

# Robust Benchmark Design<sup>\*</sup>

*Darrell Duffie<sup>†</sup> and Piotr Dworczak<sup>‡</sup>*

Forthcoming in Journal of Financial Economics

## Abstract

Scandals over the manipulation of Libor, foreign exchange benchmarks, and other financial benchmarks have spurred policy discussions over the appropriate design of benchmark fixings. We introduce a framework for the design of a benchmark fixing as an estimator of fair market value. The fixing data are the reports or transactions of agents whose profits depend on the fixing, and who may therefore have incentives to manipulate the fixing. We focus on linear fixings, which are weighted sums of transaction prices, with weights that depend on transaction sizes. We derive the optimal fixing under a simplifying assumption that weights are unidimensional, and we axiomatically characterize the unique benchmark that is robust to a certain form of collusion among traders. Our analysis provides a foundation for the commonly used volume-weighted average price (VWAP) and for a variant of VWAP based on unidimensional size weights. We characterize the relative advantages of these fixing designs, depending on the market characteristics.

**Keywords:** financial benchmarks, manipulation, collusion, mechanism design without transfers, volume-weighted average price

**JEL codes:** G12, G14, G18, G21, G23, D82.

---

<sup>\*</sup>We are grateful for useful discussions with members of the Market Participants Group on Reference Rate Reform (MPG). We have had particularly useful conversations with Matteo Aquilina, Terry Belton, Tim Bowler, David Bowman, Finbarr Hutchison, Paul Milgrom, Anthony Murphy, Holger Neuhaus, PierMario Satta, Roberto Schiavi, David Skeie, Andy Skrzypacz, Tom Steffen, Jeremy Stein, Kevin Stiroh, James Vickery, Zhe Wang, Victor Westrupp, Nate Wuerffel, Anthony Zhang, and Haoxiang Zhu.

<sup>†</sup>Stanford University, Graduate School of Business, and National Bureau of Economic Research.  
*Email address:* [duffie@stanford.edu](mailto:duffie@stanford.edu).

<sup>‡</sup>Northwestern University, Department of Economics. *Email address:* [piotr.dworczak@northwestern.edu](mailto:piotr.dworczak@northwestern.edu).

## 1 Introduction

This paper solves a problem faced by administrators of financial benchmarks. A benchmark administrator constructs a “fixing,” meaning an estimator of an asset’s market value that is based on reported transaction prices or other data. The data are generated by agents whose profits depend on the realization of the fixing. Agents may therefore misreport or trade at distorted prices in order to manipulate the fixing. We study optimal transaction weights for benchmark fixings, assuming that the benchmark administrator cannot use transfers, such as fees or subsidies. We derive the optimal fixing under a simplifying assumption that weights are unidimensional, and axiomatically characterize the unique benchmark that is robust to a certain form of collusion among traders. Our analysis provides a foundation for two versions of the commonly used volume-weighted average price (VWAP), and delivers insights about the properties of benchmark fixings necessary for ensuring their robustness against manipulation.

Financial benchmarks are prevalent in modern financial markets. Their fixings directly influence the valuation of other assets: Literally millions of different financial contracts, including interest rate swaps, futures, options, variable rate bank loans, and mortgages, have payments that are contractually linked to interest-rate benchmarks.<sup>1</sup> In addition to the role of benchmarks for the purpose of contractual settlement of financial instruments, benchmarks serve an important price transparency function.<sup>2</sup> Benchmarks are also used for evaluating investment performance and as indicators of current conditions in credit and interest-rate markets.

However, these important functions of benchmarks become distorted if the fixings no longer reflect the underlying market fundamentals. Concerns have been raised over the manipulation of benchmarks, including the London Interbank Offered Rate (Libor),<sup>3</sup> foreign exchange fixings, and various commodity benchmarks.<sup>4</sup> These concerns were followed by investigations resulting in billions of dollars of fines. By February 2017, the Commodity Futures Trading Commission, alone, had fined dealers<sup>5</sup> \$5.29 billion for manipulation of Libor, Euribor, foreign exchange benchmarks and the swap rate benchmark known as ISDAFIX. Benchmark manipulation has also been a recent concern for the equity volatility benchmark

---

<sup>1</sup>For example, the aggregate outstanding amount of Libor-linked contracts has been estimated by the [Alternative Reference Rate Committee \(2018\)](#) at \$200 trillion.

<sup>2</sup>See [Duffie, Dworzak and Zhu \(2017\)](#).

<sup>3</sup>For details, see, for example, [Hou and Skeie \(2013\)](#); [Financial Conduct Authority \(2012\)](#); [BIS \(2013\)](#); [Market Participants Group on Reference Rate Reform \(2014\)](#).

<sup>4</sup>See [Financial Stability Board \(2014\)](#).

<sup>5</sup>See “[CFTC Orders The Royal Bank of Scotland to Pay \\$85 Million Penalty for Attempted Manipulation of U.S. Dollar ISDAFIX Benchmark Swap Rates.](#)”

known as VIX,<sup>6</sup> and in the markets for various commodities,<sup>7</sup> precious metals such as gold,<sup>8</sup> and manufactured goods such as pharmaceuticals.<sup>9</sup>

Reports of systematic manipulation of benchmarks have also triggered regulatory reforms. Among other jurisdictions, the [European Union \(2016\)](#) introduced legislation<sup>10</sup> in support of robust benchmarks, which came into force on January 1, 2018. The Financial Stability Board is leading an ongoing global process to overhaul key reference rate and foreign currency benchmarks with a view to improving their robustness to manipulation. A key principle of [International Organization of Securities Commissions \(2013\)](#) is that fixings of key benchmarks should be “anchored” in actual market transactions or executable quotations.

The reasons for traders to engage in benchmark manipulations varied depending on the setting. However, the most common incentive was to profit from benchmark-linked positions held by the manipulator. For example, in a typical email uncovered by investigators, a trader at a bank whose reports are used to fix Libor wrote to the Libor rate submitter: “For Monday we are very long 3m cash here in NY and would like setting to be as low as possible...thanks.”<sup>11</sup> Going forward, Libor in various currencies will be replaced by benchmarks that are fixed on the basis of transactions prices or executable quotes. But this does not eliminate the ability or incentive to manipulate. For example, a manipulator can buy an asset at a price above its fair market value to cause an upward distortion of the fixing. The cost of over-paying can be more than offset by the manipulator’s profit on contracts whose settlement value is contractually linked to the benchmark.

This paper has a theoretical focus. Under restrictive conditions, we focus on the optimal design of a transactions-based weighting scheme. In order to illustrate the problem that we study, we ask the reader to consider the following abstract situation. An econometrician is choosing an efficient estimator of an unknown parameter  $Y$ . Data are generated by strategic

---

<sup>6</sup>See [Griffin and Shams \(2017\)](#).

<sup>7</sup>For cases of oil, natural gas, and propane benchmark manipulation, see “[Federal Court Orders \\$13 Million Fine in CFTC Crude Oil Manipulation Action against Parnon Energy Inc., Arcadia Petroleum Ltd., and Arcadia Energy \(Suisse\) SA, and Crude Oil Traders James Dyer and Nicholas Wildgoose,](#)” CFTC, August 4, 2014; “[CFTC Files and Settles Charges against Total Gas & Power North America, Inc. and Therese Tran for Attempted Manipulation of Natural Gas Monthly Index Settlement Prices,](#)” CFTC, December 7, 2015; and “[CFTC Finds Statoil ASA Attempted to Manipulate the Argus Far East Index, a Propane Benchmark, to Benefit Statoil’s NYMEX-cleared Swaps Position,](#)” CFTC, November 14, 2017.

<sup>8</sup>See [Vaughn \(2014\)](#) and “[How a Barclays’ options trader manipulated the gold price fix,](#)” Reuters, May 23, 2014.

<sup>9</sup>See [Gencarelli \(2002\)](#).

<sup>10</sup>[Financial Conduct Authority \(2016\)](#) explains how the EU regulation “aims to ensure benchmarks are robust and reliable, and to minimise conflicts of interest in benchmark-setting processes.”

<sup>11</sup>December 14, 2006, Trader in New York to Submitter; source: [Malloch and Mamorsky \(2013\)](#). Another example: “We have another big fixing tom[orrow] and with the market move I was hoping we could set the 1M and 3M Libors as high as possible.”

agents whose utilities depend on the realized outcome of the estimator  $\hat{Y}$ . Thus, the chosen estimator influences the data generating process. This game-theoretic aspect of the problem must be considered in the design of the estimator.

Our model features a benchmark administrator who acts as a mechanism designer. The designer chooses a benchmark fixing  $\hat{Y}$  that is an unbiased estimator of  $Y$  and minimizes the mean squared error of the estimation. The agents that might manipulate the benchmark could be banks, broker-dealers, asset-management firms, or individual traders within any of these types of firms. The mechanism designer observes the transactions: The data generated by each transaction consist only of the price and size (the notional amount) of the transaction. Whether or not manipulated, the transactions prices are noisy signals of the fundamental value  $Y$ . For non-manipulated transactions, noise arises from market microstructure effects, as explained by [Ait-Sahalia and Yu \(2009\)](#), and also from asynchronous reporting.<sup>12</sup>

In our model, the benchmark administrator is restricted to fixings  $\hat{Y}$  that are linear with respect to observed transactions prices, with weighting coefficients that can depend on the sizes of the transaction. If the transaction weights are required to add up to one, we call the associated benchmark a *fractional-weight* fixing. A commonly used fractional-weight fixing is the “volume weighted average price” (VWAP), for which the weight applied to a given transaction price is the size of that transaction normalized by the total of the sizes of all observed transactions. VWAP fixings are used, for example, for the settlement of futures contracts on the Chicago Mercantile Exchange.<sup>13</sup> We will characterize optimality for both fractional-weight fixings and for absolute-size-weight fixings, for which the weight on any given transaction depends only on its own absolute size, and not on the sizes of other transactions.

Agents have private information about their exposures to the benchmark. If the agent decides not to manipulate, she transacts in the market to fulfill “legitimate” business purposes (such as hedging, investment, or market making). However, the agent can choose to manipulate by conducting a transaction with an artificially inflated or reduced price in order to gain from the associated distortion of the benchmark. Manipulation is assumed to be costly. For example, in order to cause an upward distortion in the benchmark, a trader could buy the underlying asset at a price above its fair market value. In order to manipulate the price downward, the agent would sell the asset at a price below its true value. Either way, by trading at a distorted price, the agent suffers a loss. On the other hand, the agent may have

---

<sup>12</sup>For example, the WM/Reuters benchmarks for major foreign exchange rates are fixed each day based on transactions that occur within 5 minutes of 4:00pm London time. In an over-the-counter market, moreover, each pair of transacting counterparties is generally unaware of the prices at which other pairs of counterparties are negotiating trades at around the same time.

<sup>13</sup>See [Quick Facts on Settlements at CME Group](#), CME Group, October, 2014). For the [NYMEX crude oil futures contract](#), “If a trade(s) occurs on Globex between 14:28:00 and 14:30:00 ET, the active month settles to the volume-weighted average price (VWAP), rounded to the nearest tradable tick.”

pre-existing contracts, such as derivatives, whose settlement value is linked to the benchmark. On a large pre-existing derivatives position, the agent may be able to distort the benchmark enough to generate a profit that exceeds the cost of creating the fixing distortion.

This suggests the benefit of avoiding benchmarks whose underlying asset market is thinly traded relative to the market for financial instruments that are contractually linked to the benchmark. Our model implies that manipulation is unavoidable when the potential benefits from manipulation, measured by the monetary gain from changing the fixing by one unit, are large relative to (i) the cost of manipulation, (ii) the average size of the transaction, and (iii) the number of transactions in the market for the benchmark asset. As emphasized by [Duffie and Stein \(2015\)](#), when the volume of transactions in an underlying market that determines a fixing is tiny by comparison with the volume of derivatives contracts whose settlements are contractually linked to that fixing, there is a stronger incentive to manipulate. However, our framework makes clear that even if a benchmark could be found that induces only honest (unmanipulated) transactions, this is not necessarily optimal from the viewpoint of the efficiency of the estimator, because the transaction weights that are optimal for statistical efficiency can be quite different from those minimizing the incentive to manipulate.

Our formal analysis consists of two main parts. First, we solve for the optimal benchmark fixing among absolute-size-weight fixings using optimal control techniques. Second, because optimization is not longer tractable in more general classes of fixings, we axiomatically characterize a unique class of fractional-size-weight fixings that prevent traders from engaging in a particular form of collusion.

Our main findings for absolute-size-weight fixings, analyzed in [Section 4](#), are the following. First, the optimal benchmark must put nearly zero weight on small transactions. This is intuitive, and stems from the fact that it is cheap for agents to make small manipulated transactions. For instance, [Scheck and Gross \(2013\)](#) describe a strategy said to be used by oil traders to manipulate the daily oil price benchmark published by Platts: “Offer to sell a small amount at a loss to drive down published oil prices, then snap up shiploads at the lower price.” Second, although the optimal transaction weight is always non-decreasing in the size of a transaction, the optimal benchmark assigns nearly equal weight to all large transactions. This follows from the fact that the optimal weighting function is concave in size, with a slope that goes to zero as trade sizes become large. In many cases, the optimal weight is actually constant above some threshold transaction size. This avoids overweighting transactions made by agents with particularly strong incentives to manipulate. Third, our main result characterizes the exact shape of the optimal weighting function, as a solution to a certain second-order differential equation. In the examples that we study, this optimal shape is well approximated by a weighting function that is linear in size up to a threshold, and constant afterwards.

In our baseline model, we assume that each trader conducts one transaction, or that the trader’s multiple transactions are aggregated for purposes of the fixing. In Section 4.2, we relax this assumption, and allow traders to split their total desired transaction into smaller trades. For many benchmarks, such as the VIX, LBMA Gold Price, CME futures and WMR foreign exchange fixings, the underlying data are anonymous with respect to the firm or trader, thus preventing the benchmark administrator from detecting or deterring order splitting. In such settings, we show that the optimal fixing is the absolute-volume weighted average price (A-VWAP).

In Section 5, we turn our attention to fractional-size-weight fixings. We observe that because benchmark weights in this class are allowed to depend on *all* transaction sizes, agents have incentives to collude by sharing information or coordinating their trades. A simple collusive strategy for two traders is to split their total volume of trade in a way that maximizes their overall impact on the benchmark fixing, effectively by having one agent execute trades on the other’s behalf. Because this strategy does not affect the total trade volume (and hence prices) in the market, it is difficult to detect. A fixing that makes this collusive strategy unprofitable is said to “neutralize pairwise order splitting.” Remarkably, it turns out that this property, along with the weak condition requiring that positive weight is only assigned to transactions with non-negligible size, pins down a unique fixing within the class of fractional-size-weight fixings, namely, the commonly used volume-weighted average price (VWAP).

Overall, our theoretical results give support to both the commonly used VWAP benchmark, and also to the variant, A-VWAP, that is based on absolute size weights. In Section 6, we discuss the trade-offs associated with choosing between fixings based on absolute and relative size weights. Absolute weighting avoids certain types of collusion by inducing a dominant-strategy equilibrium between the agents. However, when weights depend only on absolute transaction sizes, they suffer from at least two drawbacks. First, their mean squared error is inflated by a term that depends on the variance of the fundamental asset value  $Y$ , which could be avoided by switching to a fractional-weighting scheme. Second, absolute size weighting can lead to benchmark values that are outside the range of transaction prices.<sup>14</sup> This cannot happen with fractional weights. Under conditions, we show that the A-VWAP benchmark actually admits an equivalent fractional representation, mitigating some of its undesirable properties. On the other hand, fractional-weight fixings, including VWAP, create incentives for collusive information sharing among agents, and also increase the component of the mean squared error arising from manipulation. Ultimately, whether the VWAP or the A-VWAP fixing is preferred might depend on market characteristics such as the number of traders, variability of day-to-day trading volume, and scope for collusion.

In addition to the choice of the benchmark asset and the fixing design, regulators can

---

<sup>14</sup>We are grateful to a referee for noting this.

implement a range of governance and compliance safeguards, raising the cost of manipulation, consistent with the suggestions of [Financial Conduct Authority \(2012\)](#) and the [International Organization of Securities Commissions \(2013\)](#). Our setting allows for an extra cost for trading at a price away from the fair value, associated with the risk of detection of manipulation by the authorities, and resulting penalties or loss of reputation.

