Contracting for Financial Execution*

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

November 18, 2018

Abstract

Financial contracts often specify reference prices whose values are undetermined at the time of contracting, which makes them prone to manipulation. To study such situations, we introduce a stylized model of financial contracting between a client, who wishes to trade a large position, and her broker. We find that a simple contract based on the volume-weighted average price (VWAP) emerges as the unique optimal solution to this principal-agent problem. This result explains the popularity of guaranteed VWAP contracts in practice and also suggests considerations for the optimal design of financial benchmarks.

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

JEL Codes: G11, G14, G18, G23, D82, D86

*We thank Robert Almgren, Hendrik Bessembinder, Piotr Dworczak, Burton Hollifield (discussant), Pete Kyle, Kasper Larsen, Esen Onur, 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, these 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. Manipulation of the World Markets/Reuters Closing Spot Rates (WM/R), which are common benchmark rates for foreign currency transactions, has recently led HSBC to agree to pay more than 100 million dollars to enter into a deferred prosecution agreement, and has led to the imprisonment of the firm's head of global foreign exchange trading (DOJ, 2018a,b). In addition, there are a number of similar examples of manipulation of foreign exchange rates (Bloomberg, 2013; WSJ, 2018). These 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) prevailing in the market over the intervening time interval. The aforementioned examples come from foreign exchange, and indeed, we view that market as a leading application for our analysis. However, both the problem of how a client should contract with her broker and the model that we introduce are general, and our analysis is not restricted to any single asset class.

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 temporary 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. In the first-best solution, the broker schedules trades so as to equalize the marginal cost of trading an extra unit across time periods. In the market model, this is equivalent to what is known in practice as a *volume participation strategy,* wherein the order is split over time so as to be proportional to the volume profile of the market.

Whether the broker's trading decisions actually accord with the first-best benchmark will, however, depend on the form of the contract. Since the client cannot observe the broker's trades, deviations from the first-best cannot be detected directly. Rather, the client must consider how the broker's incentives will be shaped by the contract offered. 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 client pays the broker at 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 guaranteed market-on-close contract, where the client pays the broker at the price prevailing 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 deviate from the first-best trading policy by tilting his trades toward the last period. Because his trades have temporary price impact, this will move the price in period T, thereby creating a gap between this price and the average price that he paid to acquire the position. This behavior corresponds to what is known in practice as "banging the close."

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 pursues the first-best trading policy. And unlike many other contracts that are tied to future prices, the VWAP contract does not incentivize the broker to deviate from the first-best policy. This helps explain the popularity of such contracts in practice.[1]

The contracts described in the previous paragraphs are examples of "principal trading"

[1] For example, Nomura (2014) provides some evidence of the prevalence of these contracts in the domain of equity 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 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.

We also apply our results to the question of benchmark design. This is particularly relevant to markets in which it is difficult or impossible for clients to observe prices or volumes directly. Nevertheless, a third party, such as a regulator or platform, may publish a benchmark that summarizes these quantities, and that benchmark may be observed and contracted on. Our results suggest that it may be desirable to compute this benchmark as the VWAP, since, in that case, it becomes possible for the client to propose the optimal contract by referring to the benchmark.

The remainder of this paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the baseline model. In our initial description of the model (Section 3.1), we portray the nature of price impact in a reduced form way, allowing for a large class of functional forms. However, we also provide (in Section 3.2) a micro-foundation for one particular member of this class. Section 4 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 5, 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 5.1) or when agency trading is possible (Section 5.2). However, in Section 6, 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 7 applies the model to benchmark design. 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, taking on a net position. In contrast, we analyze a setting in which a conflict may lead to suboptimal execution, even in the absence of any proprietary net positions of the broker.

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 in the trading conducted by issuers of structured equity products. Evidence of related distortions is also found by Barbon, Di Maggio, Franzoni and Landier (2017). Although the latter do not find brokers acting directly on information about an impending client transaction, they provide evidence of brokers acting *indirectly* by leaking this information to their other clients. In a similar spirit, Battalio, Corwin and Jennings (2016) find evidence of brokers routing client orders suboptimally in order to collect rebates from exchanges. Hillion and Suominen (2004) find evidence of brokers manipulating closing prices so as to enhance their reputation for execution quality. Comerton-Forde and Putniņš (2011, 2014) also analyze episodes of closing price manipulation.

Although less directly related to our analysis, conflicts also arise within the non-trading component of broker-client relationships. For example, when acting in an advisory capacity, brokers 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. Interest rate benchmarks constitute one example. Banks may have incentives 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, 2018; Coulter, Shapiro and Zimmerman, 2017). Despite many differences between this setting and that of our paper, Duffie and Dworczak (2018) 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). Finally, Duffie, Dworczak and Zhu (2017) analyze how benchmarks affect the incentives of traders in search markets, finding that the publication of a benchmark can raise social surplus.

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.

While our focus is on the contractual relationship between a client and her broker, there is also a connection to the literature on optimal trading strategies. In our model, the solution to the first-best benchmark amounts to a volume participation strategy. Similarly, Kato (2015) provides conditions under which such a participation strategy is optimal. Note that if a participation strategy is used, then the price paid for the order necessarily equates to the VWAP over the trading period. Consequently, the participation strategy can be equivalently thought of as a strategy designed to target VWAP. Given the importance of VWAP in practice, there is considerable literature on how to devise such strategies (e.g. Humphery-Jenner, 2011; Frei and Westray, 2015; Cartea and Jaimungal, 2016). We depart from this literature in the sense that our primary focus is not on the first-best benchmark, but rather on the second-best—that is, the version of the problem in which there is a client-broker relationship and an agency problem between them stemming from the client's inability to observe the broker's trading decisions. One might think that this additional friction would preclude achievement of the first-best benchmark. To the contrary, we show that first-best can be achieved in the model if the broker is incentivized with an appropriate contract. Our main contribution is to identify the guaranteed VWAP contract as the unique contract to do this.

3 Model

3.1 Setting

There are two decision makers in the model: a client (the principal) and a broker (the agent). There are finitely many discrete trading periods $t \in \{1, 2, \ldots, T\}$.[2,3] And there is a single security. The client has a need to purchase a fixed number of shares, which we normalize to one, and she offers her broker a contract regarding the intended trade. 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. The main friction is one of hidden action, because the client cannot observe how the broker trades while he is acquiring the share from the market.

Trading. The client contracts to purchase one share from the broker after time T. 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. We also require the broker to purchase no more than the required share, so that he ends with an inventory of zero. 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*.

[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).

[3]Alternatively to this discrete-time model, we could consider a continuous-time model with trading during an interval $[0, T]$; see the remark at the end of Section 5.1.

Market conditions. In addition, there are time-varying market conditions, denoted by η_t, which both affect total trading volume and moderate the price impact of the broker's trades. The market condition η_t is an exogenous random variable that takes positive values. Its realization is learned by the broker prior to the start of trading, but the client knows only its distribution. We do not impose any assumptions on the joint distribution of $(\eta_t)_{t=1}^T$. In particular, $(\eta_t)_{t=1}^T$ do not need to be independent. Nor do they need to be identically distributed; thus, our approach permits the client to have some information about market conditions, albeit less than the broker. Our approach is to treat η_t as an abstract object that affects price and volume in ways specified below. Nevertheless, one concrete example is that in which η_t represents the volume traded on the market by outside traders in period t. Another example will be pursued in Section 3.2, where η_t denotes the number of outside traders.

