056-14 Accessed: 2018-04-25.

The United States Department of Justice, "HSBC Holdings Plc Agrees to Pay More Than $100 Million to Resolve Fraud Charges," January 2018. https://www.justice.gov/opa/pr/hsbc-holdings-plc-agrees-pay-more-100-million-resolve-fraud-charges Accessed: 2018-04-25.

The Wall Street Journal, "Citigroup, J.P. Morgan Take Brunt of Currencies Settlement," November 2014. https://www.wsj.com/articles/banks-reach-settlement-in-foreign-exchange-rigging-probe-1415772504 Accessed: 2018-06-15.