For our theoretical analysis, we assume that the mechanism designer cannot use transfers. In particular, fines or litigation damages, forms of negative transfer, may affect the cost of manipulation exogenously but cannot be actively controlled by a benchmark administrator. Building on our framework, [Coulter, Shapiro and Zimmerman \(2018\)](#) address the optimal design of fines in a “revealed preference mechanism” that directly elicits private information from the agents. Because [Coulter \*et al.\* \(2018\)](#) do not study the problem of designing an optimal fixing, our approaches are complementary.

[Baldauf, Frei and Mollner \(2020\)](#) analyze a contracting problem between a client (principal) who needs to purchase a fixed quantity of an asset and a dealer (agent). The contract references a benchmark whose value can be computed from commonly observed market prices and volumes. The dealer can manipulate the benchmark using the fact that her trades are not directly observed by the client. [Baldauf \*et al.\* \(2020\)](#) derive the optimal (cost-minimizing) contract for the client and show that—under conditions—it references a VWAP benchmark. Thus, our work and theirs are complementary, both giving theoretical support for a VWAP benchmark but in different contexts and under a different objective function.

[Zhang \(2020\)](#) studies a similar manipulation problem of financial benchmarks but identifies an alternative approach to regulating them based on the measurement of manipulation incentives. He defines two measures of manipulation-induced welfare losses, and shows that they can be estimated from the commonly observed market data. His empirical framework can be used to estimate the manipulation-induced distortions that would arise under hypothetical benchmark designs, casting light on their optimal design.

Our work falls into a growing literature on mechanism design without transfers. This body of research, however, typically focuses on allocation problems (for example [Ben-Porath, Dekel and Lipman, 2014](#) and [Mylovanov and Zapechelnyuk, 2017](#)). The techniques we use are reminiscent of those used to study direct revelation mechanisms and, to some degree, principal-agent models. There are, however, essential differences. Because of the restriction on the class of mechanisms (linear estimators), we cannot rely on the Revelation Principle. The objective function is not typical. Our mechanism designer is minimizing the mean squared error of the estimator (benchmark). Agents face a cost of misreporting their type which is proportional to the deviation from the true type.<sup>15</sup> Overall, we are forced to develop new

---

<sup>15</sup>[Lacker and Weinberg \(1989\)](#) analyze a model of an exchange economy where an agent may falsify public information at a cost; [Kartik \(2009\)](#) studies a cheap talk game in which the Sender pays a cost

techniques that draw on tools from optimal control theory.

We do not analyze estimators that assign different weights to transactions based on the transactions prices themselves (that is, nonlinear estimators). In particular, our methods do not allow for trimming outlier price reports or trades. Some benchmarks dampen or eliminate the influence of outlier prices. Related to this, [Eisl, Jankowitsch and Subrahmanyam \(2014\)](#) and [Youle \(2014\)](#) argue that the median estimator can reduce the incentive to manipulate. However, the net effect on the statistical efficiency of a median-based fixing as an estimator of the underlying market value is unknown in a setting such as ours with strategic data generation. We show in [Appendix A](#) that, at least in our model, the median-based fixing does not fundamentally alter manipulation incentives relative to linear benchmarks.

The remainder of the paper is organized as follows. [Section 2](#) introduces the primitives of the model and the solution concept. [Section 3](#) offers some preliminary results that apply to a general class of linear fixings. [Section 4](#) contains our main analysis of absolute-size-weight fixings, while [Section 5](#) deals with fractional-size-weight fixings. We conclude in [Section 6](#) by discussing the trade-offs associated with the choice between the two benchmark designs that emerge from the analysis of the preceding sections. [Appendix A](#) discusses non-linear fixings, [Appendix B](#) presents two models of manipulation that micro-found our reduced-form baseline framework, while [Appendix C](#) contains additional examples. Finally, most proofs are relegated to [Appendix D](#).

## 2 The baseline model

A mechanism designer (benchmark administrator) estimates an uncertain variable  $Y$ , which can be viewed as the “true” market value of an asset (the “fundamental”). To this end, she designs a benchmark fixing, which is an estimator  $\hat{Y}$  that can depend on the transaction data  $\left\{(\hat{X}_i, \hat{s}_i)\right\}_{i=1}^n$  generated by a fixed set  $\{1, 2, \dots, n\}$  of agents. Here,  $\hat{X}_i$  is the price and  $\hat{s}_i$  is the quantity of the transaction of agent  $i$ . The size  $\hat{s}_i$  of each transaction is restricted to  $[0, \bar{s}]$ , a technical simplification that could be motivated as a risk limit imposed by a market regulator or by an agent’s available capital. The price  $\hat{X}_i$  is a noisy or manipulated signal of  $Y$ , in a sense to be defined. Agents are strategic: they have preferences, to be explained, over their respective transactions and over the benchmark  $\hat{Y}$ . The sensitivity of a given agent’s utility to  $\hat{Y}$  is known only to that agent.

We next describe in detail the problem of the benchmark administrator and the agents. Further interpretation of our assumptions is postponed to the end of the section.

---

for deviating from the truth; [Kephart and Conitzer \(2016\)](#) formulate a Revelation Principle for a class of models in which the agent faces a reporting cost.



## 2.1 The problem of the benchmark administrator

The benchmark administrator minimizes the mean squared error  $\mathbb{E}[(Y - \hat{Y})^2]$  of a benchmark fixing  $\hat{Y}$  that is a linear unbiased estimator of the fair value  $Y$  of the form

$$\hat{Y} = \sum_{i=1}^n f_i(\hat{\mathbf{s}}) \hat{X}_i, \quad (2.1)$$

where  $\hat{\mathbf{s}} = (\hat{s}_1, \dots, \hat{s}_n)$ , and for each  $i$ ,  $f_i : [0, \bar{s}]^n \rightarrow \mathbb{R}^+$  is a measurable transaction weighting function. We will consider two special classes:

1. **Absolute-size-weight fixings**, for which  $f_i(\hat{\mathbf{s}}) = w_i(\hat{s}_i)$ , where, for each  $i$ ,  $w_i : [0, \bar{s}] \rightarrow \mathbb{R}^+$  is a transaction weighting function to be chosen. That is, the weight placed on a given transaction depends only on its own size.
2. **Fractional-weight fixings**, those for which

$$\sum_{i=1}^n f_i(\hat{\mathbf{s}}) = 1,$$

for all  $\hat{\mathbf{s}} \in [0, \bar{s}]^n$ . For example, the volume-weighted average price (VWAP),  $f_i(\hat{\mathbf{s}}) = \hat{s}_i / (\sum_{j=1}^n \hat{s}_j)$ , is a fractional-weight fixing.

## 2.2 The problem of the agents

We now explain how the transaction data  $\{(\hat{X}_i, \hat{s}_i)\}_{i=1}^n$  are generated by strategic agents. We assume that an agent can conduct a manipulated transaction at some reduced-form net benefit, without explicitly modeling the market in which the transaction takes place. In Appendix B, we propose two alternative stylized models of market trading that endogenize the reduced-form costs and benefits of a manipulator.

Agent  $i$  privately observes her type  $R_i$ , which is interpreted as the agent's profit exposure to the benchmark. Specifically, the agent's payoff includes a profit component  $R_i \hat{Y}$ . This type  $R_i$  can be negative, corresponding to cases when the agent holds a short position in the asset whose value is positively correlated with the benchmark.

Having observed  $R_i$ , the agent chooses a pair  $(\hat{z}_i, \hat{s}_i) \in \{-z_i, 0, z_i\} \times [0, \bar{s}]$ , where  $\hat{z}_i$  is a price distortion and  $\hat{s}_i$  is a trade size. The absolute magnitude  $z_i > 0$  of the price distortion is assumed to be exogenous, and can be a random variable, as discussed in Appendix B. If agent  $i$  chooses not to manipulate, in that  $\hat{z}_i = 0$ , the benchmark administrator observes  $(\hat{X}_i, \hat{s}_i) = (X_i, s_i)$ , where  $(X_i, s_i)$  can be thought of the transaction (price and quantity)

of agent  $i$  that would be naturally preferred in the absence of manipulation incentives. The transaction  $(X_i, s_i)$  is determined by hedging or speculative motives that we do not model. That is, we take the distribution of  $(X_i, s_i)$  as given. On the other hand, if agent  $i$  chooses to manipulate, in that  $\hat{z}_i \neq 0$ , then the benchmark administrator observes the transaction  $(\hat{X}_i, \hat{s}_i) = (X_i + \hat{z}_i, \hat{s}_i)$ .

Intuitively, the agent can trade some quantity  $\hat{s}_i$  at a distorted price to manipulate the benchmark fixing. This substitution, however, induces a cost  $\gamma \hat{s}_i |X_i - \hat{X}_i|$  to the agent that is proportional to the size of the transaction and to the deviation of the price from the market level  $X_i$ , where  $\gamma > 0$  is a fixed parameter.

Without loss of generality, we normalize to zero the payoff to the agent associated with the truthful reporting choice  $(\hat{X}_i, \hat{s}_i) = (X_i, s_i)$ . Therefore, agent  $i$  solves the following problem

$$\max_{\hat{z}_i \in \{-z_i, 0, z_i\}, \hat{s}_i \in [0, \bar{s}]} \left[ R_i \mathbb{E} \left[ \Delta_{\hat{z}_i, \hat{s}_i} \hat{Y} \right] - \gamma \hat{s}_i |\hat{z}_i| \right] \mathbf{1}_{\{\hat{z}_i \neq 0\}}, \quad (\mathcal{A})$$

where  $\mathbb{E} \left[ \Delta_{\hat{z}_i, \hat{s}_i} \hat{Y} \right]$  is the equilibrium expected manipulative impact on the benchmark fixing relative to choosing  $\hat{z}_i = 0$  and  $\hat{s}_i = s_i$ . (Thus,  $\Delta_{0, s_i} \hat{Y} = 0$ .) We assume that agents play a Bayes Nash equilibrium. Formally, each agent  $i$  evaluates  $\mathbb{E} \left[ \Delta_{\hat{z}_i, \hat{s}_i} \hat{Y} \right]$  assuming equilibrium strategies for other agents, taking expectations over their unobserved characteristics  $(R_j, s_j, X_j)_{j \neq i}$ . To simplify the problem faced by the agent, we further assume that the transaction price  $X_i$  is unobserved at the time of choosing whether to manipulate or not. (We discuss and motivate this assumption in Subsection 2.4.)

### 2.3 The distribution of transactions data

The unmanipulated transactions  $\{(X_i, s_i)\}_{i=1}^n$  are generated as follows. First,  $Y$  is drawn from some probability distribution with mean normalized to zero, and with some finite variance  $\sigma_Y^2$ . Then, a pair  $(\epsilon_i, s_i)$  is drawn for every agent, *i.i.d.* across agents and independently of  $Y$ . We assume that  $\mathbb{E}(\epsilon_i | s_i) = 0$ , and that  $\text{var}(\epsilon_i | s_i) = \sigma_\epsilon^2$  for some  $\sigma_\epsilon^2 > 0$ . The size  $s_i$  has a cumulative distribution function (cdf)  $G$  with a continuous density  $g$  that is strictly positive on  $[0, \bar{s}]$ . The unmanipulated price is  $X_i = Y + \epsilon_i$ , which is therefore a noisy and unbiased signal of  $Y$ , with variance  $\sigma_U^2 \equiv \sigma_Y^2 + \sigma_\epsilon^2$ . The subscript  $U$  is a mnemonic for “unmanipulated.”

The exposure types  $R_1, \dots, R_n$  are *i.i.d.* and independent of all other primitive random variables in the model. The cdf  $\tilde{H}$  of  $R_i$  has support contained by some interval  $[-\bar{R}, \bar{R}]$ . We allow the case of  $\bar{R} = \infty$ . We assume that the probability distribution of  $R_i$  is symmetric around zero, and that bigger incentives to manipulate are relatively less likely to occur than smaller incentives. That is,  $\tilde{H}$  has a finite variance and a density  $\tilde{h}$  that is symmetric around zero and strictly decreasing on  $(0, \bar{R})$ . Examples include the normal and Laplace (“double

exponential”) distributions. Given the symmetry of  $\tilde{H}$ , we can define a cdf  $H$  on  $[0, \bar{R}]$  such that

$$\tilde{H}(R) = \begin{cases} \frac{1}{2} - \frac{1}{2}H(-R) & \text{if } R < 0, \\ \frac{1}{2} + \frac{1}{2}H(R) & \text{if } R \geq 0. \end{cases}$$

That is,  $H$  is the distribution of  $R_i$  conditional on  $R_i \geq 0$ . We let  $h$  denote the density of  $H$ , and we assume that  $h$  is twice continuously differentiable.

The agents’ respective price-distortion magnitudes  $z_1, \dots, z_n$  are *i.i.d.*, and are independent of all other primitive model variables. The variance  $\sigma_z^2$  of  $z_i$  is finite and strictly positive. We let  $\sigma_M^2 \equiv \sigma_U^2 + \sigma_z^2 > \sigma_U^2$  denote the variance of the reported price  $X_i + \hat{z}_i$  conditional on the event  $\{\hat{z}_i \neq 0\}$  of a manipulation. The subscript  $M$  is thus a mnemonic for “manipulated.”

## 2.4 Comments on assumptions

For tractability, we restrict attention to benchmark fixings that are linear with respect to price, with weights depending only on the sizes of transactions. We do not consider nonlinear benchmarks, that is, fixings whose weights depend jointly on the sizes and prices of transactions. We do, however, discuss nonlinear designs in Appendix A.

For absolute-size-weight fixings, a trader’s problem is particularly simple, since her weight in the fixing is independent of the choices of other traders. In contrast, for general fixings such as VWAP, for which the weight placed on a given agent’s transaction may depend on the sizes of the transactions of other agents, there can be non-trivial strategic interaction, because the optimal manipulation strategy of an agent may depend on how other agents behave. As a result, some fixings create natural scope for collusion. Because we are not able to accommodate all classes of benchmark fixings in a single optimization framework, we instead proceed in three steps. First, in Section 4, we solve for the optimal absolute-size-weight fixing using optimal control techniques. Second, we adopt an axiomatic approach in Section 5, where we characterize the unique class of fixings that are robust to certain types of collusive behaviors among traders. Third, in Section 6, we compare the optimal designs that emerge from the analysis of Section 4 and 5, and explain when using one versus the other is preferred.

The commonly used volume-weighted average price (VWAP) benchmark fixing method uses the relative size weight  $\hat{s}_i / (\sum_{j=1}^n \hat{s}_j)$  and is thus a special case of a fractional-weight fixing. In the class of absolute-size-weight fixings, we can accommodate a version of the VWAP benchmark, in which  $\sum_{j=1}^n \hat{s}_j$  in the denominator of the VWAP benchmark is replaced with its expectation (under equilibrium strategies). We will refer to this fixing as absolute-VWAP or “A-VWAP.”

We have assumed that the variance of the price noise  $\epsilon_i$  associated with an unmanipulated

transaction has a variance that does not depend on the size of the transaction. It would be more realistic to allow the price precision to be increasing with the size of the transaction, as implicitly supported by volume-weighted-average-price (VWAP) schemes often used to report representative prices in financial markets.<sup>16</sup> Let  $\kappa(s_i) = \text{var}(X_i | s_i)^{-1}$  denote the precision of the unmanipulated price  $X_i$  conditional on the transaction size  $s_i$ . Focusing on the case of a constant  $\kappa(s)$  allows us to greatly simplify our arguments and sharpen the results. However, as shown in a preliminary version of this paper (Duffie and Dworczak, 2014), our qualitative conclusions remain valid provided that  $\kappa(s)$  is a non-decreasing and concave function, and that  $\sigma_z^2$  is large enough that a manipulated transaction has more price noise than an unmanipulated transaction, regardless of its size.

The problem faced by each agent is stylized. We aim to capture some of a manipulator’s key incentives. The trader’s type  $(X_i, s_i)$  can be interpreted as the transaction that the agent would make, given current market conditions, to fulfill her usual “legitimate” business purposes. For example, such a trade could be the result of a natural speculative, market making, or hedging motive. The assumption that each agent can make only one transaction is relaxed in Section 4.2. Formally, this assumption is justified if all of the transactions of a given agent are first aggregated and only then provided as a single input to the estimator. This is the method currently used in the fixing of Libor by ICE Benchmark Administrator. Section 4.2 considers the problem of a benchmark administrator when such an aggregation is infeasible or undesirable.