Price. The price that prevails in period t is

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

where h is a strictly increasing function such that $yh(y)$ is strictly convex; $(\varepsilon_t)_{t=1}^T$ are random variables that all have the same expectation conditional on $(\eta_t)_{t=1}^T$, namely, $\mathbb{E}[\varepsilon_t|\eta_1, \ldots, \eta_T] = \mu$ almost surely for all t and some constant μ. However, we impose no further assumptions on $(\varepsilon_t)_{t=1}^T$: they do not need to be independent of each other or $(\eta_t)_{t=1}^T$, and neither do they need to be 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 η_t, and not on x_t itself. In light of this, the market condition η_t can be thought of as parametrizing how steep price impact will be in period t. In the case where η_t represents outside volume, using precisely the ratio x_t/η_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 $h(x_t/\eta_t) = (x_t/\eta_t)^a$ 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).[4]

Note also that the price depends only on contemporaneous values of x_t and η_t. In that sense, our baseline model should be interpreted as one of temporary price impact. This is not inconsistent with the literature, where both empirically and theoretically, different forms of price impact have been analyzed without clear conclusions as to its functional form (or how that form might vary with the setting). Nevertheless, see Section 6.2 for a discussion of the issues that arise when price impact has a permanent component.

Volume. The total volume (including the broker's trades) at time t is given by $v(x_t, \eta_t)$, where v is a positive function with domain $\text{dom}(v) \subseteq \mathbb{R}_+ \times \mathbb{R}_{++}$ that for all $x \neq 0$ takes the form $v(x, \eta) = xV(x/\eta)$ for some function $V(y)$ that is strictly decreasing for $y \neq 0$. For example, in the case where η_t represents the outside volume traded in period t, $v(x_t, \eta_t) = x_t + \eta_t$, which is indeed of the desired form, with $V(y) = 1 + 1/y$ for $y > 0$.

This assumption implies that total volume $v(x, \eta)$ is increasing in η. This is natural: in light of the fact that η_t can be thought of as parametrizing the steepness of price impact in period t, we obtain the intuitive relationship that, holding fixed the broker's volume, price impact is smaller when volume is larger. The assumption also implies that total volume $v(x, \eta)$ is homogenous of degree one. This is natural as well, since it implies that price impact is invariant to the scale of the market. Indeed, in the case where $v(x, \eta)$ is

[4] 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).

homogenous of degree one, proportionate increases in x and η correspond to an increase in scale of the market: both the broker's volume and outside volume increase at the same rate. And at the same time, price impact as measured by $h(x/\eta)$ remains constant. The following proposition establishes that the above assumption on $v(x,\eta)$ not only implies these two realistic properties but also is implied by them.

Proposition 1. *Let v be a positive function with domain $dom(v) \subseteq \mathbb{R}_+ \times \mathbb{R}_{++}$. The following are equivalent:*

(i) $v(x,\eta) = xV(x/\eta)$ for $x \neq 0$ and a function $V(y)$ that is strictly decreasing for $y \neq 0$.

(ii) $v(x,\eta)$ is homogeneous of degree one and strictly increasing in η for $(x,\eta) \in dom(v)$ with $x \neq 0$.

Whereas we have modeled prices as stochastic, our baseline formulation assumes that volumes are a deterministic function of the broker's trading schedule and market conditions.[5] This is, of course, not completely realistic: even the most knowledgeable traders are unable to perfectly forecast volumes. However, this formulation is consistent with the existence of known empirical regularities in trading volume (e.g. the so-called "liquidity smile"), whereas there are fewer such patterns in prices (and in fact no such patterns under the weak form efficient market hypothesis).

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

[5] Our later arguments go through if the price impact and total volume are of the forms $h\big(\frac{x_t}{\Xi\eta_t}\big)$ and $v(x_t,\eta_t) = x_t V\big(\frac{x_t}{\Xi\eta_t}\big)$, respectively, for a random variable Ξ that is independent from $\boldsymbol{\varepsilon}$. Under these forms of price impact and total volume, the true market conditions would be $(\Xi\eta_t)_{t=1}^T$. The broker observes $(\eta_t)_{t=1}^T$, which therefore provides only a noisy signal of market conditions, yet does nevertheless fully reveal the relative market conditions $\left(\frac{\Xi\eta_t}{\sum_{s=1}^T \Xi\eta_s}\right)_{t=1}^T$.

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 $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\eta})$ at time T. We allow the client to offer contracts that are arbitrary functions of these market outcomes.[6] 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 $\tau(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\eta}))$ 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,[7] (*iii*) the time-weighted average price (TWAP), which corresponds to $\tau = \frac{1}{T}\sum_{t=1}^{T} p_t$, and (*iv*) the VWAP, which corresponds to $\tau = \frac{\sum_{t=1}^{T} p_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)}$.

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 5.2, 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,

[6] In contrast, we imagine that other quantities are not publicly observable, and so they cannot be contracted upon (except through their influence on prices and volumes). For example, the realized market conditions $\boldsymbol{\eta}$ cannot be contracted upon because the client does not observe them. Similarly, the client also does not observe the broker's trading schedule $\boldsymbol{x}$, which is the sense in which this is a model of hidden action.

[7] For example, a so-called arrival price contract would be one in which τ_0 is the market price prevailing at the time of contracting, with the possible addition of some commission.

then he receives an outside option of 0. If he accepts the contract, then he learns $\boldsymbol{\eta}$ and chooses a trading schedule $\boldsymbol{x}$. Next, $\boldsymbol{\varepsilon}$ 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 $\boldsymbol{\eta}$ 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 5.1, 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 $\boldsymbol{\eta}$ but also on information observed in previous periods.

The broker's payoffs. Since the broker may schedule trades in a way that depends on $\boldsymbol{\eta}$, he can be thought of as choosing a *trading policy*, which is a measurable function $\boldsymbol{x}(\cdot)$ that maps each $\boldsymbol{\eta}$ 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

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

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

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

The client of our model most closely resembles a large institution, but one that is unsophisticated as regards financial trading. "Large" because the client's order generates a significant amount of temporary price impact and because of our assumption that the client is risk neutral. And "unsophisticated" because the client chooses to access the market through an external broker rather than one in house and because the client has no information about fundamentals. As a concrete example, consider Cairn Energy Plc, a British oil explorer that was harmed by HSBC in an episode alluded to in the introduction. In connection with selling part of its ownership interest in a subsidiary company, it realized a need to convert \$3.5 billion USD into British pounds, and it contracted with HSBC to execute that transaction.

3.2 Micro-foundation

In the baseline model presented in Section 3.1, prices and volumes in a period t are jointly influenced by both the broker's trades x_t and market conditions η_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 η_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 η_t in a way that is nested by our previous analysis.

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, a positive measure η_t of outside traders arrive, who then depart the market after the period has ended. Consistent with our previous notation, the realization of $\boldsymbol{\eta}$ is learned prior to the start of trading by the broker, but it is not learned by the client, and we do not impose any assumptions on the joint distribution of $(\eta_t)_{t=1}^T$.

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. We assume that an outside trader i arriving in period t submits the schedule

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

where $\theta_i = \psi_i + \varepsilon_t$.[8] The idiosyncratic component ψ_i is an independent draw from a standard normal distribution, and the common components $(\varepsilon_t)_{t=1}^T$ are random variables that all have the same expectation conditional on $(\eta_t)_{t=1}^T$, which is also consistent with our previous assumption. 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.

[8]There might be many ways to micro-found this demand schedule, but one is as follows. Suppose the security has a liquidation value of $V \sim N(0,1)$. Suppose that trader i arrives to the market having received an endowment shock of $-\theta_i$ units of the security and that his utility over final wealth is given by $u_i(w) = -\exp(-w)$. From acquiring y shares at the per-share price p_t, he receives expected utility $v_i(y, p_t) = \mathbb{E}\left[-\exp(-[V(y-\theta_i) - p_t y])\right] = -\exp(p_t y + (y-\theta_i)^2/2)$. Since there is a continuum of outside traders, trader i acts as a price taker. Taking the first-order condition with respect to y, we obtain that trader i optimally submits the demand schedule given in the text.

Solution. In period t, η_t outside traders are active. Substituting for $\theta_i = \psi_i + \varepsilon_t$ in the demand schedule submitted by outside trader i, the market-clearing condition becomes

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

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

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

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

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

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

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

This function is positive since $\eta_t > 0$ and of the form $v(x_t, \eta_t) = x_t V\left(\frac{x_t}{\eta_t}\right)$ for a strictly decreasing function V, so that the expression for volume is also nested by our previous analysis. Indeed, the function $V(y) = \Phi(y) + \frac{1}{y}\phi(y)$ is strictly decreasing because

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

4 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.

4.1 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 $\boldsymbol{\eta}$,

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

Thus, a first-best trading policy minimizes expected trading cost conditional on all realizations of the market conditions $\boldsymbol{\eta}$. 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 η_t. This trading policy is efficient because it equates the marginal cost of trading an extra unit across time periods. The final part of the lemma states that this policy leads the broker to use a volume participation strategy: he trades in proportion to the total volume profile of the market. As the relative volume curve of his trades corresponds to that of the market, the trading cost incurred by this policy, $\boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta})$, will always be equal to the market VWAP.

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

$$\boldsymbol{x}^{FB}(\boldsymbol{\eta}) = \left(\frac{\eta_t}{\sum_{s=1}^{T} \eta_s} \right)_{t=1}^{T}.$$

The expected trading cost incurred by this policy is $\mathbb{E}[\boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta})] = \mu + \mathbb{E}\Big[h\Big(\frac{1}{\sum_{t=1}^{T} \eta_t}\Big)\Big]$. *Moreover, the following equality holds*

$$\boldsymbol{x}^{FB}(\boldsymbol{\eta}) = \left(\frac{v(x_t^{FB}(\boldsymbol{\eta}), \eta_t)}{\sum_{s=1}^{T} v(x_s^{FB}(\boldsymbol{\eta}), \eta_s)} \right)_{t=1}^{T}.$$

4.2 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:

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

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

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

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 $\hat{\boldsymbol{x}}(\cdot)$:

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

(ii) $\tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})\right) = \boldsymbol{p}\cdot\boldsymbol{x}^{FB}(\boldsymbol{\eta})$ *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. Indeed, 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.

4.3 Optimality of VWAP

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

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

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* $\boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})$ *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. To see that condition (ii) of Lemma 3 is satisfied, recall that, by the last part of Lemma 2, the first-best trading policy corresponds to a participation strategy. As a result, the costs incurred by the policy, $\boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta})$, are always equal to the market VWAP, and thus always equal to the payment $\tau^{VWAP}(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})$. The meat of the argument lies in establishing that condition (i) of Lemma 3 is satisfied.

To see why condition (i) of the lemma is satisfied, suppose that a broker who is compensated according to τ^{VWAP} considers deviating from $\boldsymbol{x}^{FB}(\cdot)$ to trade $\delta > 0$ fewer shares at time t and δ more shares at time t'. By the definition of $\boldsymbol{x}^{FB}(\cdot)$, this will raise the broker's expected costs for acquiring the position from the market, $\mathbb{E}[\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{\eta})]$. Indeed, this deviation will lower the price at t and raise the price at t' so that more shares are being acquired at a higher price and fewer shares at a lower price. But for a similar reason, this will also raise the broker's expected revenue received as payment from the client, $\mathbb{E}\left[\frac{1}{\sum_{s=1}^{T} v(x_s,\eta_s)}\boldsymbol{p}\cdot\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\eta})\right]$. The key observation is that costs increase by more than revenue. Indeed, what matters for the magnitude of the change in costs is the fraction of the broker's volume that is shifted from t to t', which in this case is δ. But what matters for the change in revenue is the fraction of aggregate volume that is shifted. This is a smaller fraction than δ for the reason that the effect is muted by the volume accounted for by outside traders.

For proving the uniqueness result of 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.[9]

Some intuition for the uniqueness result in Theorem 5 can be gleaned by studying why contracts tied to other common benchmark prices fail to be optimal in our model. First, consider a contract benchmarked to the closing price: $\tau = p_T$. This contract, which corresponds to what is known in practice as a guaranteed market-on-close contract, incentivizes the broker to deviate from the first-best trading policy by tilting his trades toward the last period, a behavior known in practice as "banging the close." For example, a risk-neutral broker would trade so as to maximize the expected gap between the closing price and his average price paid.[10]

At the other extreme, an infinitely risk-averse broker would only trade in the last period and thereby completely ignore his knowledge of market conditions. This failure to induce

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

[10]This type of suboptimal execution also appears in Saakvitne (2016) who develops a model of brokers that are incentivized on the basis of such contracts and therefore trade in this way.

the first-best trading policy destroys optimality. Next, consider a contract benchmarked to the TWAP: $\tau = \frac{1}{T}\sum_{t=1}^{T} p_t$. This contract also incentivizes the broker to deviate from the first-best trading policy—in this case by smoothing trading across time periods more than the first-best policy prescribes—and so it also fails to be optimal. Finally, consider fixed price contracts. By paying the broker a pre-determined amount, regardless of the prices or volumes that are realized, these contracts require the broker to bear some price risk. And if the broker is risk averse, this constitutes inefficient risk sharing, which destroys optimality.

The previous discussion also highlights why the uniqueness result requires risk aversion: an appropriately specified fixed price (i.e. "sell the firm") contract would also attain optimality 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.

Finally, note that to the extent our model is only an approximation of reality, the aforementioned results on the guaranteed VWAP contract might be expected to hold only in approximation. And in particular, it might be necessary to add a small commission in order to ensure that the (IR) constraint is satisfied. Consistent with this, such "VWAP plus commission" contracts are commonly observed in practice.