The assumption that  $X_i$  is unobserved at the time of solving problem  $\mathcal{A}$  means that the agent commits to a direction and magnitude of manipulation before executing the trading strategy. For example, manipulation can often be generated through the effect of “price impact.” A manipulator who wishes to create an upward distortion of the price could submit a large market order to buy to a limit-order-book exchange. The buy order would “travel up” the limit order book to the price level at which it is filled, which cannot be predicted when the order is submitted. More generally, it is typically difficult to predict one’s execution price in a financial market at the point in time at which the trading decision is made (see Appendix B for a formal model supporting this claim). Alternatively, this assumption can be micro-founded via an agency friction. Suppose that there are two traders at bank  $i$ : Trader  $A$  is responsible for trading the underlying asset, while trader  $B$  operates in the derivative market. It is then trader  $B$  who observes  $R_i$  and stands to gain from the manipulation. The literal interpretation of our model is that trader  $B$  submits a request to trader  $A$  with the recommended size and price distortion  $(\hat{z}_i, \hat{s}_i)$ , and then trader  $A$  observes  $(X_i, s_i)$  and executes the trade  $(\hat{X}_i, \hat{s}_i)$  in the market for the underlying asset.<sup>17</sup>

<sup>16</sup>See, for example, Berkowitz *et al.* (1988).

<sup>17</sup>While trader  $B$  would in general benefit from conditioning her manipulation strategy on the

For simplicity, we have also assumed that the size of a price manipulation is bounded by  $z_i$ . Then, the restriction to the choice from  $\{-z_i, 0, z_i\}$  is without loss of generality given the linearity of the agent’s objective function in  $\hat{z}_i$ . Alternatively, we could assume that there is an increasing cost  $\psi(|z|)$  of manipulation, based for instance on an increasing probability of detection. Formally, in our setting,  $\psi(|z|) = c\mathbf{1}_{\{z \notin [-z_i, z_i]\}}$ , for some large  $c > 0$ . The results depend mainly, in this regard, on the assumption that the manipulation levels chosen by agents are high enough that manipulated transactions are less precise signals of price than unmanipulated transactions. This property would hold across many plausible alternative model specifications.

The cost of manipulation reflects the losses that the trader incurs when trading away from fair-value prices in order to manipulate the fixing. We take a partial-equilibrium approach, relegating an endogenous model of trading and payoffs to Appendix B. Our particular functional form for the cost of manipulation, chosen in large part for its tractability, can be further justified by an alternative interpretation of the nature of manipulations. Namely, imagine that traders can submit “shill trades,” in the form of fictitious transactions at distorted prices, with reimbursements, “kickbacks,” arranged through side payments. Then  $\hat{s}_i|X_i - \hat{X}_i|$  is precisely the kickback cost of manipulation.<sup>18</sup> Assuming that the cost of manipulation is linear, rather than strictly convex, with respect to size and price distortion is a conservative approach in that it allows for relatively higher profits associated with larger manipulations.

Finally,  $R_i$  can be thought of as the position that the trader holds in contracts whose settlement price is tied to the administered benchmark fixing. For example, manipulators of oil price benchmarks (Footnote 7) were allegedly motivated by the fact that they held oil futures contracts whose settlement payments are contractually based on the oil price fixing. For positions such as options whose market values are nonlinear with respect to a benchmark, one can view  $R_i$  as the so-called “delta” (first-order) sensitivity of the position value to the benchmark. The assumption that  $R_i$  is symmetric around zero is, in effect, a belief by the benchmark administrator that upward and downward manipulative incentives are similar, other than with respect to their signs.

### 3 Statistical efficiency versus manipulation deterrence

This section considers basic properties of a general benchmark fixing  $\hat{Y} = \sum_i f_i(\hat{\mathbf{s}})\hat{X}_i$ . We present solutions to some preliminary cases that provide intuition as well as elements on which

---

current market information about prices  $X_i$ , such back-and-forth communication between traders might be too risky. Indeed, Libor litigation and fines have mostly arisen when such private communication between traders was uncovered—see footnote 11.

<sup>18</sup>This assumption is that the cost is linear in size. We can view  $\gamma$  as a per-dollar cost of using an illegal transfer channel, for example resulting from the possibility of detection and punishment.

to build when solving the general case.

### 3.1 Solution without manipulation

For comparison purposes, we first solve the problem assuming that agents do not manipulate. That is, for each  $i$ ,  $\hat{X}_i = X_i$  and  $\hat{s}_i = s_i$ . The law of iterated expectation implies that

$$\mathbb{E}[\hat{Y}|Y] = Y\mathbb{E}\left[\sum_{i=1}^n f_i(\mathbf{s})\right]. \quad (3.1)$$

Thus,  $\hat{Y}$  is unbiased if and only if  $\mathbb{E}\sum_{i=1}^n f_i(\mathbf{s}) = 1$ . Then, it follows from our distributional assumptions that

$$\mathbb{E}\left[(Y - \hat{Y})^2\right] = \left(\mathbb{E}\left[\left(\sum_{i=1}^n f_i(\mathbf{s})\right)^2\right] - 1\right)\sigma_Y^2 + \mathbb{E}\left[\sum_{i=1}^n f_i^2(\mathbf{s})\right]\sigma_\epsilon^2.$$

**Proposition 1** *Absent manipulation, the weighting function  $f$  that minimizes the mean squared error of the benchmark fixing subject to unbiasedness is given by  $f_i(\mathbf{s}) = 1/n$ .*

The proof is skipped: This problem can be viewed as a simple case of ordinary-least-squares estimation. The benchmark administrator's optimal weights are proportional to the precision of each price observation. Because the precisions are assumed to be identical and in particular invariant to the sizes of transactions, the optimal weights are equal. There is an obvious generalized-least-squares extension to the case of a general covariance structure for the observation "noises"  $\epsilon_1, \dots, \epsilon_n$ .

### 3.2 Manipulation deterrence

We now turn to the manipulation problem  $\mathcal{A}$  facing an agent. By symmetry, we may concentrate on the event of a positive manipulation incentive,  $R_i \geq 0$ . Using the assumptions of Subsections 2.2 and 2.3, we can express the problem  $\mathcal{A}$  of agent  $i$  as

$$\max_{\hat{z}_i \in \{0, z_i\}, \hat{s}_i \in [0, \bar{s}]} [R_i \bar{f}_i(\hat{s}_i) - \gamma \hat{s}_i] \hat{z}_i \mathbf{1}_{\{\hat{z}_i \neq 0\}}, \quad (3.2)$$

where  $\bar{f}_i(\hat{s}_i) = \mathbb{E}[f_i(\hat{s}_i, \hat{\mathbf{s}}_{-i})]$  is the interim expected weight for agent  $i$ 's transaction with expectation taken over the vector of transaction sizes  $\hat{\mathbf{s}}_{-i}$  of agents other than  $i$ . In particular, we have used the assumption that, from the perspective of agent  $i$ ,  $\hat{X}_j$  has zero mean, and

thus agent  $i$  can ignore the influence of her  $\hat{s}_i$  on the weight on agent  $j$ 's transaction, for any  $j \neq i$ .

An agent with type  $R_i$  manipulates if and only if there is some  $s \in [0, \bar{s}]$  such that  $R_i \bar{f}_i(s) > \gamma s$ , that is, if there is a size for the manipulated trade at which the impact on the benchmark fixing is high enough to cover the associated manipulation cost. If type  $R_i$  chooses to manipulate, then all types higher than  $R_i$  also manipulate. Similarly, if type  $R_i$  chooses not to manipulate, all types below  $R_i$  also choose not to manipulate. Moreover, because agents are ex-ante identical, if the designer wants to implement truthful reporting by all agents, she may without loss of generality focus on symmetric weights:  $f_i(\mathbf{s}) = f(s_i, \mathbf{s}_{-i})$  for some  $f : [0, \bar{s}]^n \rightarrow \mathbb{R}_+$ . It follows that with any such weighting function  $f$  we may associate a unique threshold  $R_f$  defined by

$$R_f = \sup\{R \leq \bar{R} : R \bar{f}(s) \leq \gamma s, \quad s \in [0, \bar{s}]\},$$

where  $\bar{f}(s) = \mathbb{E}[f(s, \hat{\mathbf{s}}_{-i})]$ . That is, given  $f$ , the types above  $R_f$  manipulate and the types below  $R_f$  do not. This easy observation leads to the following result.

**Proposition 2** *The benchmark administrator can ensure that there are no manipulations if and only if  $\bar{R} \leq n\gamma\mathbb{E}[s_1]$ .*

*Proof:* No manipulation means that  $\hat{s}_i = s_i$  for all  $i$ . By the fact that the benchmark fixing is unbiased, we must have that  $\mathbb{E}\bar{f}(s_1) = 1/n$ . By the above discussion, it is possible to implement no-manipulation (truthful reporting) if and only if, for every  $s \in [0, \bar{s}]$ , we have  $\bar{R}\bar{f}(s) \leq \gamma s$ . Thus, a necessary condition to implement no manipulation is that

$$\bar{R}\mathbb{E}\bar{f}(s_1) \leq \gamma\mathbb{E}[s_1]$$

from which it follows that  $\bar{R} \leq n\gamma\mathbb{E}[s_1]$ . The condition is also sufficient. Indeed, the designer can simply use an absolute-size-weight fixing with weights  $w_1(s_i), \dots, w_n(s_n)$  of the form  $w_i(s_i) = s_i/(n\mathbb{E}[s_1])$  which are unbiased, and do not lead to any manipulation since in this case  $\bar{f} = w_i$  and

$$\bar{R}\bar{f}(s) = \bar{R}\frac{s}{n\mathbb{E}[s_1]} \leq n\gamma\mathbb{E}[s_1]\frac{s}{n\mathbb{E}[s_1]} = \gamma s.$$

□

The result states that implementing truthful reporting may sometimes be possible. However, the condition  $\bar{R} \leq n\gamma\mathbb{E}[s_1]$  is likely to be violated in practice, especially when the underlying asset market for the benchmark is thinly traded relative to the market for instruments that determine the incentives to manipulate. A thinly traded underlying asset market

corresponds to the case in which  $\bar{R}$  is large relative to  $\mathbb{E}[s_1]$ , the expected size of a typical transaction in the benchmark market. This condition is even less likely to be satisfied when manipulation is relatively cheap ( $\gamma$  is small) or when there are few transactions ( $n$  is small). The latter case is indeed a practical concern. For example, as documented by Brundsen (2014) for the case of EURIBOR, banks are increasingly reluctant to support benchmarks in the face of potential regulatory penalties and the risk of private litigation.

We note that the weight used in the proof of Proposition 2 differs significantly from the optimal weight used when agents do not manipulate from Proposition 1. This is a reflection of the fundamental trade-off that our designer faces between statistical efficiency (which dictates using constant weights) and manipulation deterrence (which requires weights that are small for small transactions). In particular, this means that the benchmark administrator *should not* restrict attention to weighting functions that fully deter manipulation, even when this is feasible.<sup>19</sup> The optimal weighting function instead influences the *degree* to which agents manipulate. At the same time, it is never optimal to ignore the manipulation problem altogether by using the statistically efficient fixing from Proposition 1. The following simple observation illustrates this fact.

**Observation 1** *If the weight chosen by the designer satisfies  $\limsup_{s_k \rightarrow 0} \bar{f}(s_k) > 0$ , then all agents choose to manipulate with probability 1.*

Observation 1 is a straightforward consequence of the assumption that it is much less costly to manipulate small transactions. As a result, if the fixing attaches a strictly positive weight to arbitrarily small transactions, each agent will always choose to bias the price in her favor, no matter how small their incentive to manipulate  $R_i$ . Because manipulated transactions are less precise signals of the fundamental  $Y$ , such a fixing cannot be optimal. The designer could set a zero weight for transaction sizes below some arbitrarily small threshold, and lower the mean squared error by improving the precision of signals with strictly positive probability (since agents with sufficiently low  $R_i$  would no longer choose to manipulate).

For clarity of exposition, we henceforth assume that  $\bar{R} = \infty$ , so that implementing truthful reporting is not possible.<sup>20</sup>

## 4 The optimal absolute-size-weight fixing

In this section, we work with the class of absolute-size-weight fixings, and derive the optimal benchmark using optimal control techniques. All proofs are relegated to the appendix.

<sup>19</sup>See an earlier draft of our paper Duffie and Dworczak (2014) for a formal numerical example illustrating this point.

<sup>20</sup>Mathematically, this is essentially without loss of generality because we can specify the cdf  $H$  of  $R_i$  to place arbitrarily small probability mass above any finite  $\bar{R}$ .



Because agents are ex-ante identical, it is without loss of generality to restrict attention to symmetric weights. We will abuse notation slightly by using  $f : [0, \bar{s}] \rightarrow \mathbb{R}_+$  to denote the one-dimensional weight function in this section. That is, we study benchmarks of the form  $\hat{Y} = \sum_{i=1}^n f(\hat{s}_i) \hat{X}_i$ .

We impose a mild regularity condition that is needed to ensure the existence of a solution to the administrator's problem. Let  $\mathcal{C}^{K,M}$  be the set of upper semi-continuous  $f : [0, \bar{s}] \rightarrow \mathbb{R}_+$  with the property that there exist at most  $K$  points  $0 = s_1 < s_2 < \dots < s_{K-1} < s_K = \bar{s}$  such that  $f$  is Lipschitz continuous with Lipschitz constant  $M$  in each  $(s_i, s_{i+1})$ . The constants  $K$  and  $M$  are assumed to be finite but large.<sup>21</sup> This regularity allows the weighting function  $f$  to have finitely many jump discontinuities.

We summarize the problem of the benchmark administrator as

$$\inf_{f \in \mathcal{C}^{K,M}} \mathbb{E} \left[ \left( Y - \sum_{i=1}^n f(\hat{s}_i) \hat{X}_i \right)^2 \right] \quad \text{subject to} \quad \mathbb{E} \left[ \sum_{i=1}^n f(\hat{s}_i) \right] = 1. \quad (\mathcal{P})$$

We now derive a concise mathematical formulation of the problem  $\mathcal{P}$  faced by the benchmark administrator under the assumptions of Subsections 2.2 and 2.3.

First, we use our symmetry assumptions to simplify the problem. From the viewpoint of the benchmark administrator, the events  $\hat{z}_i = z_i$  and  $\hat{z}_i = -z_i$  are equally likely, even after conditioning on  $\hat{s}_i$ . Therefore, equation (3.1) still holds if we replace  $s_i$  by  $\hat{s}_i$ . That is, forcing the estimator  $\hat{Y}$  to be unbiased is equivalent to the requirement that  $\mathbb{E}[\sum_{i=1}^n f(\hat{s}_i)] = 1$ . We denote by  $\Psi_f(\cdot)$  the cdf of the transaction size  $\hat{s}_i$ , conditional on its manipulation. That is,

$$\Psi_f(s) = \mathbb{P}_{R_i \sim H}(\operatorname{argmax}_{\hat{s}} R_i f(\hat{s}) - \gamma \hat{s} \leq s \mid R_i > R_f). \quad (4.1)$$

By the law of iterated expectations and because of arguments presented in Subsection 3.2,

$$\mathbb{E} \left[ (Y - \hat{Y})^2 \right] = \sum_{i=1}^n \int_0^{\bar{s}} f^2(\hat{s}_i) \left[ \sigma_U^2 H(R_f) g(\hat{s}_i) d\hat{s}_i + \sigma_M^2 (1 - H(R_f)) d\Psi_f(\hat{s}_i) \right] - \frac{\sigma_Y^2}{n}.$$

The displayed equation states that if  $|R_i| \leq R_f$  (which happens with probability  $H(R_f)$ ), then the transaction of agent  $i$  is unmanipulated, in that  $\hat{s}_i = s_i$ ,  $\hat{X}_i$  has variance  $\sigma_U^2$ , and  $\hat{s}_i$  has probability density  $g$ . On the other hand, if the transaction is manipulated, which happens with probability  $1 - H(R_f)$ , then  $\hat{X}_i$  has variance  $\sigma_M^2$  from the viewpoint of the benchmark administrator.

<sup>21</sup>Formally, all our results hold for large enough  $M$  and  $K$ .