5 Extensions

We now turn to the question of when our main results generalize beyond our baseline model. In this section, we consider two modifications of the model in which our main result—that the VWAP contract is optimal—does indeed extend. In Section 5.1, we enrich the set of deviations available to the broker by allowing him to condition his trading decisions on the realizations of previous prices. Under an additional assumption on the volume function

$v(x, \eta)$, we show that the VWAP contract remains optimal. In Section 5.2, 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.

5.1 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 $\boldsymbol{\eta}$ and $(p_s)_{s=1}^{t-1}$, (*ii*) for all t, $x_t\big(\boldsymbol{\eta}, (p_s)_{s=1}^{t-1}\big) \geq 0$ almost surely, and (*iii*) $\sum_{t=1}^T x_t\big(\boldsymbol{\eta}, (p_s)_{s=1}^{t-1}\big) = 1$ almost surely.

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 $v(x, \eta) = x + \eta$. In Section 6.1, 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

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

for all $t = 1, \ldots, T-1$. Condition (1) 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 $\varepsilon_{t+1} = \varepsilon_t + \psi_{t+1}$ for $\psi_1, \ldots, \psi_T$ independent zero-mean random variables and with a constant ε_0. Under condition (1), all ε_t have the same expectation conditional on $\boldsymbol{\eta}$, which we still denote by μ. Note that given the market conditions $\boldsymbol{\eta}$, dependence on the previous prices $(p_s)_{s=1}^{t-1}$ is equivalent to dependence on $(\varepsilon_s)_{s=1}^{t-1}$. Therefore, we can equally well condition in (1) on $\boldsymbol{\eta}, p_1, \ldots, p_{t-1}$, which represents the information available to the broker when making the time-t trading decision. Thus, (1) means that the price shocks between t and $t+1$ are unpredictable with the knowledge of the broker before time t.

All results stated in Section 4 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, the same result obtains:

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

Remark. It is also possible to analyze versions of the model with a continuous time trading period $[0, T]$ rather than the discrete trading periods $\{1, 2, \ldots, T\}$. In such a continuous-time model, a trading policy prescribes a trading rate. Our results continue to hold for continuous-time versions of both our baseline model and its extensions, provided that the setting is suitably adjusted. One subtlety is that for the continuous-time version of the dynamic trading policies extension considered in this section, condition (1) becomes the martingale property for $\boldsymbol{\varepsilon}$ (with respect to the σ-algebras generated by itself and the entire process $\boldsymbol{\eta}$). By contrast, condition (1) in discrete time is slightly more general than the martingale property because the conditioning is only over $\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}$ and not $\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t$. We refrain from spelling out the details for the continuous-time model because it would not offer additional insight compared to our discrete-time model.

5.2 Agency trading

In the baseline model, the feasible contracts were all functions of the sequence of prices and total volumes. This would be the appropriate set of contracts to consider if that were all the client could observe and contract on at time T.

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 $\boldsymbol{x}$).[11]

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 4.2, 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.

[11] 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.

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

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.[12] 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:

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

In addition to the distinction 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 $\boldsymbol{\eta}$, 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. (But as previously observed, individual rationality can be restored through the addition of a suitable risk premium.) On the other hand, the agency contract completely insulates the broker from risk, even under imperfect knowledge of $\boldsymbol{\eta}$. And mainly for that reason, the agency contract would continue to be optimal.

As another example of how these two contracts differ in robustness to unmodeled elements, suppose there are many brokers who have the same level of risk aversion but are

[12] Thus, the agency contract fails to be robust to many perturbations of the model. For instance, suppose the model is perturbed by adding vanishingly small effort costs for the broker, where different trading policies may require different amounts of effort. It is natural to assume that the first-best policy would not minimize these effort costs. In this perturbation, the first-best policy would then no longer be incentive compatible under the agency contract.

heterogeneous in terms of skill, which we model as knowledge of market conditions. For concreteness, suppose that there are two types of brokers: (*i*) high-skill brokers, who have perfect knowledge of $\boldsymbol{\eta}$, as in the baseline model, and (*ii*) low-skill brokers, who have imperfect knowledge. Then the VWAP contract has the added advantage that it may serve as a screening device: high-skill brokers would accept it, while low-skill brokers would not. The agency contract, on the other hand, would be equally acceptable to all types of brokers, which could prove expensive to the client if accepted by a low-skill broker.

6 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 $\boldsymbol{\eta}$, 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. In highlighting some of the conditions required for the VWAP contract to be optimal, this analysis also illustrates some of the economic forces at play.

6.1 Dynamic trading policies with general volume functions

The baseline model allows for a very general class of volume functions $v(x, \eta)$, but restricts the set of trading policies to functions of $\boldsymbol{\eta}$. In contrast, the version of the model considered in Section 5.1 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 $v(x, \eta) = x + \eta$. 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 $\sum_{t=1}^{T} v(x_t, \eta_t)$ 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 $v(x, \eta) = x + \eta$, as in Section 5.1. Although trading decisions can depend on ε_1 in that version of the model, they cannot create the requisite distortion, since the total volume $\sum_{t=1}^{T} v(x_t, \eta_t) = \sum_{t=1}^{T} x_t + \sum_{t=1}^{T} \eta_t = 1 + \sum_{t=1}^{T} \eta_t$ does not depend on the trading policy. However, if neither of these conditions holds, then a profitable deviation of this form might exist, in which case the VWAP contract would no longer be optimal.[13]

6.2 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

[13]To elaborate, τ^{VWAP} would not be optimal because it would be dominated by a "VWAP minus commission" contract $\tau^{VWAP} - c$ for an appropriately chosen commission c. Indeed, if $(\tau^{VWAP}, \boldsymbol{x}(\cdot))$ satisfies (IC), then so does $(\tau^{VWAP} - c, \boldsymbol{x}(\cdot))$. Furthermore, the existence of the profitable deviation described in the text means that $(\tau^{VWAP}, \boldsymbol{x}(\cdot))$ satisfies (IR) with slack. Thus, for an appropriately chosen c, $(\tau^{VWAP} - c, \boldsymbol{x}(\cdot))$ also satisfies (IR) while being cheaper for the client. Note that while this establishes that τ^{VWAP} is not optimal, it does not solve for the optimal contract, which could be an even more substantial deviation from VWAP.

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 $v(x, \eta) = x + \eta$; (*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

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

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 η_1, η_2 are such that $c < \min\{2/\eta_2, 1/\eta_1 + 1/(\eta_1 + 2\eta_2)\}$ 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 broker's profit-maximizing trading strategy under a VWAP contract. Under this contract, the payment made by the client to the broker is $\tau^{\text{VWAP}} = \frac{(x+\eta_1)p_1+(1-x+\eta_2)p_2}{\eta_1+\eta_2+1}$, so that the broker optimizes

$$\max_{x\in[0,1]} \mathbb{E}\big[\tau^{\text{VWAP}} - p_1 x - p_2(1-x)\big|\eta_1, \eta_2\big].$$

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

$$\hat{x} = \frac{\eta_1}{\eta_1+\eta_2} + \frac{c\eta_1}{2(1/\eta_1 + 1/\eta_2 - c)(\eta_1+\eta_2)}. \tag{2}$$

As in our baseline analysis, if the broker were to trade $\frac{\eta_1}{\eta_1+\eta_2}$ 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. From the expression for $\hat{x}$ given above, we can see that the broker trades slightly more than $\frac{\eta_1}{\eta_1+\eta_2}$ in the first period. This "manipulation" leads to a slightly higher price in the second period (compared to the case in which $x_1 = \frac{\eta_1}{\eta_1+\eta_2}$). 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. Since the broker can therefore earn a positive expected profit, the VWAP contract cannot be optimal for the reasons laid out in footnote 13.

Note, moreover, that the VWAP contract also does not induce the broker to pursue the first-best trading policy. To see this, we derive the first-best trading policy:

$$\min_{x\in[0,1]} \mathbb{E}\big[p_1 x + p_2(1-x)\big|\eta_1, \eta_2\big],$$

which yields the optimal number of shares in the first period

$$x^{FB} = \frac{1/\eta_2 - c/2}{1/\eta_1 + 1/\eta_2 - c} = \hat{x} - \frac{c\eta_2}{2(1/\eta_1 + 1/\eta_2 - c)(\eta_1 + \eta_2)},$$

where $\hat{x}$ is the broker's profit-maximizing choice from (2). If $c > 0$, then $\hat{x} > x^{FB}$, which shows that the VWAP contract induces the broker to distort his trading decision away from the first-best trading policy when there is permanent price impact.

Note that this counterexample is presented in a risk-neutral setting. In that setting, it is possible to sign the direction of the distortion: the broker trades more at $t = 1$ than he would under first best. When the broker is risk averse as in the model of Section 3, the distortion could have either sign.

7 Application to benchmark design

In our baseline model, we assumed that any measurable function of prices and total volumes is a feasible contract. This large set of feasible contracts may be appropriate for modeling asset classes with transparent and publicly available trading data, such as equities. But for other asset classes, data is more opaque and difficult to access so that it is not possible to contract on prices and total volumes in arbitrary ways. Nevertheless, it is often possible to contract on a benchmark that a third party with access to data—perhaps a platform or regulator—computes and makes available. In effect, the types of contracts that are feasible depend upon how that benchmark is computed. And to the extent that clients and brokers are limited to contracting on the benchmark, market outcomes will be influenced by how the benchmark is computed, for example, in ways highlighted by our preceding analysis. This, in turn, raises questions about optimal benchmark design from the perspective of client-broker relationships.

As a specific example of the latter type of market, consider foreign exchange, where a prominent benchmark is the WM/Reuters London 4 pm fix ("the fix"). In recent years, industry participants have begun to rethink how this benchmark ought to be computed. For example, in 2015, WM/Reuters widened the relevant window for collecting prices from one to five minutes (e.g., Michelberger and Witte, 2016). What is more, a primary motive for this change seems to have been concern about brokers manipulating the fix to the detriment of their clients, a concern shared by the members of the Foreign Exchange Joint Standing Committee (FXJSC, 2008).

Analysis. If they are to facilitate the writing of desirable contracts, then how should benchmarks like the fix be computed? A simple way of altering our model so as to provide an answer to this question would be as follows. Add a benchmark administrator whose role is to compute and publish a benchmark $b(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\eta}))$. Restrict the set of feasible contracts to functions of the benchmark $\tau(b)$. Then ask: how should the benchmark administrator design the benchmark function b?

If the benchmark administrator's objective were to maximize the welfare of the client, then our previous results immediately imply that an optimal benchmark is the VWAP

$$b^{VWAP} = \frac{\sum_{t=1}^{T} p_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)}.$$

By choosing $\tau(b) = b$, the client would effectively be writing a VWAP contract, so that Theorem 4 implies that the optimum is being achieved. Likewise, if the benchmark administrator's objective were, alternatively, to maximize the sum of broker and client welfare, then similar arguments would nevertheless continue to imply optimality of the VWAP benchmark.

Discussion. According to this analysis, an optimal benchmark is the price that would be referenced in the optimal contract under full data availability for prices and total volumes. This enables clients to overcome their data limitations and to propose the optimal contract. Though not captured directly in our model, reducing transaction costs of clients in this way would, presumably, allow more gains from trade to be realized. Traded volume would likely increase as well, and for that reason, platforms might also have an incentive to publish such a benchmark. To the extent that the foreign exchange market resembles the setting of our model, our analysis suggests that the definition of the fix should be amended to more closely resemble a VWAP. 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).

One caveat that should be mentioned is that ours is a partial equilibrium model in that we assume market conditions are exogenous. While this seems reasonably appropriate for our baseline application of bilateral contracting between one client and one broker—other traders would not seem likely to be influenced by, or even aware of, the agreed-upon contract—it may be less appropriate for studying market-wide changes, such as benchmark design. In the language of our model, the choice of benchmark may affect the distribution over $\boldsymbol{\eta}$ in ways that we do not capture. Thus, the above analysis is most applicable to markets in which a large fraction of trading volume is driven by traders whose incentives are not tied directly to the benchmark, so that the aforementioned feedback effects are more likely to be relatively small.

Another caveat is that, in practice, banks may have financial interests tied to the realization of the benchmark beyond the trades that they will directly conduct at that benchmark. In the setting of foreign exchange, for example, banks may possess financial obligations that are denominated in a particular currency. Interests like these may create incentives to manipulate the benchmark beyond what is captured by our model. And the LIBOR scandal constitutes a particularly stark example of the power of these incentives. Duffie and Dworczak (2018) also tackle the question of benchmark design; one difference between their approach and ours is that they do allow for these incentives. Another difference lies in the criteria by which benchmarks are judged. Their aim is to design a benchmark that is resistant to manipulation and thus close to and informative about an underlying value. In contrast, our aim is to design a benchmark that facilitates the writing of contracts that lead to desirable outcomes. Despite these differences, Duffie and Dworczak (2018) find that in some cases—namely, when agents are able to split their trades undetected—a benchmark that resembles VWAP emerges as optimal.

Another application of our model is to "Trading At Settlement" (TAS), which is an order matching procedure that is available for some commodity futures. It allows market participants to trade futures contracts at the yet-to-be determined daily settlement price. It sometimes happens that a trader will conduct a TAS trade, then pursue offsetting trades over the course of the day. This resembles the setting of our model, with the aforementioned trader in the role of the broker and that trader's TAS counterparty in the role of the client. In such markets, a question may concern how best to compute the daily settlement price. To the extent that our results speak to this issue, they suggest that it would be appropriate to compute it as the VWAP.