Similarly, we can express the constraint in problem  $\mathcal{P}$  as

$$1 = \mathbb{E} \left[ \sum_{i=1}^n f(\hat{s}_i) \right] = \sum_{i=1}^n \int_0^{\bar{s}} f(\hat{s}_i) [H(R_f)g(\hat{s}_i) d\hat{s}_i + (1 - H(R_f)) d\Psi_f(\hat{s}_i)].$$

To characterize the optimal benchmark, we use an approach familiar from principal-agent models. We address the best way, given some target manipulation threshold  $R$ , for the administrator to implement an outcome in which an agent with  $|R_i| \leq R$  chooses not to manipulate. As we saw before, this requires that the benchmark weight function satisfies the additional constraint  $f(s) \leq (\gamma/R)s$ . Solving this auxiliary problem is a key step towards solving the original problem  $\mathcal{P}$ . This auxiliary problem is illuminating in its own right. For example, the benchmark administrator may have exogenous preferences for deterring manipulation, which could be modeled by setting a high manipulation threshold  $R$ . Formally, using the assumption that agents are symmetric, we can formulate the auxiliary problem as

$$\inf_{f \in \mathcal{C}^{K,M}} \int_0^{\bar{s}} f^2(s) [\sigma_U^2 H(R)g(s) ds + \sigma_M^2 (1 - H(R)) d\Psi_f(s)] \quad (\mathcal{P}(R))$$

subject to<sup>22</sup>

$$f(s) \leq \frac{\gamma}{R}s, \quad s \in [0, \bar{s}], \quad (4.2)$$

$$\int_0^{\bar{s}} f(s) [H(R)g(s) ds + (1 - H(R)) d\Psi_f(s)] = \frac{1}{n}. \quad (4.3)$$

If the target manipulation threshold  $R$  is too high, then no function  $f$  inducing that threshold (that is, satisfying constraint (4.2)) will satisfy constraint (4.3). Among all weighting functions  $f$  satisfying (4.2),  $f(s) = (\gamma/R)s$  maximizes the left hand side of (4.3). In particular, under this transaction weighting all manipulators choose the maximal transaction size  $\bar{s}$ . Therefore, if we define

$$\hat{R} = \max \left\{ R \geq 0 : \frac{\gamma H(R)}{R} \mathbb{E}(s_1) + \frac{\gamma(1 - H(R))}{R} \bar{s} \geq \frac{1}{n} \right\},$$

then the condition  $R \leq \hat{R}$  is both necessary and sufficient for the set of feasible  $f$  to be non-empty for the problem  $\mathcal{P}(R)$ .

## 4.1 The optimal benchmark

In this section we present the solution to the problem faced by the benchmark administrator. Theorem 1a lists the main properties of the optimal benchmark. Theorem 1b describes the exact shape of the optimal fixing under a technical assumption. When this technical assump-

<sup>22</sup>We abuse notation slightly by treating  $\Psi_f(s)$  as being defined by (4.1) with  $R_f$  replaced by  $R$ .

tion fails, the optimal fixing can still be described as a solution to a parameterized differential equation.

**Theorem 1a** *For any  $R \in (0, \hat{R})$ , there exists a unique solution  $f^*$  to problem  $\mathcal{P}(R)$ . Moreover,  $f^*$  is non-decreasing, concave, continuously differentiable, and satisfies  $(f^*)'(\bar{s}) = 0$ . There is some  $s_0 > 0$  such that  $f^*(s)$  coincides with  $(\gamma/R)s$  whenever  $s \leq s_0$ .*

The following ordinary differential equation (ODE), which will play a key role in determining the shape of the optimal fixing, is indexed by two parameters:  $s_0$  and  $s_1$ , with  $0 < s_0 < s_1 < \bar{s}$ . Consider

$$f''(s) = -\frac{[f(s_1) - f(s)]H(R)g(s) + \frac{\sigma_M^2}{\sigma_U^2}\gamma h\left(\frac{\gamma}{f'(s)}\right)}{\left[\frac{\sigma_M^2}{\sigma_U^2}f(s) - f(s_1)\right]\left(-h'\left(\frac{\gamma}{f'(s)}\right)\right)\frac{\gamma^2}{(f'(s))^3}}, \quad (4.4)$$

with boundary conditions

$$f(s_0) = (\gamma/R)s_0, \quad f'(s_0) = (\gamma/R).$$

**Theorem 1b** *Suppose that there exist  $s_0 < s_1 < \bar{s}$  such that  $f^*$ , defined by*

$$f^*(s) = \begin{cases} \frac{\gamma}{R}s, & s \in [0, s_0] \\ \text{solves (4.4),} & s \in (s_0, s_1) \\ f^*(s_1), & s \in [s_1, \bar{s}], \end{cases}$$

*is continuously differentiable and satisfies (4.3). Then, for any  $R \in (0, \hat{R})$ ,  $f^*$  is the unique solution to the optimal fixing problem  $\mathcal{P}(R)$ .*

Intuitively, in Theorem 1b, for any given  $s_0$ , the point  $s_1$  is chosen so that  $f'(s_1) = 0$ . (This is called the “shooting method.”) This construction guarantees that  $f^*$  is continuously differentiable. Then  $s_0$  can be chosen to satisfy (4.3). However, especially when  $\bar{s}$  is relatively small, suitable choices for  $s_0$  and  $s_1$  might not exist. In such a case, the optimal  $f^*$  asymptotes to a constant function without being constant on any interval;  $f^*$  satisfies a generalized version of (4.4) given in Appendix C.1, which depends on an additional parameter chosen to satisfy the boundary condition  $(f^*)'(\bar{s}) = 0$ .

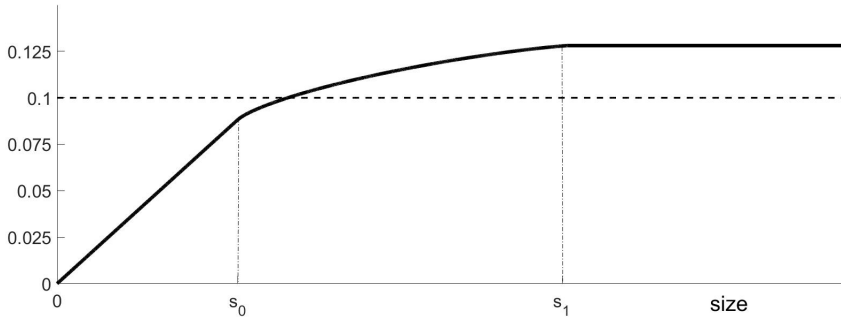
We refer a reader interested in our techniques to Appendix D.1 that presents an informal sketch of the proof of the above results, as well as the formal supporting arguments.

Because there is no explicit solution to the differential equation (4.4), a closed-form solution for the optimal fixing-weight function  $f^*$  is not available. However, Theorem 1a provides a number of economic predictions about the form of the optimal benchmark. One robust

finding is that the optimal weighting function becomes flat as the transaction size increases, as captured by the property  $(f^*)'(\bar{s}) = 0$ . The optimal  $f^*$  is typically flat after some threshold transaction size  $s_1 < \bar{s}$ , as predicted by Theorem 1b and as illustrated in Example 1, to follow. Intuitively, assigning too much weight to very large transactions is suboptimal because it induces agents with high manipulation incentives to choose large transaction sizes, resulting in overweighting such large transactions in the estimator.

Another general feature of the optimal benchmark is that  $f^*(s)$  coincides with  $(\gamma/R)s$  for a sufficiently small transaction size  $s$ . In particular,  $f^*$  attaches small weight to small transactions. The shape of the optimal fixing for small transactions is pinned down by the binding constraint that an agent with the cutoff type  $R$  prefers to avoid manipulation. This is intuitive. If a benchmark fixing places small weight on small transactions, then unmanipulated transactions are underweighted compared to the weight that they would receive in a statistically efficient estimator. Therefore, it is optimal to place the maximal weight on small transactions that is consistent with deterring manipulation by types above  $R$ .

Fig. 4.1: Optimal weighting function for Example 1. The dotted line depicts the optimal solution in the absence of manipulation.



Finally, Theorem 1a indicates that the optimal benchmark provides an incentive for “smoothing out” manipulations, preventing them from “bunching” around a given transaction size. This is perhaps somewhat surprising. The manipulated transactions have the same precisions as signals of  $Y$ , and yet it is optimal to attach different weights to them. As added intuition, we note that local behavior of  $f'$  has only second-order effects on the incremental variance term  $\sigma_Y^2 f^2(s) H(R) dG(s)$  associated with unmanipulated transactions. In contrast, the incremental variance term  $\sigma_M^2 f^2(s) (1 - H(R)) d\Psi_f(s)$  for manipulated transactions is sensitive to the local behavior of  $f'$ . This follows from the influence of  $f'$  on the distribution  $\Psi_f$ , in that relatively small changes in the slope of  $f$  can lead to large changes in the optimal transaction volume chosen by a manipulator. Under our assumptions, this variance term is convex in  $f'$ . Thus, minimizing the variance term requires minimizing the

variation of  $f'$  (subject to meeting other criteria). As a result,  $f'$  changes continuously rather than exhibiting discrete jumps.

When the fixing function  $f^*$  is that given by Theorem 1b, all manipulators choose a transaction size in  $[s_0, s_1]$ , and the distribution of sizes has full support in that interval. However, as shown in Figure 4.2, the distribution of manipulated transaction sizes is often concentrated around  $s_0$ . Under these conditions, the optimal benchmark can be well-approximated by a simple fixing that is linear up to a threshold, and constant afterwards (we comment further on this point in Example 3 in Appendix C.2)

To illustrate the above discussion we consider the following numerical example.

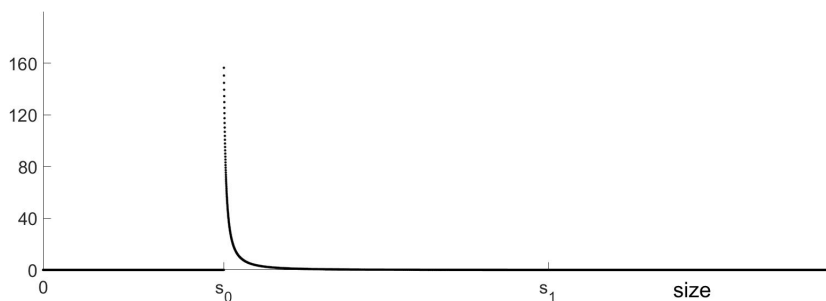
**Example 1** Suppose that  $\gamma = 1$ ,  $n = 10$ ,  $\bar{R} = 5$ ,  $\sigma_V^2 = \sigma_\epsilon^2 = \sigma_z^2 = 1$ ,  $g$  is the uniform density on  $[0, 1]$ , and  $h(x) = \exp(-x/2)/2$ . The given density  $h$  implies that, on average, the exposure to the benchmark asset is equal to 2. We set the manipulation threshold to be twice the mean,  $R = 4$ . The type threshold  $R = 4$  corresponds to a probability of manipulation of around 14%. The optimal weighting function is depicted in Figure 4.1. This function is smooth ( $C^1$ ), but its first derivative changes rapidly close to  $s_0 \approx 0.40$ . All of the manipulated transactions are in the interval  $[s_0, s_1]$ . Figure 4.2 illustrates that manipulations are in fact highly concentrated around  $s_0$ .

While the qualitative properties of the optimal weighting function are intuitive, the particular form of the ODE (4.4) is less clear. To gain intuition, we can rewrite (4.4) as

$$\underbrace{[f(s)\sigma_M^2 - f(s_1)\sigma_U^2]}_{I_M} dH\left(\frac{\gamma}{f'(s)}\right) + \underbrace{[f(s) - f(s_1)]\sigma_U^2 H(R)g(s)}_{I_U} = \underbrace{\frac{d}{ds} \left[ (f(s)\sigma_M^2 - f(s_1)\sigma_U^2) h\left(\frac{\gamma}{f'(s)}\right) \frac{\gamma}{f'(s)} \right]}_{I_A}.$$

In the above formula, one may think of  $f(s_1)$  as the optimal constant weight that would be assigned to unmanipulated transactions for the efficient estimator (fixing) that would be chosen in the absence of manipulation incentives (see Proposition 1). The term  $I_U$  is zero, that is  $f(s) = f(s_1)$ , when the weight is chosen optimally from the point of view of unmanipulated transactions. This term is proportional to the density of sizes corresponding to unmanipulated transactions. On the other hand, the term  $I_M$  is zero, that is  $f(s)\sigma_M^2 = f(s_1)\sigma_U^2$ , when the weight is chosen optimally from the point of view of manipulated transactions. This term is proportional to the density of sizes that arises from manipulated transactions. In both of these cases, individually, the term  $I_A$  is also zero because  $h(\gamma/f'(s))(\gamma/f'(s)) = 0$  when  $f$  is

Fig. 4.2: Density of manipulations



constant.<sup>23</sup> Ideally the benchmark administrator would like to set both of the terms  $I_M$  and  $I_U$  to zero, but this is impossible when  $\sigma_M^2 > \sigma_U^2$ . Thus, the administrator faces a trade-off. She either puts insufficient weight on unmanipulated transactions, which are relatively precise signals of the fundamental value, or she puts too much weight on manipulated transactions, which are relatively noisy signals of the fundamental value  $Y$ .

In balancing these two effects, the administrator takes into account the term  $I_A$ . By assumption, types in  $[R, \infty)$  manipulate. By controlling  $f'$ , the administrator controls the sizes of transactions chosen by types  $R_i$  in  $[R, \infty)$ . Because the optimal fixing is concave and differentiable, the optimal size of a manipulation is pinned down by the first-order condition for the manipulator's problem (3.2). Thus, an agent with type  $R_i = \gamma/f'(s)$  chooses size  $s$ . The term  $\gamma/f'(s)$  starts at  $R$  when  $s = 0$ , and ends at  $\infty$  when  $s = \bar{s}$ . It follows that  $dH(\gamma/f'(s))$  describes the density of manipulated transactions. The term  $I_A$  accounts for the fact that when the benchmark administrator chooses  $f(s)$  at  $s$ , she considers the effect of the speed with which the slope changes on the distribution of the remaining mass of manipulated transactions.

To complete the characterization of the optimal fixing function  $f^*$ , we observe that for the case  $R = 0$  (at which every type manipulates), the optimal solution is  $f^*(s) = 1/n$ . This is analogous to Proposition 1, replacing  $\sigma_U^2$  with  $\sigma_M^2$ . For the case  $R = \hat{R}$ , there is only one feasible fixing function, that with  $f(s) = (\gamma/R)s$ , which is thus trivially optimal.

Having characterized the solution to problem  $\mathcal{P}(R)$  for a fixed manipulation threshold  $R$ , one can solve the original problem  $\mathcal{P}$  by choosing an optimal threshold  $R^*$ . This involves computing the optimal weighting function  $f^*$  for every  $R \in [0, \hat{R}]$ , evaluating the objective function, and finding the maximum over all  $R$ , achieved at some  $R^*$ . This optimum is attained, by Berge's Theorem. While analytic solutions are infeasible, this step can be done numerically. However, we prove below that the optimal  $R^*$  is interior, and thus deters some manipulation

<sup>23</sup>While this seems to present some issues associated with division by zero, the result follows from integrability of  $h$ , and is formally stated in Lemma 7 of Appendix D.

but does not minimize the probability of manipulation among all feasible weighting functions.

**Proposition 3** *The optimal manipulation threshold for problem  $\mathcal{P}$  is interior:  $R^* \in (0, \hat{R})$ .*

A consequence of Proposition 3 is that the predictions of Theorem 1a about the shape of the fixing hold at the optimal  $R^*$ . Appendix C.2 contains an additional numerical example.

## 4.2 Robustness to order splitting

A practical concern related to the design of benchmarks based on transaction data is that agents intending to trade total quantity  $s$  of the asset may split the order into several smaller “chunks” to influence the benchmark fixing. For example, the fixing of the ICE Swap rate is based on transaction data from multiple order books.<sup>24</sup> Because orders are submitted anonymously, instead of submitting one order to one exchange, a trader could instead submit a few smaller orders to several exchanges. Similarly, the VIX benchmark,<sup>25</sup> the LBMA gold price fixing,<sup>26</sup> CME futures settlement benchmarks,<sup>27</sup> and WMR FX fixings<sup>28</sup> are based on anonymous user inputs, such as trades, quotes, or bids, and are not amenable to fixing approaches that would combine different inputs from the same entity. Such concerns are less relevant in settings where transaction data are submitted by well-identified traders – for example, from a panel of banks in the case for Libor<sup>29</sup> or registered participants of the Platts’

<sup>24</sup>As ICE describes it, “At each snapshot, we combine the order books from all the trading venues to create a synthetic order book that represents the best prices (and accompanying volumes) available in the market at that time. We then calculate the volume weighted prices at which you could fill a trade in SMS [of a standard nominal size] from this synthetic order book on both the bid and offer side and these effective prices are used to calculate the VWAMP [volume weighted average mid-price].” See “ICE Benchmark Administration Calculation of ICE Swap Rate from Tradeable Quotes Overview,” at [https://www.theice.com/publicdocs/ICE\\_Swap\\_Rate\\_Full\\_Calculation\\_Methodology.pdf](https://www.theice.com/publicdocs/ICE_Swap_Rate_Full_Calculation_Methodology.pdf)

<sup>25</sup>Chicago Board Options Exchange, “Settlement Information for VIX Derivatives,” at <http://cfe.cboe.com/cfe-products/vx-cboe-volatility-index-vix-futures/settlement-information-for-vix-derivatives>.