Our analysis also highlights that other manners of computing the daily settlement price would leave it susceptible to a certain type of manipulation. For example, if the settlement price were the closing price, then a trader might achieve a positive expected profit through the following scheme: conduct a TAS trade, then pursue offsetting trades over the course of the day, concluding with a very large trade at the close so as to create a gap between the closing price and the average price of the offsetting trades. This is consistent with behavior observed in the market for crude oil futures. There, the settlement price is computed as the VWAP between 2:28 and 2:30 pm, the last two minutes before the close, so that it is approximately the closing price. And in fact, the CFTC has litigated some traders for manipulation of the kind described above (CFTC, 2008, 2013).[14] Approximately the same economics applies to on-close orders offered by equities exchanges. And indeed, the SEC has also litigated traders for similar types of manipulation (SEC, 2014).

8 Conclusion

Institutional investors often delegate the execution of their trades to brokers. But in many markets, it is difficult for such investors to monitor their brokers throughout the execution process. Even though brokers are typically bound by 'best interest' or 'best execution' obligations, these responsibilities are often vague and leave the broker with some leeway to act in ways that may harm his client.

Potential solutions to this conflict of interest include requiring more transparency, so that brokers could be more easily monitored, as well as more vigorous enforcement of the obligations of brokers. But another solution, on which we focus in this paper, is to search for a contract that mitigates the conflict itself by aligning the broker's interests with those of the client. We see two applications for our results. The most direct application pertains to the question of what sort of contract a client should push for in markets with public data availability. In such markets, it is possible to contract on prices and volumes, as our model presumes. And our results suggest that the guaranteed VWAP contract is optimal for the client.

A second application pertains to the question of how to design a benchmark. This

[14] Market on close (MOC) orders create similar incentives and have led to similar episodes of manipulation. One example of such behavior pertains to metals futures (WSJ, 2011).

question is particularly relevant to markets without public data availability. In such markets, participants cannot contract in arbitrary ways on prices and volumes, but they can often contract on a benchmark published by a platform or a regulator. Our results suggest that a VWAP benchmark best facilitates the writing of desirable contracts.

In markets such as foreign exchange, the prevailing benchmark more closely resembles the closing price than the VWAP. In consequence, current principal trading arrangements often effectively take the form of guaranteed market-on-close contracts. However, such contracts may induce the broker to distort his trading away from the efficient policy, instead trading an overly large quantity at the close. This prediction of the model is consistent with behavior often observed in such markets, including some of the episodes of manipulation mentioned in Section 1. To reduce the distortions stemming from such manipulation, our results recommend that the definitions of these benchmarks should be amended to more closely resemble a volume-weighted average price.

A Auxiliary lemmas

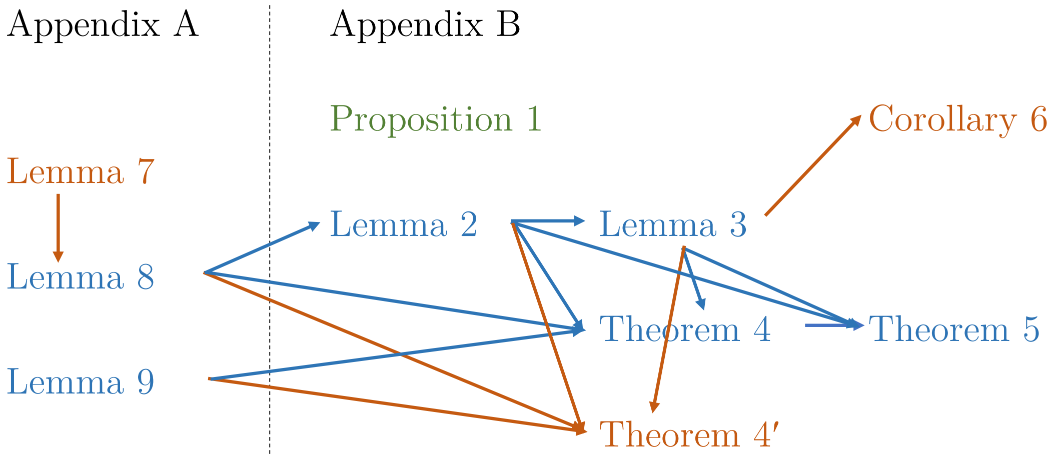

Figure 1: Roadmap of the Appendix: green indicates the result is used to motivate a model assumption, blue relates to a main result, and red is about model extensions.

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

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

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

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

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

Conversely, assume that (b) holds so that

$$\mathbb{E}[\varepsilon_t - \varepsilon_T|\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}] = \mathbb{E}\big[\varepsilon_t - \underbrace{\mathbb{E}[\varepsilon_T|\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}, \ldots, \varepsilon_{T-2}]}_{=\mathbb{E}[\varepsilon_{T-1}|\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}, \ldots, \varepsilon_{T-2}] \text{ by (b)}}\big|\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}\big]$$
$$= \mathbb{E}[\varepsilon_t - \varepsilon_{T-1}|\boldsymbol{\eta}, \varepsilon_1, \ldots, \varepsilon_{t-1}].$$

From this, we deduce

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

by repeatedly applying (b). □

Lemma 8. *Consider random variables $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$. Assume that either*

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

Then

$$\sum_{t=1}^T \mathbb{E}[\varepsilon_t x_t|\boldsymbol{\eta}] = \mu \quad \textit{almost surely,}$$

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

Proof of Lemma 8. We also suppress the argument in x_t in this proof, and all equalities are meant to hold almost surely. In the first case, we compute

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

using $\mathbb{E}[\varepsilon_t|\boldsymbol{\eta}] = \mu$. In the second case, we argue as follows:

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

where the final step is thanks to (1) and Lemma 7. □

Lemma 9. *When $v(x, \eta)$ satisfies the equivalent statements of Proposition 1, then*

$$\left(\sum_{t=1}^{T} x_t\right)\left(\sum_{t=1}^{T} h\left(\frac{x_t}{\eta_t}\right)v(x_t, \eta_t)\right) \leq \left(\sum_{t=1}^{T} v(x_t, \eta_t)\right)\left(\sum_{t=1}^{T} h\left(\frac{x_t}{\eta_t}\right)x_t\right) \tag{3}$$

for all $T \in \mathbb{N}$ and $(x_1, \eta_1), \dots, (x_T, \eta_T) \in dom(v)$. For $(x_1, \eta_1), \dots, (x_T, \eta_T) \in dom(v)$ with $\sum_{t=1}^{T} x_t = 1$, equality in (3) holds if and only if $x_t = \frac{\eta_t}{\sum_{s=1}^{T}\eta_s}$ for all $t = 1, \dots, T$.

Proof of Lemma 9. We prove (3) by induction over T.

Induction base: For $T = 1$, (3) becomes

$$x_1 h\left(\frac{x_1}{\eta_1}\right)v(x_1, \eta_1) \leq v(x_1, \eta_1)h\left(\frac{x_1}{\eta_1}\right)x_1,$$

which holds with equality.

Induction step: We can write (3) as

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

Using the induction hypothesis, it is enough to show

$$h\left(\frac{x_T}{\eta_T}\right)v(x_T,\eta_T)x_t + x_T h\left(\frac{x_t}{\eta_t}\right)v(x_t,\eta_t) \leq h\left(\frac{x_T}{\eta_T}\right)x_T v(x_t,\eta_t) + v(x_T,\eta_T)h\left(\frac{x_t}{\eta_t}\right)x_t \tag{4}$$

for every $t = 1, 2, \ldots, T-1$. Rearranging terms, (4) is equivalent to

$$\big(x_t v(x_T,\eta_T) - x_T v(x_t,\eta_t)\big)\left(h\left(\frac{x_T}{\eta_T}\right) - h\left(\frac{x_t}{\eta_t}\right)\right) \leq 0. \tag{5}$$

If $x_t = 0$ or $x_T = 0$, then (5) holds. In other cases, using $v(x,\eta) = xV(x/\eta)$ for $x \neq 0$, it becomes

$$x_t x_T \left(V\left(\frac{x_T}{\eta_T}\right) - V\left(\frac{x_t}{\eta_t}\right)\right)\left(h\left(\frac{x_T}{\eta_T}\right) - h\left(\frac{x_t}{\eta_t}\right)\right) \leq 0,$$

which is satisfied for all $(x_t,\eta_t), (x_T,\eta_T) \in \mathrm{dom}(v)$ because V is decreasing and h is increasing.

We now turn to the second part. It is straightforward to check that if $x_t = \frac{\eta_t}{\sum_{s=1}^T \eta_s}$ for all $t = 1, \ldots, T$, then (3) holds with equality. For the converse, consider $(x_1,\eta_1), \ldots, (x_T,\eta_T) \in \mathrm{dom}(v)$ with $\sum_{t=1}^T x_t = 1$ and suppose that (3) holds with equality. By the above induction hypothesis, this can be the case only if (5) holds with equality for all $t = 1, \ldots, T-1$. Note that (5) holds with equality if and only if $x_t v(x_T,\eta_T) = x_T v(x_t,\eta_t)$ or $x_T\eta_t = x_t\eta_T$. However, $x_t v(x_T,\eta_T) = x_T v(x_t,\eta_t)$ implies $x_T\eta_t = x_t\eta_T$. To see this, suppose that $x_T\eta_t > x_t\eta_T$. This can be the case only if $x_T \neq 0$. We separately consider two cases. In the first, suppose further that $x_t = 0$. Then we obtain $x_T v(x_t,\eta_t) > 0 = x_t v(x_T,\eta_T)$, where the inequality follows because $v(x_t,\eta_t)$ is positive. In the second case, suppose instead that $x_t \neq 0$. Then we obtain

$$x_T v(x_t,\eta_t) = x_T x_t V(x_t/\eta_t) > x_T x_t V(x_T/\eta_T) = x_t v(x_T,\eta_T),$$

where the inequality follows because V is strictly decreasing. By symmetry, $x_T\eta_t < x_t\eta_T$ implies $x_T v(x_t,\eta_t) < x_t v(x_T,\eta_T)$ so that the equality $x_t v(x_T,\eta_T) = x_T v(x_t,\eta_t)$ can hold only if $x_T\eta_t = x_t\eta_T$. Hence, we can have equality in (3) only if $x_T\eta_t = x_t\eta_T$ for all t, which means $x_t = \frac{\eta_t}{\sum_{s=1}^T \eta_s}$ since $\sum_{t=1}^T x_t = 1$. □

B Proofs

Throughout this Appendix, we will typically write a trading policy as $\boldsymbol{x}(\boldsymbol{\eta})$ for notational convenience, despite the potential dependence on previous prices (*viz.* for the version of the model considered in Section 5.1).

Proof of Proposition 1. It is straightforward to check that (i) implies (ii). For the converse, we define a function g by $g(x,\eta) = v(x,\eta)/x$ for $(x,\eta) \in \mathrm{dom}(v)$ with $x \neq 0$. For $(x',\eta'),(x'',\eta'') \in \mathrm{dom}(v)$ with $x'x'' \neq 0$ and $x'/\eta' = x''/\eta''$, we deduce

$$g(x',\eta') = \frac{v(x',\eta')}{x'} = \frac{v\left(\frac{\eta'}{\eta''}x'', \frac{\eta'}{\eta''}\eta''\right)}{\frac{\eta'}{\eta''}x''} = \frac{\frac{\eta'}{\eta''}v(x'',\eta'')}{\frac{\eta'}{\eta''}x''} = \frac{v(x'',\eta'')}{x''} = g(x'',\eta''), \tag{6}$$

where the third equality uses the fact that $v(x,\eta)$ is homogeneous of degree one. We partition $\mathrm{dom}(v)$ into sets $D_y = \{(x,\eta) \in \mathrm{dom}(v) : x/\eta = y\}$ for $y \in \mathbb{R}_+$. If there are no $(x,\eta) \in \mathrm{dom}(v)$ with $x/\eta = y$, we set $D_y = \emptyset$. Note that $\mathrm{dom}(v) = \bigcup_{y\in\mathbb{R}_+} D_y$ and $D_y \cap D_z = \emptyset$ for $y \neq z$. For every y with $D_y \neq \emptyset$ with $y \neq 0$, (6) implies that $g(x,\eta)$ takes the same value for all $(x,\eta) \in D_y$. Therefore, we can write $g(x,\eta) = V\left(\frac{x}{\eta}\right)$ for a function V and $(x,\eta) \in \mathrm{dom}(v)$ with $x \neq 0$, so that $v(x,\eta) = xV\left(\frac{x}{\eta}\right)$ for all $(x,\eta) \in \mathrm{dom}(v)$. Because $v(x,\eta)$ is strictly increasing in η for $(x,\eta) \in \mathrm{dom}(v)$ with $x \neq 0$, we obtain that $V(y)$ is strictly decreasing for $y \neq 0$. □

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

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

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

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

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 $x_1/\eta_1 = x_2/\eta_2 = \cdots = x_T/\eta_T$. Since we must have $\sum_{t=1}^T x_t = 1$, the expected trading cost conditional on $\boldsymbol{\eta}$ is minimized if and only if the trading schedule is $\boldsymbol{x} = \left(\frac{\eta_t}{\sum_{s=1}^T \eta_s}\right)_{t=1}^T$. Thus, $\boldsymbol{x}^{FB}(\cdot)$ is the first-best trading policy, and it results in the unconditional expected trading cost

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

The last statement of Lemma 2 follows from

$$\frac{v(x_t^{FB}(\boldsymbol{\eta}),\eta_t)}{\sum_{s=1}^T v(x_s^{FB}(\boldsymbol{\eta}),\eta_s)} = \frac{x_t^{FB}(\boldsymbol{\eta})V\left(\frac{x_t^{FB}(\boldsymbol{\eta})}{\eta_t}\right)}{\sum_{s=1}^T x_s^{FB}(\boldsymbol{\eta})V\left(\frac{x_s^{FB}(\boldsymbol{\eta})}{\eta_s}\right)}$$

$$= \frac{\frac{\eta_t}{\sum_{r=1}^T \eta_r}V\left(\frac{1}{\sum_{r=1}^T \eta_r}\right)}{\sum_{s=1}^T \frac{\eta_s}{\sum_{r=1}^T \eta_r}V\left(\frac{1}{\sum_{r=1}^T \eta_r}\right)} = \frac{\eta_t}{\sum_{r=1}^T \eta_r} = x_t^{FB}(\boldsymbol{\eta}),$$

where the first equality uses that $v(x,\eta) = xV\left(\frac{x}{\eta}\right)$ for all $(x,\eta) \in \text{dom}(v)$ by assumption. □