<sup>26</sup>LBMA Gold Price fixing is the clearing price of an anonymous submission double auction, under rules described at <http://www.lbma.org.uk/lbma-gold-price>.

<sup>27</sup>See “CME Group Daily Settlement Procedures,” which are typically VWAP based on anonymous transactions, at [https://www.cmegroup.com/content/dam/cmegroup/market-data/files/CME\\_Group\\_Settlement\\_Procedures.pdf](https://www.cmegroup.com/content/dam/cmegroup/market-data/files/CME_Group_Settlement_Procedures.pdf).

<sup>28</sup>All transactions are anonymous and enter the fixing algorithm individually, not grouped by investor. They are used to estimate an effective bid and an effective offer for each trade. The fixing is roughly the median of the implied midpoints of all of the transactions. See “Refinitiv Benchmark Services Limited, “WM/Reuters FX Benchmarks – Spot and Forward Rates Methodology Guide,” at [https://www.refinitiv.com/content/dam/marketing/en\\_us/documents/methodology/wm-reuters-methodology.pdf](https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/wm-reuters-methodology.pdf).

<sup>29</sup>Since about 2015, each bank submits all of its transactions to the benchmark administrator, IBA. So, “order splitting” is detectable. But the benchmark is merely the unweighted average of one submitted rate from each bank, from among those banks whose submitted rates are not outliers,

Market on Close<sup>30</sup> process for energy price benchmarks – because the administrator can aggregate the data from each trader before calculating the benchmark. Even then, however, robustness to order-splitting might be a desirable property if two or more traders could collude and effectively behave as a single trader who can submit more than one transaction.

So far, we have ruled out the possibility of strategic order splitting by assuming that each agent conducts exactly one transaction. If we relax this assumption, it turns out that the optimal benchmark from Section 4.1 is susceptible to this type of manipulation. To see this, imagine that agent  $i$ , with a positive manipulation incentive  $R_i$ , intends to trade the quantity  $s_1$ . (See Figure 4.1.) Beyond merely distorting the price of this transaction, the agent can additionally influence the benchmark fixing by submitting two smaller transactions, each with quantity  $s_1/2$ . Such a manipulation is costless relative to making one transaction with size  $s_1$  given our linear cost function, and yields the agent a benefit of

$$R_i(2f^*(s_1/2) - f^*(s_1)) > 0,$$

because of the concavity of the weighting function  $f^*$ .

By an extension of this argument, if the designer chooses a benchmark fixing  $f$ , the effective weighting function that will arise under optimal order-splitting takes the form

$$\tilde{f}(s) = \sup\{f(q_1) + \dots + f(q_k) : q_i \in [0, \bar{s}], q_1 + \dots + q_k = s\}. \quad (4.5)$$

Therefore, if order-splitting is allowed and costless,<sup>31</sup> it is without loss of generality to require

---

defined for example as the highest and lowest quintiles. A bank's submitted rate generally can be judgmental, or based on an algorithm used by the submitter. This is up to each submitter, subject to auditing.

<sup>30</sup>See Platts, “An Introduction to Platts’ Market-on-Close Process in Petroleum,” at <https://www.platts.com/IM.Platts.Content/aboutplatts/mediacenter/PDF/intromocoil.pdf> “Platts Market On Close (MOC): is the process Platts’ editors use to assess prices for crude oil, petroleum products and related swaps. ... The MOC is a structured, highly transparent process in which bids, offers and transactions are submitted by participants to Platts’ editors and published in real-time throughout the day until the market close. Following the close, Platts editors examine the data gathered through the day, conduct their analysis and develop price assessments that reflect an end-of-day value.” ... “the guidelines mandate that the MOC process be fully transparent. Thus, Platts requires full disclosure on the details of any bid, offer or transaction submitted into the MOC process, and does not allow participants to submit bids, offers and transactions anonymously. Also, Platts requires that the data submitted reflect actual verifiable transactions and/or bids and offers that are ‘firm,’ open to the market at large, and therefore executable. This is to prevent a participant from distorting the process (and thus a Platts assessment) with a bid or offer that is not representative of market value. On occasion, Platts’ editors will request supporting documentation from the counterparties in order to verify a transaction.

<sup>31</sup>Costless order splitting amounts to assuming that agents have no price impact in the underlying market. With price impact, submitting smaller orders might actually improve the price received by the agent, which further encourages order-splitting.



that the benchmark administrator chooses a weighting function  $f$  that leaves no incentive for this type of order-splitting manipulation. This property is easily seen to be equivalent to the condition that  $f$  is superadditive. In particular, for any positive integer  $k$ ,

$$f(ks) \geq kf(s), \quad s \leq \frac{\bar{s}}{k}. \quad (4.6)$$

Superadditivity is a cumbersome constraint in optimal control problems because it is a global property, ruling out characterizations based on local behavior. Therefore, for tractability, we will assume a slightly stronger mathematical condition by requiring (4.6) to hold for all real  $k \geq 1$ , and not only for integer  $k$ .

**Definition 1** *A benchmark weighting function  $f$  is robust to order-splitting if*

$$f(ks) \geq kf(s),$$

for all  $k \in [1, \infty)$  and all  $s$  such that  $ks \leq \bar{s}$ .

It is clear that the optimal weighting function found in Theorem 1a-1b is not robust to order splitting. In fact, if  $f$  is concave but not linear, it cannot satisfy Definition 1.

**Theorem 2** *For any  $R \leq \hat{R}$ , the optimal solution  $f^*$  to problem  $\mathcal{P}(R)$ , subject to robustness to order splitting, is given by  $f^*(s) = (\gamma/\hat{R})s = s/(n\mathbb{E}[\hat{s}_i])$ . Then, the optimal manipulation threshold  $R^*$  is equal to  $\hat{R}$ , and the associated benchmark is the absolute-volume weighted average price (A-VWAP).*

The proof can be found in Appendix D.3. Theorem 2 states that if the benchmark administrator cannot deter or detect order-splitting, then the optimal benchmark is the absolute-volume weighted average price. The intuition for this result is relatively straightforward. When agents engage in strategic order splitting, the optimal weighting function cannot be concave unless it is linear. At the same time, it is not optimal for the weighting function to be strictly convex in any interval, since no manipulator would choose a transaction size in the interior of that interval, resulting in inefficient underweighting of the unmanipulated transactions with these sizes. Thus, it is optimal to choose a linear weighting function.

Under a linear weighting function, all manipulators choose the largest feasible transaction size, which hence receives the highest possible weight. Therefore, the optimal benchmark that is unbiased and robust to order splitting minimizes the unconditional probability of manipulation. In particular, the probability of manipulation is smaller under the A-VWAP benchmark than under the optimal benchmark from Theorems 1a-1b. At the same time, the A-VWAP gives a much higher weight to manipulated transactions because if it did not, the

manipulator would split her order into multiple smaller ones. We illustrate this point with Example 3 in Appendix C.2.

## 5 Manipulation-robust fractional-weight fixings

In this section, we turn our attention to fractional-weight fixings, so that  $f_i(\hat{\mathbf{s}})$  may now depend on the entire vector  $\hat{\mathbf{s}}$  of transaction sizes, subject to summing to one for every realization of  $\hat{\mathbf{s}}$ . As discussed, this opens up incentives for collusion, and we are unable to solve an optimization problem as we did for the case of absolute-size-weighting.<sup>32</sup> Instead, we adopt an axiomatic approach. We characterize the unique class of fractional-weight fixings that deter agents from engaging in a particularly simple form of collusion. For ease of exposition, we assume for this section that  $\bar{s} = \infty$ , so that there is no upper bound on transaction sizes.

We adopt the notation  $\mathbf{s}_K = (s_i)_{i \in K}$  and, in particular sometimes write  $(\mathbf{s}'_K, \mathbf{s}_{-K})$  to denote a vector of transaction sizes in which the coordinates corresponding to a subset  $K \subset \{1, \dots, n\}$  are taken from the vector  $\mathbf{s}'$ , and the transaction sizes for the complementary subset are taken from the vector  $\mathbf{s}$ . We will also use  $\mathbf{s}_{-ij}$  as shorthand notation for the vector  $\mathbf{s}$  without its  $i$ -th and  $j$ -th coordinates, and write

$$(\mathbf{s}'_i, \mathbf{s}'_j, \mathbf{s}_{-ij}) = (s_1, \dots, s_{i-1}, s'_i, s_{i+1}, \dots, s_{j-1}, s'_j, s_{j+1}, \dots, s_n).$$

Finally, we let  $N = \{1, \dots, n\}$ . With this, we can define a property of benchmark fixings that will play a key role in our characterization.

**Definition 2** *A benchmark fixing neutralizes pairwise order splitting if for any  $\mathbf{s} \in \mathbb{R}_+^n$ ,  $(i, j) \in N^2$ , and  $x \leq s_i$ ,*

$$f_i(\mathbf{s}) + f_j(\mathbf{s}) = f_i(s_i - x, s_j + x, \mathbf{s}_{-ij}) + f_j(s_i - x, s_j + x, \mathbf{s}_{-ij}).$$

To understand the definition, suppose that for some  $i$  and  $j$  in  $N$ , we have  $X_i = X_j$ . That is, traders  $i$  and  $j$  face the same price in the market.<sup>33</sup> Then, traders  $i$  and  $j$  could conduct the following collusion. Trader  $i$  reduces her order by the quantity  $x$  and trader  $j$  increases her order by the same amount. Because the total of the order sizes of  $i$  and  $j$  is unchanged, the price would be invariant to this arrangement in some typical market mechanisms. This would be true, for example, if trading is organized with a limit order book in which the price

<sup>32</sup>The fundamental reason for the intractability of this case is that it would require solving a multivariate optimal control problem; strategic interactions between agents complicate the problem but could be handled (see Section 6).

<sup>33</sup>It is not required that the traders actually observe these prices, only that they expect to trade at similar terms.

depends on the aggregate demand and supply curves built up from orders. This collusion would not be difficult to coordinate. If by this arrangement traders  $i$  and  $j$  could increase their total influence on the benchmark fixing— as determined by the sum of their weights— then they would have the incentives to do so. Definition 2 characterizes a fixing scheme with the property that, regardless of how traders  $i$  and  $j$  split their total trade volume, the total of their trade weights remains the same, thus negating any potential collusive profits.

An alternative interpretation of the assumption can be given if  $i$  indexes individual market transaction sizes rather than the trade amount of a trader. In that case, Definition 2 is a variant of the no-order-splitting condition from Section 4.2. The condition is that if a given trader makes two transactions at the same price, then she should not be able to increase her influence on the benchmark fixing by splitting the volume across these transactions in an alternative way.

In the following result, we assume that a fixing assigns weights only to transactions of non-zero size, in that for each  $i$  and any  $\mathbf{s}_{-i}$ , we have  $f_i(0, \mathbf{s}_{-i}) = 0$ .

**Theorem 3** *A fractional-weight fixing  $f$  neutralizes pairwise order splitting and assigns weights only to transactions of non-zero size if and only if  $f$  is the VWAP fixing, that is, for each  $i$ ,  $f_i(\mathbf{s}) = s_i / (\sum_{j=1}^n s_j)$ .*

The restriction that a fixing prevents pairwise order-splitting seems weak. In particular, it only constrains a particular type of deviation between a pair of traders, and does not make any explicit assumption about the possibility of collusion among a larger group of traders, or about more elaborate collusive schemes. The assumption of assigning no weight to transactions with negligible size is also natural, for as we have argued in Observation 1, in the absence of this property of a fixing, all traders would *always* choose to manipulate because the cost of manipulating with a small transaction is also negligible. Yet, we find that these two apparently permissive assumptions pin down a unique candidate benchmark fixing, which is the VWAP.

We have not even made an assumption that the benchmark fixing must treat all transactions symmetrically. It turns out that symmetry is a necessary implication of the two assumptions of Theorem 3. Intuitively, if two traders were treated asymmetrically, they would have an incentive to collude, because the trader with a larger influence could make a larger transaction “on behalf” of both traders, effectively giving the same influence to the “less influential” trader.

In the proof of this result, a key step is to show that robustness to pairwise order-splitting implies the seemingly stronger property that the fixing neutralizes the impact of order-splitting among any subset of the traders.

**Definition 3** A benchmark fixing  $f$  neutralizes group order splitting if for any  $\mathbf{s} \in \mathbb{R}_+^n$  and any subset  $K \subseteq \{1, \dots, n\}$  of traders, the total weight assigned to traders in  $K$ ,  $\sum_{i \in K} f_i(\mathbf{s})$ , is unchanged if traders in  $K$  reallocate transaction sizes among themselves while keeping their total transaction volume  $\sum_{i \in K} s_i$  constant.

Obviously, a VWAP benchmark neutralizes group order splitting.

As a corollary of our proof, we can show that even without the condition of assigning no weight to transactions with negligible size, any fractional-weight fixing  $(f_i)_{i=1}^n$  that neutralizes pairwise order-splitting must take the form

$$f_i(\mathbf{s}) = \left( 1 - \frac{1}{n} \sum_{i \in N} \beta_i \left( \sum_{j=1}^n s_j \right) \right) \frac{s_i}{\sum_{j=1}^n s_j} + \beta_i \left( \sum_{j=1}^n s_j \right) \frac{1}{n}, \quad (5.1)$$

for some function  $\beta : \mathbb{R}_+ \rightarrow \mathbb{R}_+^n$ . If the fixing is also symmetric across  $i$ , then (5.1) shows that the weight  $f_i(\mathbf{s})$  is itself a weighted average of the VWAP weight and the constant weight  $1/n$ , where the relative importance of these two components can depend on the total volume of trade in the market.

## 6 Discussion

We developed a simple model for the design of robust benchmark fixings in settings for which incentives to manipulate the benchmark arise from a profit motive related to investment positions that are valued according to the benchmark. We have restricted attention to fixings that are given by a size-dependent weighted average price, an important limitation that we will discuss shortly. For the class of absolute-size-weight fixings, we characterized the optimal weight for each size of transaction. We showed that an optimal benchmark fixing must in general allow some amount of manipulation, puts negligible weight on small transactions, and nearly equal weight on large transactions. When order-splitting cannot be detected or otherwise deterred, the absolute-volume-weighted average price (A-VWAP) emerges as the optimal design. For the class of relative-size weighted fixings, we have shown that the conventional VWAP, whose weights are *relative* volume shares, emerges as the unique fixing that is fractional, assigns weights only to transactions of non-zero size, and neutralizes pairwise order-splitting.

### 6.1 A-VWAP versus VWAP

Since our two main results characterize Absolute-VWAP and the conventional VWAP as the designer-preferred benchmarks among absolute-size-weight and fractional-weight fixings,

respectively, we now discuss the trade-offs associated with choosing between these two designs, and identify directions for future research.

**Statistical efficiency.** We first show that fractional-weight fixings (in particular, VWAP) can mechanically lower the mean squared error of the benchmark fixing compared to an absolute-size-weight fixing, by eliminating the component of the error that depends on the variance of the fundamental  $\sigma_Y^2$ . At the same time, the result shows that this is the only improvement, and that the improvement disappears when  $n$ , the number of transactions, grows large, as long as the fractional-weight fixing is asymptotically efficient.