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

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

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

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

Since u is concave and strictly increasing, this requires

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

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

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

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

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

Combining (7) and (8),

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

Because u is concave and strictly increasing, this implies that

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

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 $\tau'(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{\eta}), \boldsymbol{\eta})) = \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{\eta})$ almost surely. If u is strictly concave, then Jensen's inequality implies that

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

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

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

Combining (7) and (9), we obtain

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

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

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

using $\mathbb{E}[\varepsilon_t|\boldsymbol{\eta}] = \mu$ 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

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

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 (3) holds almost surely, hence if and only if $x_t = \frac{\eta_t}{\sum_{s=1}^{T}\eta_s}$ almost surely, by the second part of Lemma 9. Note that in this case, we also have $x_t = \frac{v(x_t,\eta_t)}{\sum_{s=1}^{T}v(x_s,\eta_s)}$ almost surely by Lemma 2, so that there is equality in (12) as well. Thus, a trading policy $\boldsymbol{x}(\cdot)$ maximizes $\mathbb{E}\big[u\big(\tau^{VWAP}(\boldsymbol{p},\boldsymbol{v}(\boldsymbol{x}(\boldsymbol{\eta}),\boldsymbol{\eta}))-\boldsymbol{p}\cdot\boldsymbol{x}(\boldsymbol{\eta})\big)\big]$ if and only if it implies that $x_t = \frac{\eta_t}{\sum_{s=1}^T \eta_s}$ 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), which implies 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:

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

The same computation reveals that $\tau^{VWAP}\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})\right) - \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta}) = 0$. 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 $\boldsymbol{\eta}$ 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 that both of the following hold almost surely:

$$\begin{aligned} \tau\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})\right) &= \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta}) \\ \tau^{VWAP}\left(\boldsymbol{p}, \boldsymbol{v}(\boldsymbol{x}^{FB}(\boldsymbol{\eta}), \boldsymbol{\eta})\right) &= \boldsymbol{p} \cdot \boldsymbol{x}^{FB}(\boldsymbol{\eta}) \end{aligned}$$

Using $\boldsymbol{\iota}$ to denote a vector of ones, we conclude that the following holds almost surely:

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

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

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

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

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

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{t=1}^{T} \varepsilon_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{t=1}^{T} \varepsilon_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)}\middle|\boldsymbol{\eta}\right]\right] \\
&= \mathbb{E}\left[\frac{\sum_{t=1}^{T} \mathbb{E}[\varepsilon_t x_t|\boldsymbol{\eta}] + \sum_{t=1}^{T} \mathbb{E}[\varepsilon_t \eta_t|\boldsymbol{\eta}]}{1 + \sum_{s=1}^{T} \eta_s}\right] \\
&= \mathbb{E}\left[\frac{\mu + \mu \sum_{t=1}^{T} \eta_t}{1 + \sum_{s=1}^{T} \eta_s}\right] = \mu, \qquad (13)
\end{aligned}$$

using Lemma 8 and $\mathbb{E}[\varepsilon_t|\boldsymbol{\eta}] = \mu$ 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

$$\begin{aligned}
&\mathbb{E}\big[u\big(\tau^{VWAP} - \boldsymbol{p} \cdot \boldsymbol{x}\big)\big] \\
&= \mathbb{E}\left[u\left(\sum_{t=1}^{T} \frac{\big(h\big(\frac{x_t}{\eta_t}\big) + \varepsilon_t\big) v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)} - \sum_{t=1}^{T} \left(h\left(\frac{x_t}{\eta_t}\right) + \varepsilon_t\right) x_t\right)\right] \\
&\leq \mathbb{E}\left[u\left(\sum_{t=1}^{T} \frac{\varepsilon_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)} - \sum_{t=1}^{T} \varepsilon_t x_t\right)\right] \qquad (14) \\
&\leq u\left(\mathbb{E}\left[\sum_{t=1}^{T} \frac{\varepsilon_t v(x_t, \eta_t)}{\sum_{s=1}^{T} v(x_s, \eta_s)} - \sum_{t=1}^{T} \varepsilon_t x_t\right]\right) \\
&= u(0),
\end{aligned}$$

where (14) follows from the first part of Lemma 9; and the last equality is implied by Lemma 8 and equation (13). 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.

Barbon, Andrea, Marco Di Maggio, Francesco Franzoni, and Augustin Landier, "Brokers and Order Flow Leakage: Evidence from Fire Sales," *Working Paper*, 2017. http://ssrn.com/abstract=2991617.

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.

Comerton-Forde, Carole and Tālis J Putniņš, "Measuring Closing Price Manipulation," *Journal of Financial Intermediation*, 2011, *20* (2), 135–158.

_ and _ , "Stock Price Manipulation: Prevalence and Determinants," *Review of Finance*, 2014, *18* (1), 23–66.

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*, 2018. http://ssrn.com/abstract=2505846.

_ , _ , and Haoxiang Zhu, "Benchmarks in Search Markets," *The Journal of Finance*, 2017, *72* (5), 1983–2044.

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.

Saakvitne, Jo, "'Banging the Close': Price Manipulation or Optimal Execution?," *Working Paper*, 2016. http://ssrn.com/abstract=2753080.

The United States Commodity Futures Trading Commission, "CFTC v Optiver US LLC, Optiver Holding BV, Optiver VOF, Christopher Dowson, Bastiaan van Kempen, and Randal Meijer," July 2008. https://www.cftc.gov/sites/default/files/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfoptiveruscomplaint072408.pdf Accessed: 2018-08-09.

_ , "CFTC v Daniel Shak and SHK Management LLC," November 2013. http://dodd-frank.com/wp-content/uploads/2013/12/Shak-SHK-banging-the-close.pdf Accessed: 2018-08-09.

The United States Department of Justice, "Former Head of HSBC's Global Foreign Exchange Cash-Trading Sentenced to Prison for Multimillion-Dollar Front-Running Scheme," April 2018. https://www.justice.gov/opa/pr/former-head-hsbc-s-global-foreign-exchange-cash-trading-sentenced-prison-multimillion-dollar Accessed: 2018-08-08.

_ , "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 United States Securities and Exchange Commission, "SEC Order: Athena Capital Research, LLC, Release No. 34-73369," October 16 2014. https://www.sec.gov/litigation/admin/2014/34-73369.pdf Accessed: 2018-11-18.

The Wall Street Journal, "Big Fine in Metals Case," July 2011. https://www.wsj.com/articles/SB10001424053111904772304576468530892198942 Accessed: 2018-08-06.

_ , "Deutsche Bank to Pay $205 Million Fine to End N.Y. Currency-Trading Probe," June 2018. https://www.wsj.com/articles/deutsche-bank-to-pay-205-million-fine-to-end-n-y-currency-trading-probe-1529517428 Accessed: 2018-08-06.