**Proposition 4** *For any fractional-weight fixing  $f$  with a mean squared error  $V$ , there exists an unbiased absolute-size-weight fixing  $\bar{f}$  with a mean squared error strictly smaller than  $V + \delta_n \sigma_Y^2$  with  $\delta_n < 1$ . Moreover,  $\bar{f}$  induces the same equilibrium among agents as  $f$ , but in dominant strategies. Furthermore, if  $f$  is asymptotically efficient (its mean squared error converges to 0 as  $n \rightarrow \infty$ ), then  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ .*

It is instructive to go through the main idea behind the proposition. Starting with some fractional weight  $f_i(\hat{\mathbf{s}})$  and the corresponding Bayes-Nash equilibrium among the traders, the designer can instead present each agent  $i$  with the interim expected weight  $\bar{f}_i(\hat{s}_i) = \mathbb{E}[f_i(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_{-i})]$  that only depends on the transaction size  $\hat{s}_i$  of agent  $i$ . Intuitively, the designer simulates the expected weight, using the equilibrium distribution, instead of conditioning the weight on the transaction of agent  $i$  on the realization of other traders' transactions. As we have shown in Section 3, the manipulation incentives of agent  $i$  depend only on the interim expected weight of agent  $i$ . Hence, agent  $i$  has exactly the same best response as before. However, her best response is now a dominant strategy, since her payoff is trivially independent of the strategies adopted by other traders under the absolute-size-weight fixing. The new weighting scheme is unbiased because it is an expectation over the original fixing which is unbiased (by the fact that it is fractional). The potential increase in the mean squared error associated with adopting the absolute-size-weight fixing is due to the fact that the component  $Y$  in every transaction price  $\hat{X}_i$  now receives a total weight of  $\sum_{i=1}^n \bar{f}_i(\hat{s}_i)$ , which is typically different from one, resulting in an additional error term  $Y(1 - \sum_{i=1}^n \bar{f}_i(\hat{s}_i))$ . In contrast, that term would be 0 under any fractional-weight fixing, by definition.

Proposition 4 does not preclude that the improvement achieved by allowing the benchmark weights to depend on all transaction sizes could be quantitatively significant. However, it implies that this larger class of fixings cannot reduce the variance associated with either noise ( $\sigma_\epsilon^2$ ) or manipulation ( $\sigma_z^2$ ). Consequently, if  $\sigma_Y^2$  were negligible compared to  $\sigma_\epsilon^2$  and  $\sigma_z^2$ , then it would be *approximately optimal* to choose benchmark weights that take the absolute-size-weight form. Moreover, the proof of Proposition 4 implies that any benchmark that

completely eliminates the noise due to  $\sigma_Y^2$  (for example, VWAP) must necessarily *increase* the total noise due to  $\sigma_\varepsilon^2$  or  $\sigma_z^2$ . The intuition is that the incentives to manipulate in our model only depend on the properties of the *interim expected* weight, and a fractional weight introduces additional variance due to the fluctuation of the ex-post weight around its interim expected value. In summary, while fractional weights help reduce the mean squared error of the fixing, absolute-size weights—including those of the A-VWAP—are actually superior in terms of reducing the variance due to manipulation.

**Fixings outside of the range of transaction prices.** A potential disadvantage of absolute-size-weight benchmarks is that they can generate fixings that lie outside of the range of observed transaction prices. For a simple example, consider the trivial case in which all transaction prices are the same, and equal to  $X$ , while all traders choose transaction sizes that are twice the average size. Then, the A-VWAP fixing is  $2X$ . While this outcome is unlikely when transaction sizes are independent (as assumed in our formal model) or if  $n$  is large, it becomes more plausible in settings for which trade volume is strongly correlated across agents, or if there are few active traders. A fractional-weight benchmark eliminates this problem.

**Observation 2** *A general fixing  $\hat{Y}$  of the form (2.1) has the in-range property, in that  $\hat{Y} \in [\min_i \hat{X}_i, \max_i \hat{X}_i]$  for all realizations of the primitive random variables, if and only if it is a fractional-weight fixing.*

The proof of Observation 2 is immediate and hence skipped. As a corollary, the VWAP emerges as the unique linear fixing that satisfies the in-range property, neutralizes pairwise order-splitting, and assigns weights only to transactions of non-zero size.

Given Observation 2 and preceding discussions, one could ask whether there exists a fractional-weight representation of the A-VWAP benchmark. That is, does there exist a fractional-weight fixing such that the interim expected weights faced by agents are the same as those of the A-VWAP? As we have already argued, this would preserve the equilibrium among the agents, and could actually lower the mean squared error by eliminating the term that depends on  $\sigma_Y^2$ . We give a positive answer to this question, under stated conditions.

**Proposition 5** *Let  $f^*(s) = s/(n\mathbb{E}[\hat{s}_1])$  be the optimal A-VWAP benchmark from Section 4.2. For large enough  $n$ , there exists a fractional-weight fixing  $f = (f_i)_{i=1}^n$  such that for any agent  $i$ , the expected equilibrium influence on the benchmark of misreporting is the same as under  $f^*$ , in that  $\mathbb{E}[f_i(\hat{s}_i, \hat{\mathbf{s}}_{-i}) \mid \hat{s}_i] = f^*(\hat{s}_i)$ . In particular, the induced equilibrium behavior of agents is the same under  $f$  and  $f^*$ . Moreover, the same conclusion holds true for any  $n \geq 2$ , if  $g$  has the uniform distribution.*

Proposition 5 can be seen a converse to Proposition 4. Proposition 4 explains how we can convert a fractional-weight fixing into an absolute-size-weight fixing while preserving the equilibrium, without a large increase in the mean squared error. Proposition 5 explains how we can convert the A-VWAP fixing into a fractional-weight fixing while preserving the equilibrium, and also while potentially lowering the mean squared error.

The proof, found in Appendix D.6, applies the Matthews-Border condition from the literature on reduced-form auctions (Matthews, 1984, and Border, 1991): If  $\Lambda(s)$  is the distribution of the (potentially manipulated) transaction size, then a non-decreasing weight function  $f(s)$  is induced by some fractional weight (in the sense of Proposition 5) if and only if

$$\int_{\tau}^{\bar{s}} f(s) d\Lambda(s) \leq \frac{1 - \Lambda^n(\tau)}{n}, \quad \tau \in [0, \bar{s}), \quad (6.1)$$

with equality for  $\tau = 0$ . We conjecture that condition (6.1) might be useful in characterizing the optimal benchmark in the class of fractional-weight fixings, a task that we leave for future research. (The main complication is that the objective function of the designer—the mean squared error of the fixing—depends on the entire weight function, and not only on its interim expectation, resulting in a multi-variate optimal control problem.)

**Manipulation and collusion.** As we emphasized in Section 2.4, the benefit of the potentially lower variance of a benchmark  $f$  that is not restricted to be an absolute-size-weight fixing might be offset by the associated incentives for collusion. When the weights depend on all transaction sizes, strategic players have incentives to collude by sharing data and coordinating actions. (For a given benchmark design, a collusive model of manipulation is suggested by Osler (2016).) Even with a VWAP benchmark fixing, which was shown to neutralize group order splitting (Theorem 3), agents can still benefit from sharing information, because manipulation becomes more effective when other traders trade less. In contrast, as shown formally by Proposition 4, with an absolute-size-weight fixing, the benchmark designer can ensure that agents have dominant strategies.

A setting that in practice has activated traders' incentives for collusive information sharing and joint setting of transaction sizes is the daily fixing of the WM/Reuters FX benchmarks.<sup>34</sup>

<sup>34</sup>See, for example, the U.K. Financial Conduct Authority's findings, at <http://www.fca.org.uk/news/fca-fines-five-banks-for-fx-failings>, which include: *Traders at different Banks formed tight knit groups in which information was shared about client activity, including using code names to identify clients without naming them. These groups were described as, for example, 'the players', 'the 3 musketeers', '1 team, 1 dream', 'a co-operative' and 'the A-team'. Traders shared the information obtained through these groups to help them work out their trading strategies. They then attempted to manipulate fix rates and trigger client "stop loss" orders (which are designed to limit the losses a client could face if exposed to adverse currency rate movements). This involved traders attempting to manipulate the relevant currency rate in the market, for example, to ensure that the rate at which*

It was discovered by regulators that groups of dealing traders have shared information about their customers' orders, which they will later fill at the benchmark price, acting as direct counterparties to their own respective customers. The traders are able to influence the fixing with their own trades and quotes during the minutes just before and during the fixing time window. Through information sharing, they have allegedly profited by colluding in a manner that reduces the sizes of the transactions or quotes of some traders relative to the sizes of those of other traders. This does not require that all traders profit on each occasion. They can "take turns." For example, it could be discovered through information sharing that a large manipulation profit would likely be achieved by traders with a buying interest, provided that other traders in the group reduce the sizes of their sell transactions or quotes. They can collusively agree that on this occasion, the traders who wish to buy shall take precedence. The WM/Reuters fixing method is not of the VWAP form, but involves similar incentives for information sharing and collusive setting of transaction sizes.

**Summary.** We conclude that the A-VWAP benchmark might be an attractive alternative for benchmark administrators in settings where collusion (in particular, data sharing) among traders is a real concern, the number of transactions is relatively large, and the day-to-day variability of total volume is not so high that an A-VWAP benchmark would likely generate fixings outside of the range of prices. In some cases, the designer might be able to choose a fractional-weight representation of the A-VWAP benchmark (see Proposition 5) to avoid the undesirable out-of-range property. On the other hand, a conventional VWAP benchmark might perform better in a setting with few traders, because it lowers the variance associated with the variability in the fundamental  $Y$ , which could be significant with a small number of transactions. At the same time, the VWAP neutralizes some forms of collusion (see Theorem 3). While the VWAP does create incentives for information sharing across agents, this type of collusion might be effectively addressed through audits and fines.

## References

- Aït-Sahalia, Y. and Yu, J. (2009) High Frequency Market Microstructure Noise Estimates and Liquidity Measures, *Annals of Statistics*, **3**, 422–457.
- Alternative Reference Rate Committee (2018) [Second Report](#), Federal Reserve Bank of New York.
- Aquilina, M., Ibikunle, G., Mollica, V., Pirrone, A. and Steffen, T. (2018) [The Visible Hand: Benchmarks, Regulation and Liquidity](#), Working paper, University of Edinburgh.
- Baldauf, M., Frei, C. and Mollner, J. (2020) [Principal Trading Arrangements: When Are Common Contracts Optimal?](#), Working paper.

---

*the bank had agreed to sell a particular currency to its clients was higher than the average rate it had bought that currency for in the market. If successful, the bank would profit.*



- Bank of England (2014) [Recommendations on additional financial benchmarks to be brought into UK regulatory scope](#), Fair and Effective Markets Review, Report to Her Majesty's Treasury, August.
- Ben-Porath, E., Dekel, E. and Lipman, B. L. (2014) Optimal allocation with costly verification, *American Economic Review*, **104**, 3779–3813.
- Berkowitz, S., Logue, D. and Noser, E. (1988) The total cost of transactions on the nyse, *Journal of Finance*, **43**, 97–112.
- BIS (2013) [Towards Better Referece Rate Practices: A Central Bank Perspective](#), Bank for International Settlements, Basel, Working Group report, Bank for International Settlements.
- Border, K. C. (1991) Implementation of reduced form auctions: A geometric approach, *Econometrica*, **59**, 1175–1187.
- Brousseau, A., V. Chailloux and Durre, A. (2009) [Interbank Offered Rate: Effects of the Financial Crisis on the Information Content of the Fixing](#), IESEG School of Management Working Paper No. 2009-ECO-10.
- Bruckner, A. M. and Ostrow, E. (1962) Some function classes related to the class of convex functions., *Pacific J. Math.*, **12**, 1203–1215.
- Bruntsen, J. (2014) [ECB Seeks Rules to Stem Bank Exodus from Benchmark Panels](#), Bloomberg, June 19.
- Chen, J. (2013) [LIBOR's Poker: Interbank Borrowing Costs and Strategic Reporting](#), Working paper, Haas School of Business, University of California, Berkeley., UC Berkley Haas School of Business Working Paper.
- Coulter, B., Shapiro, J. and Zimmerman, P. (2018) A mechanism for libor, *Review of Finance*, **22**, 491–520.
- Duffie, D. and Dworczak, P. (2014) Robust benchmark design, Working Paper, Stanford Graduate School of Business.
- Duffie, D., Dworczak, P. and Zhu, H. (2017) Benchmarks in search markets, *The Journal of Finance*, **72**, 1983–2044.
- Duffie, D. and Stein, J. (2015) Reforming LIBOR and other financial market benchmarks, *Journal of Economic Perspectives*, **29**, 191–212.
- Eisl, A., Jankowitsch, R. and Subrahmanyam, M. (2014) [The Manipulation Potential of Libor and Euribor](#), Working paper, New York University.
- European Union (2016) [Regulation \(EU\) 2016/1011 of the European Parliament and of the Council](#), European Union, Brussels.
- Financial Conduct Authority (2012) [The Final Report of the Wheatley Review of LIBOR](#), Financial Conduct Authority, Her Majesty's Treasury, October.
- Financial Conduct Authority (2014) [Final Notice; Reference Number 121882](#), Financial Conduct Authority, Financial Services Authority, London, England.
- Financial Conduct Authority (2016) [EU Benchmark Regulation](#).
- Financial Stability Board (2014) [Foreign Exchange Benchmarks Consultative Document](#), Financial Stability Board, Basel.

- Finch, G. and Larkin, N. (2014) [U.K. Seeks to Criminalize Manipulation of 7 Benchmarks](#), Bloomberg, September 25.
- Foreign Exchange Benchmark Group (2014) [Foreign Exchange Benchmarks Consultative Report](#), Financial Stability Board, July.
- Gencarelli, D. (2005) [One Pill, Many Prices: Variation in Prescription Drug Prices in Selected Government Programs](#), National Health Policy Forum, Issue Brief Number 807, August 29, George Washington University.
- Gencarelli, D. M. (2002) [Average Wholesale Price for Prescription Drugs: Is There a More Appropriate Pricing Mechanism](#), *NHPF Issue Brief*, No. 775.
- Griffin, J. and Shams, A. (2017) Manipulation in the VIX?, *Review of Financial Studies*, **31**.
- Hou, D. and Skeie, D. (2013) LIBOR: origins, economics, crisis, scandal and reform, *The New Palgrave Dictionary of Economics, Online Edition*, edited by Steven N. Durlauf and Lawrence E. Blume.
- Hurtado, P. (2014) [Deutsche Bank, HSBC Accused of Silver Fix Manipulation](#), Bloomberg, July 25.
- Innes, R. D. (1990) Limited-liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory*, **52**, 45–67.
- International Organization of Securities Commissions (2013) [Principles for Financial Benchmarks: Final Report](#), International Organization of Securities Commissions, FR07/13.
- Kartik, N. (2009) Strategic communication with lying costs, *The Review of Economic Studies*, **76**, 1359–1395.
- Kephart, A. and Conitzer, V. (2016) The revelation principle for mechanism design with reporting costs, in *Proceedings of the 2016 ACM Conference on Economics and Computation*, ACM, New York, NY, USA, EC '16, pp. 85–102.
- Lacker, J. M. and Weinberg, J. A. (1989) Optimal contracts under costly state falsification, *Journal of Political Economy*, **97**, 1345–1363.
- Malloch, T. R. and Mamorsky, J. D. (2013) *The End of Ethics and a Way Back: How To Fix a Fundamentally Broken Global Financial System*, John Wiley & Sons Singapore Pte. Ltd.
- Market Participants Group on Reference Rate Reform (2014) [Final Report](#), Financial Stability Board.
- Matthews, S. (1984) On the implementability of reduced form auctions, *Econometrica*, **52**, 1519–22.
- Milgrom, P. and Shannon, C. (1994) Monotone comparative statics, *Econometrica*, **62**, 157–180.
- Mylovanov, T. and Zapechelnyuk, A. (2017) Optimal allocation with ex post verification and limited penalties, *American Economic Review*, **107**, 2666–94.
- Osler, C. (2016) [Dealer Trading at the Fix](#), Working paper, Brandeis University.
- Patterson, S. and Burne, K. (2013) [CFTC Probes Potential Manipulation](#), *Wall Street Journal*, April 8.
- Rockafellar, T. (1970) *Convex Analysis*, Princeton: Princeton University Press.

- Scheck, J. and Gross, J. (2013) [Traders Try to Game Platts Oil-Price Benchmark](#), *Wall Street Journal*, June 19.
- Seierstad, A. and Sydsaeter, K. (1987) *Optimal Control Theory with Economic Applications*, Elsevier, North-Holland, Advanced Textbooks in Economics, volume 24.
- Stahl, I. D. O. (1989) Oligopolistic pricing with sequential consumer search, *American Economic Review*, **79**, 700–712.
- Vaughn, L. (2014) [Gold Fix Study Shows Signs of Decade of Bank Manipulation](#), *Bloomberg*, February 8.
- Youle, T. (2014) [How Much Did Manipulation Distort the Libor?](#), Mimeo, Univeristy of Minnesota.
- Zhang, A. L. (2020) [Competition and Manipulation in Derivative Contract Markets](#), Working paper.

## A Non-linear fixings

In this appendix, we briefly address the issue of non-linear fixings, focusing on the most famous example: the median-based fixing. While comparing the statistical efficiency of the optimal linear benchmark to median-based fixings is beyond the scope of this paper, we use our framework to argue that the potential benefit of such benchmarks – reducing the incentives to manipulate (see for example [Eisl \*et al.\*, 2014](#), and [Youle, 2014](#)) – might actually be lower than expected.

First, assuming  $n$  is odd for simplicity, define the *median benchmark fixing*  $\hat{Y}^{MED}$  as

$$\hat{Y}^{MED} = \hat{X}^{(\frac{n+1}{2})},$$

where  $\hat{X}^{(k)}$  is the  $k$ -th order statistic of  $\hat{X}_1, \dots, \hat{X}_n$ .

**Claim 1** *Under the benchmark  $\hat{Y}^{MED}$ , every agent chooses to manipulate.*

The proof is straightforward: An agent can have a non-negligible influence on the benchmark by submitting an arbitrarily small (thus arbitrarily cheap) transaction with a maximally distorted price. While a heavily distorted price is unlikely to be equal to the benchmark, there is a positive probability (bounded away from zero) that it will affect which price becomes the median observation (relative to not manipulating). Thus, even for very small exposure  $R_i$ , agent  $i$  would want to manipulate in this way.

Of course, this undesirable property of the median benchmark, as defined above, is due to its insensitivity to transaction sizes. The logic behind [Claim 1](#) suggests that we should instead consider the *volume-weighted* median benchmark fixing defined by

$$\hat{Y}^{VMED} = \hat{X}^{(j)},$$

where

$$j = \min \left\{ k : \sum_{j=1}^k \hat{s}^{(j)} \geq \frac{1}{2} \sum_{i=1}^n \hat{s}_i \right\},$$

with  $\{(\hat{X}^{(j)}, \hat{s}^{(j)})\}_{j=1}^n$  ordered by decreasing prices.

**Claim 2** *Let  $p_n = \mathbb{P}\left(\frac{1}{n-1} \sum_{j=1}^{n-1} s_j \leq \mathbb{E}[s_1]\right)$ . Then, a necessary condition for truthful reporting to be an equilibrium under  $\hat{Y}^{VMED}$  is that*

$$\bar{R} \leq \frac{n-1}{p_n} \gamma \mathbb{E}[s_1].$$

Recall from Proposition 2 that a sufficient condition for truthful reporting under a linear benchmark is  $\bar{R} \leq n\gamma\mathbb{E}[s_1]$ . Because  $p_n$  will typically be close to  $1/2$ , Claim 2 shows that truthful reporting under the volume-weighted median benchmark requires parameter values that are of the same order of magnitude, and can in fact be more difficult to achieve if the sample size is small and the distribution  $g$  of unmanipulated transaction sizes has a thick right tail.

To prove Claim 2, suppose that everyone but agent  $i$  reports truthfully. If agent  $i$ 's exposure is  $\bar{R}$ , one possible manipulation is to submit a transaction with distortion  $z_i$  and size  $s_i = (n-1)E[s_1]$ . Then, there is probability at least  $p_n$  that the fixing will be equal to agent  $i$ 's transaction price, so agent  $i$ 's expected influence on the benchmark is at least  $p_n z_i$ . The inequality from Claim 2 says that the benefit from manipulating,  $\bar{R} p_n z_i$  is smaller than the cost of manipulation,  $\gamma(n-1)E[s_1]z_i$ .

The necessary condition from Claim 2 is not sufficient because (i) the simple calculation in the proof ignores the fact that the manipulation can also influence the benchmark indirectly, even when agent  $i$ 's price does not become the benchmark, and (ii) there could be other, more subtle, manipulation strategies. Therefore, the scope of the median benchmark to reduce manipulations is quite limited.

Many benchmarks, including Libor, lie in between the volume-weighted average price and the volume-weighted median price by excluding only some ‘‘outlier’’ prices. An even more sophisticated approach would be to compute, for every transaction, the posterior probability that the transaction is manipulated, and to use this information to construct weights. We leave these directions for future research.

## B Models of manipulation

This section presents two stylized models of trading and manipulation that give rise to the functional forms for costs and incentives assumed in Section 2. Apart from providing a

microeconomic foundation for our assumptions, these models give more precise meanings to some model parameters.

## B.1 Committed quotes and costly search

We first consider a framework in which manipulation is costly because agents are committed to offering execution at the price quotes they submit to the benchmark administrator. In this framework, as is common in some actual benchmark settings, the submitting agents are dealers whose quotes are used to fix the benchmark. This was the case for the main industry benchmark for interest rate swaps known as ISDAFIX, whose manipulation<sup>35</sup> triggered more than \$600 million in fines for several dealers, Deutsche Bank, Goldman Sachs, Royal Bank of Scotland, Citibank, and Barclays, and to a more robust benchmark design, as outlined by [Aquilina, Ibikunle, Mollica, Pirrone and Steffen \(2018\)](#).

Manipulation consists in quoting a price that is an overestimate or underestimate of the true value of the asset to the dealer. If the values for the asset are highly correlated among market participants, then a mispriced quote is likely to be executed by a different investor, yielding a loss to the quoting bank. In an instance of manipulation of ISDAFIX by Deutsche Bank Securities Inc., the CFTC found<sup>36</sup> that “DBSI Swap traders would tell the Swaps Broker their need for a certain swap level at 11:00 a.m. or their need to have the level moved up or down. On at least one occasion, the Swaps Broker expressed the need to know how much ‘ammo’ certain DBSI traders had to use in order to move the screen at 11:00 a.m.” The “ammo” presumably refers to losses that the DBSI would incur from trades at manipulated quotes.

The probability of an execution at a distorted quote depends both on the degree of distortion and also on the transparency of the market. If quotes are public (as would be the case in a centralized limit order book), a significantly distorted quote would be executed with a probability close to one. If the market is more opaque or less active, and especially if quotes are revealed to traders only upon request (as in bilateral over-the-counter markets and on multilateral request-for-quote platforms), then the probability of incurring a loss by offering a distorted quote would be lower.

In our model, dealer  $i$  chooses  $\hat{s}_i \in [0, \bar{s}]$  and  $\hat{z}_i \in \{-\bar{z}, 0, \bar{z}\}$ , for some constant  $\bar{z} > 0$  which we could set to  $\sigma_z$  to match the notation from the baseline model. The variable  $X_i$  is interpreted as the actual per-unit value of the asset to dealer  $i$ . The dealer commits to trade up to  $\hat{s}_i$  units at a price  $\hat{X}_i = X_i + \hat{z}_i$ , where the pair  $(\hat{X}_i, \hat{s}_i)$  is used as a benchmark

<sup>35</sup>See [“CFTC Orders The Royal Bank of Scotland to Pay \\$85 Million Penalty for Attempted Manipulation of U.S. Dollar ISDAFIX Benchmark Swap Rates.”](#)

<sup>36</sup>See [CFTC Orders Deutsche Bank Securities Inc. to Pay \\$70 Million Penalty for Attempted Manipulation of U.S. Dollar ISDAFIX Benchmark Swap Rates](#), CFTC, February 1, 2018.

submission. For simplicity, we set the bid-ask spread to zero, that is,  $\hat{X}_i$  is both a bid and an ask. We assume that  $Y$  has unbounded support, while  $\epsilon_i$  has a symmetric distribution on an interval  $[-\bar{\epsilon}, \bar{\epsilon}]$ , for some  $\bar{\epsilon} \leq \bar{z}/2$ . This captures the idea that the distortion in prices due to manipulation is larger than the distortion due to idiosyncratic differences in the value of the asset to different traders.

We adopt a stylized search protocol to determine the probability that a committed quote is executed. Before observing its manipulation incentive type  $R_i$ , dealer  $i$  chooses a search intensity  $\lambda_i \in [0, 1]$ , paying a cost  $c(\lambda_i) = \frac{1}{2}\bar{c}\lambda_i^2$ . Here,  $\lambda_i$  is the probability that the dealer will be allowed to trade at the committed quotes of some other (randomly chosen) dealer  $j$ . We assume that each dealer is contacted at most once.<sup>37</sup> Upon contacting  $j$ , dealer  $i$  maximizes the value of its chosen transaction. Because  $X_i$  is the unit value of the asset to dealer  $i$ , the resulting payoff of dealer  $i$  is

$$\max \left\{ \max_{s \leq \hat{s}_j} (X_i - \hat{X}_j) s, \max_{s \leq \hat{s}_j} (\hat{X}_j - X_i) s \right\}.$$

Here, dealer  $i$  buys or sells the maximum quantity  $\hat{s}_j$  to which dealer  $j$  has committed, due to linearity in value. The difference between the value  $X_i$  and the quote  $\hat{X}_j$  determines the direction of trade.

### B.1.1 Solution

We focus on symmetric Nash equilibria. Dealer  $i$  makes two choices, the search intensity  $\lambda_i$  and the manipulation levels  $(\hat{z}_i, \hat{s}_i)$ . Regarding the first choice, the expected payoff to a dealer conditional on a successful search depends on the probability that other banks choose to manipulate. If  $p_M$  denotes the equilibrium probability of manipulation, then that expected payoff is

$$\begin{aligned} \mathbb{E} \left( (1 - p_M)|X_i - X_j| + \frac{1}{2}p_M|X_i - \bar{z} - X_j| + \frac{1}{2}p_M|X_i + \bar{z} - X_j| \right) \mathbb{E}(\hat{s}_j) \\ = [(1 - p_M)\mathbb{E}(|\epsilon_i - \epsilon_j|) + p_M\bar{z}] \mathbb{E}(\hat{s}_j) \equiv \phi. \end{aligned}$$

The optimal choice of search intensity is thus  $\lambda^* = \min \{1, \phi\bar{c}^{-1}\}$ .

As for the choice of manipulation, the dealer can always guarantee a zero payoff by quoting a price equal to the true value  $X_i$ , regardless of the size  $\hat{s}_i$ , by choosing  $\hat{z}_i = 0$ . On the other hand, choosing  $\hat{z}_i \in \{-\bar{z}, \bar{z}\}$  yields a payoff  $-\bar{z}\hat{s}_i$  in the event of being contacted by another

<sup>37</sup>Formally, imagine the following iterative procedure. Dealer 1 contacts one of the in dealer  $N \setminus \{1\}$  with probability  $\lambda_1$ . If dealer 1 contacts dealer  $j$ , then dealer 2 contacts one of the dealers in  $N \setminus \{2, j\}$  with probability  $\lambda_2$ , and so on.

dealer. The probability of being contacted is

$$\sum_{k=1}^{n-1} \binom{n-1}{k} (\lambda^*)^k (1 - \lambda^*)^{n-1-k} \frac{k}{n-1} = \lambda^*.$$

Taking into account the payoff generated by influencing the benchmark, and normalizing the payoff from not manipulating to zero, we see that the payoff from choosing  $(\hat{s}_i, \hat{z}_i)$  is equal to

$$(R_i f(\hat{s}_i) - \lambda^* \hat{s}_i) \hat{z}_i$$

which is exactly the expression assumed in Section 2, when taking  $\gamma = \lambda^*$ .

### B.1.2 Discussion

Based on the simple model of the previous subsection, the parameter  $\gamma$  can be interpreted as the probability of execution of a manipulated quote. If trade takes place on an active limit order book, then it is natural to assume that the cost  $\bar{c}$  of search is nearly zero, and hence that  $\gamma = \lambda^*$  is close to 1. That is, manipulation would almost always yield a trading loss. On the other hand, in an opaque over-the-counter markets,  $\bar{c}$  may be relatively large, and hence manipulation is less costly – a manipulated quote might not always be executed. As a consequence, holding the benchmark fixed, the probability of manipulation is higher in an opaque market.

If  $\lambda^*$  is less than one, there is an additional feedback effect between the benchmark fixing and the probability of manipulation. The ex-ante probability  $p_M$  of manipulation by any dealer is  $1 - H(R_f)$ , which is the probability that the dealer's exposure type  $R$  exceeds the threshold  $R_f$  determined by the weighting function  $f$  used in the fixing. If  $f$  is changed to reduce manipulation, then  $R_f$  goes up and  $p_M$  goes down. This, however, implies that the incentive to search is reduced, because the probability of encountering a profitable distorted quote gets smaller. As a consequence,  $\lambda^*$  decreases, and manipulation becomes cheaper. In a sense, the benchmark fixing and the market forces act as substitutes in preventing manipulation when the market is relatively opaque.

This discussion suggests that moving from a centralized to an opaque market may have an ambiguous influence on the shape of the optimal benchmark. On one hand, because a given manipulation of the price is less costly in an opaque market, the fixing that should be chosen in an opaque market would place a relatively smaller weight on small transactions. On the other hand, a fixing that deters manipulation lowers the cost of manipulating through the equilibrium effect on the search intensity of other market participants.

## B.2 An auction model

In this subsection, we consider an alternative trading model. When a liquidity shock hits a dealer, it may request quotes from other dealers, as is typical on electronic request-for-quote (RFQ) platforms. We model this as a sealed-bid auction. Absent incentives to manipulate, the dealer will accept the most attractive quote, for example, the lowest ask when it needs to buy the asset. The execution price, along with the corresponding trade volume, is then used to calculate the benchmark fixing. If, however, the dealer wants to inflate the fixing in order to take advantage of a long position in benchmark-linked assets, the dealer has an incentive to trade at the highest ask offered in the auction. This induces a tradeoff between the loss incurred in the auction and the gain associated with distorting the benchmark fixing.

We build a stylized model that aims to capture the main incentives. Dealer  $i$  is hit by a liquidity shock  $\delta_i$  that takes one of the values  $\{-\Delta, \Delta\}$  with equal probability, for some  $\Delta > 0$ . Dealer  $i$  then values each unit of the asset at  $Y + \delta_i$ , for quantities up to  $s_i$ . Whenever a dealer is hit by a shock, it requests quotes from two other dealers who have access to an unlimited supply of the asset at the common-value price  $Y$ . (The restriction to only two other dealers is not essential for the qualitative results but will yield explicit analytic solutions.) We model the competition between the two quoting dealers as a first-price auction (Bertrand competition). Absent incentives to manipulate, the dealer requesting the quote chooses the more attractive of the quotes, and thus Bertrand forces push the price to  $Y$ . However, when the quote-requesting dealer is a manipulator, it chooses the least attractive of the quotes, creating an incentive for dealers to provide quotes further away from the value  $Y$ .

### B.2.1 Solution

For concreteness, consider the case in which dealer  $i$  requests quotes to buy the asset (the opposite case is symmetric). Let  $p_M$  be the equilibrium probability that dealer  $i$  manipulates by accepting the higher of the quotes, corresponding to the case of a positive exposure  $R_i$ . In the unique symmetric equilibrium of the auction, the two dealers that provide quotes randomize their offers according to a continuous distribution function  $F$  with support  $[Y + \lambda\Delta, Y + \Delta]$ , where  $\lambda$  is determined in equilibrium. Following the line of argument in [Stahl \(1989\)](#), this requires each of the two dealers to be indifferent between all per-unit quotes  $q$  in the support of  $F$ , so that

$$[(1 - p_M)(1 - F(q)) + p_M F(q)](q - Y) = p_M \Delta.$$

Solving, we obtain

$$F(q) = 1 - \frac{p_M}{1 - 2p_M} \frac{Y + \Delta - q}{q - Y}$$



which is a well defined cdf when  $p_M < 1/2$ . Moreover, we have  $\lambda = p_M/(1 - p_M)$ . If  $p_M$  is small, the quotes are close to  $Y$ . When  $p_M$  is relatively high (but below  $1/2$ ), the quotes are close<sup>38</sup> to  $Y + \Delta$ . With the above description, we can calculate equilibrium payoffs, and the distribution of transaction data. Let  $\varepsilon_i^k$ , for  $k = 1, 2$ , and  $i = 1, 2, \dots, n$ , be the profit margin charged by dealer  $k$  in the auction requested by dealer  $i$ . That is,  $Y + \varepsilon_i^1$  and  $Y + \varepsilon_i^2$  are the quotes received by dealer  $i$ . Normalizing the payoff from not manipulating to zero, we take the cost of manipulation to be equal to the extra profit margin conceded by dealer  $i$  through choosing the less attractive quote for  $\hat{s}_i$  units of the asset. This concession is  $\hat{s}_i \mathbb{E} [\max\{\varepsilon_i^1, \varepsilon_i^2\} - \min\{\varepsilon_i^1, \varepsilon_i^2\}]$ . Taking into account the benefit from influencing the fixing, the net expected payoff from manipulation is equal to

$$(R_i f(\hat{s}_i) - \hat{s}_i) \mathbb{E} z_i,$$

where the random variable  $z_i$  is defined by  $z_i = [\max\{\varepsilon_i^1, \varepsilon_i^2\} - \min\{\varepsilon_i^1, \varepsilon_i^2\}]$ . This setting can therefore be viewed as a version of our basic model for the case  $\gamma = 1$ .

### B.2.2 Discussion

The model of this section endogenizes the noise structure assumed in Section 2. The noise term  $\varepsilon_i$  reflects the dispersion in bids and asks quoted in the auction requested by dealer  $i$ . Manipulated transactions are more noisy than unmanipulated transactions because the worst price is further away from the mean  $Y$  than the best price. The noise term  $\varepsilon_i$  is  $\pm \min\{\varepsilon_i^1, \varepsilon_i^2\}$ , with symmetric probability. Manipulated transactions contain an additional noise term  $z_i = \max\{\varepsilon_i^1, \varepsilon_i^2\} - \min\{\varepsilon_i^1, \varepsilon_i^2\}$ . Thus, we provided a game-theoretic foundation for our assumption that manipulation reduces the signal-to-noise ratio of a benchmark.

In the framework modeled in this section, there is an additional distortionary channel for manipulation, through its impact on the probability distribution of unmanipulated data. When it is more likely that a counterparty in a transaction is a manipulator, a trader might provide a noisy quote, hoping that it will be accepted when the price distortion happens to be of the sign preferred by the manipulator. As a result, even when the quote requester is not a manipulator, and would take the most attractive quote, the distribution of quotes is more dispersed. As the probability  $p_M$  of manipulation rises, the probability distribution  $F$  of quotes shifts towards quotes further away from the true value  $Y$ . Hence the variance of  $\varepsilon_i$  rises, in that  $|\varepsilon_i|$  is distributed according to the CDF  $1 - (1 - F_\varepsilon(\varepsilon))^2$ , where

$$F_\varepsilon(\varepsilon) = 1 - \frac{p_M}{1 - 2p_M} \frac{\Delta - \varepsilon}{\varepsilon},$$

<sup>38</sup>We leave out a description of the equilibrium for the case  $p_M \geq 1/2$  which is less relevant for our application. In that case, we would observe bids above  $Y + \Delta$ .

implying that

$$\sigma_\epsilon^2 = 2\Delta^2 \left( \frac{p_M}{1-2p_M} \right)^2 \left[ -\log \left( 1 + \frac{2p_M-1}{1-p_M} \right) + \frac{2p_M-1}{1-p_M} \right].$$

The noise level  $\sigma_\epsilon^2$  is increasing in  $\Delta$  and  $p_M$ . In particular,  $\lim_{p_M \rightarrow 1/2} \sigma_\epsilon^2 = \Delta^2$ .

In this auction setting, because manipulation adversely impacts the precision of unmanipulated price signals, the slope of the optimal benchmark weighting function  $f$  is lowered in order to mitigate the risk of manipulation. The benchmark designer can affect the distribution of  $\epsilon_i$  by choosing  $f$  so that  $p_M = 1 - H(R_f)$  is relatively low. As a result, the probability of manipulation is smaller than in the baseline model in which the distribution of unmanipulated transaction data is exogenous.

## C Supplemental Material for Section 4

### C.1 The generalized statement of Theorem 1a-1b

We first define a generalization of ODE (4.4). The differential equation is indexed by two parameters: the starting point  $s_0$  and a constant  $\eta > 0$ :

$$f''(s) = - \frac{[\eta - 2f(s)\sigma_U^2] H(R)g(s) + 2\gamma\sigma_M^2 h\left(\frac{\gamma}{f'(s)}\right)}{[2f(s)\sigma_M^2 - \eta] \left(-h'\left(\frac{\gamma}{f'(s)}\right)\right) \frac{\gamma^2}{(f'(s))^3}} \quad (\text{C.1})$$

with boundary conditions  $f(s_0) = (\gamma/R)s_0$ ,  $f'(s_0) = \gamma/R$ .

**Theorem 1** *For any  $R \in (0, \hat{R})$ , there exists a unique optimal solution  $f^*$  to problem  $\mathcal{P}(R)$ . The optimal weighting function  $f^*$  is non-decreasing, concave, continuously differentiable everywhere, and given by*

$$f^*(s) = \begin{cases} \frac{\gamma}{R}s & s \in [0, s_0] \\ \text{solution to (C.1)} & s \in (s_0, s_1) \\ f^*(s_1) & s \in [s_1, \bar{s}] \end{cases}.$$

The parameter  $\eta$  in (C.1) and the cutoff point  $s_1$  are chosen so that  $(f^*)'(\bar{s}) = 0$ : either  $s_1 < \bar{s}$  in which case  $\eta = 2f^*(s_1)\sigma_U^2$ , or  $s_1 = \bar{s}$  in which case  $\eta \in [2f^*(s_1)\sigma_U^2, 2f^*(s_0)\sigma_M^2]$  is chosen so that the solution to (C.1) on  $[s_0, \bar{s}]$  satisfies  $f'(\bar{s}) = 0$ . Finally, the cutoff point  $s_0 \in (0, \bar{s})$  is chosen to satisfy the constraint (4.3).<sup>39</sup>

<sup>39</sup>The existence of such  $\eta$ ,  $s_0$ ,  $s_1$  and the existence of a solution to (C.1) will be proven.

Clearly, Theorem 1 implies Theorem 1a. To see that it also implies Theorem 1b, note that  $f^*$  described by Theorem 1b corresponds exactly to the first case described by Theorem 1:  $s_1 < \bar{s}$  and  $\eta = 2f^*(s_1)\sigma_U^2$ . Because the solution is unique, when  $f^*$  described by Theorem 1b exists, it must be optimal. In this case, the ODE (4.4) is obtained from (C.1) by plugging in the above expression for  $\eta$ , and dividing the numerator and the denominator by  $2\sigma_U^2$ .

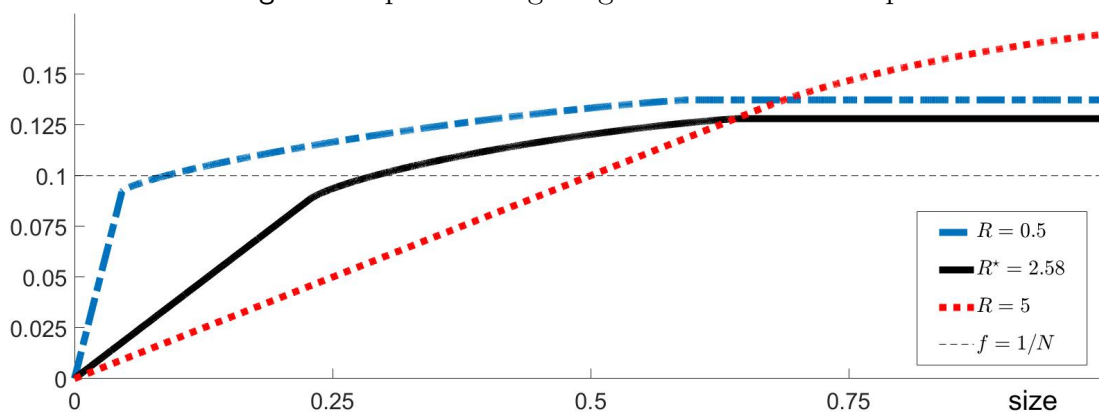
The derivation of the optimal benchmark in Appendix D.1 establishes the generalized version of Theorem 1a-1b described above.

## C.2 Additional examples

In this appendix, we present two additional examples that illustrate the results in Section 4. The first example illustrates Proposition 3.

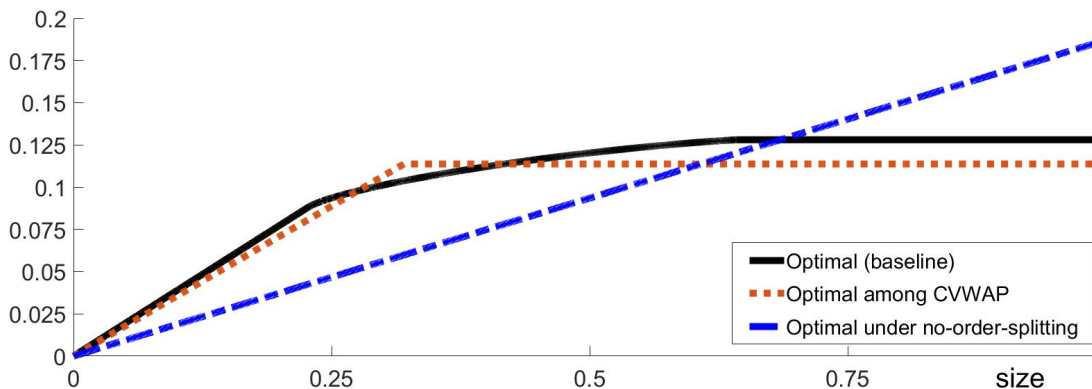
**Example 2** Under the parametric assumptions of Example 1, we can numerically compute the optimal manipulation threshold:  $R^* \approx 2.58$  achieves the minimum for the benchmark administrator's problem  $\mathcal{P}$ . We can then apply Theorem 1b to compute the optimal fixing. Figure C.1 presents the optimal weighting function for  $R = 0.5$ ,  $R = 2.58$ , and  $R = 5$ . The ex-ante probabilities of manipulation under these target levels are approximately 0.78, 0.28, and 0.08, respectively.<sup>40</sup> Figure C.1 shows that it is possible for two feasible weighting functions to never cross. If the distribution of sizes  $\hat{s}_i$  were fixed, this would clearly be impossible because any two such functions could not have the same expectation with respect to the distribution of  $\hat{s}_i$ . However, this is possible when the distribution of  $\hat{s}_i$  depends on the shape of  $f$ .

Fig. C.1: Optimal weighting functions for Example 2



<sup>40</sup>Although Figure C.1 may suggest otherwise, the function corresponding to  $R = 5$  has a zero derivative at  $s = \bar{s}$ . The second derivative gets large close to  $s = \bar{s}$ , so the first derivative changes rapidly in a small neighborhood of  $\bar{s}$ . This is the case in which Theorem 1b does not apply and the solution is described by Theorem 1 in Appendix C.1.

Fig. C.2: Optimal weighting functions for Example 3



The second example is an illustration of Theorem 2.

**Example 3** We adopt the parameters of Examples 1 and 2.<sup>41</sup> The optimal benchmark fixing in the baseline model leads to the threshold  $R^* \approx 2.58$  which induces 28% of agents to manipulate. The optimal benchmark that deters order splitting is that which minimizes the probability of manipulation subject to unbiasedness. This yields a manipulation incentive threshold  $\hat{R}$  of about 5.35, leading to manipulation with a probability of about 7%. The minimized objective function (mean squared error of the estimator) is 0.142 in the baseline case, and 0.19 when restricted to Condition 1, robustness to order splitting. This sharp increase in benchmark noise is caused by attaching a higher weight to manipulated transactions and inefficiently small weight to small unmanipulated transactions.

To put this in context, consider the optimal benchmark fixing in the class of capped volume-weighted average price fixings, those with a weighting function that is linear in transaction size  $s$  up to some maximal transaction size, after which the weight remains constant. The best such fixing has a mean squared error of 0.149 and induces manipulation by an agent in the event that the agent's manipulation incentive  $R$  exceeds 2.81, which has a probability of about 24%. These three weighting functions are depicted in Figure C.2.

## D Proofs

### D.1 Proof of Theorem 1a and 1b

In this appendix, we prove Theorems 1a and 1b. We first sketch the main argument, and then fill in the details by proving a few lemmas.

<sup>41</sup>Because we solve the example numerically, all numerical results reported in this and other examples are approximate.

To solve the problem  $\mathcal{P}(R)$ , we must first determine  $\Psi_f(\cdot)$  for each admissible  $f \in \mathcal{C}^{K,M}$ . This is complicated by the fact that  $f$  need not be well behaved. For example,  $f$  is not necessarily differentiable or even concave. However, we can use the structure of the manipulation problem faced by agents to overcome this difficulty. We do this in a series of Lemmas which establish that the optimal benchmark exists, and the weighting function  $f$  must be continuous, non-decreasing, and concave.

**Lemma 1** *The problem  $\mathcal{P}(R)$  admits a solution for any  $R \leq \hat{R}$ .*

Our proof of this lemma is relatively involved because the standard argument (exploiting upper semi-continuity of the objective function on a compact domain) does not apply directly. The weighting functions are allowed to have jump discontinuities, which can lead to discontinuities in the objective function (especially if a small change in the weighting function induces a large change in the behavior of manipulators) and failure of compactness. We deal with these difficulties by exploiting the special structure of the problem and the regularity conditions imposed on feasible  $f$ . For unmanipulated transactions, due to the continuous distribution of trade sizes, the properties of  $f$  on a measure-zero set (in particular at the finitely many points of discontinuity) are irrelevant. For manipulated transactions, we observe that discontinuities in the choice of the optimal size  $\hat{s}_i$  can occur only in cases for which the manipulator is indifferent between several transaction sizes. However, such cases are non-generic with respect to  $R_i$ . Because  $R_i$  has a continuous distribution, any such cases can be ignored when computing the expected payoff.

A simple corollary of Lemma 1 is that the full problem  $\mathcal{P}$  also admits a solution. This follows from the Maximum Theorem (also known as Berge's Theorem) which implies continuity of the value of the problem  $\mathcal{P}(R)$  in the threshold type  $R$ .

Having established existence, we can derive a series of restrictions on the shape of the optimal weighting function.

**Lemma 2** *If  $f$  is a solution to problem  $\mathcal{P}(R)$ , then  $f$  is non-decreasing.*

The proof of this lemma is technical and thus relegated to the Appendix, but the intuition behind this result is straightforward and instructive. Suppose that a feasible weighting function  $f$  is not non-decreasing. Then we can find an interval  $[s_0, s_1] \subset [0, \bar{s}]$  such that no manipulator chooses a transaction size in this interval. Intuitively, manipulators never choose transactions that give them the same influence on the benchmark as some smaller (hence less costly) transaction. Absent manipulation, however, we saw in Proposition 1 that the optimal weight is constant. Thus, we can modify  $f$  in such an interval so as to retain feasibility but improve the value of the program  $\mathcal{P}$ . This rules out the optimality of  $f$ .

**Lemma 3** *If  $f$  is a solution to problem  $\mathcal{P}(R)$ , then  $f$  is continuous.*

By Lemma 2 and the regularity conditions imposed on any weighting function, we can prove Lemma 3 merely by ruling out cases in which  $f$  jumps up at some  $s_0$ . If there is a jump at  $s_0$ , then there are no manipulations in  $(s_0 - \epsilon, s_0)$  for small  $\epsilon > 0$  because the manipulator can discretely increase the influence on the benchmark by choosing  $s_0$  instead, at a negligibly higher cost. Absent manipulations, the jump in  $f$  is suboptimal because the optimal weight for unmanipulated transactions is constant. So, we can improve on a discontinuous  $f$  by “smoothing it out” in the neighborhood of  $s_0$ .

**Lemma 4** *If  $f$  is a solution to problem  $\mathcal{P}(R)$ , then  $f$  is concave.*

To prove Lemma 4, we use the fact that there can be no manipulations in intervals over which the weighting function  $f$  fails to be concave, that is, where  $f$  lies below some affine function. This follows from the linearity of costs. In such cases, we can modify  $f$  in such an interval without inducing manipulation, so as to improve the weighting of the non-manipulated transactions.

Given Lemmas 1-4, it is without loss of generality that we consider only weighting functions in the set

$$\mathcal{F} = \{f \in \mathcal{C}^{K,M} : f \text{ is continuous, nondecreasing, and concave}\}.$$

The concavity of  $f$  implies that we can use first-order conditions to solve the agent’s manipulation problem. However,  $f$  is not necessarily differentiable, so we use “superdifferential” calculus.<sup>42</sup> We denote by  $\partial f(s)$  the superdifferential of  $f$  at the point  $s$ . A function  $f \in \mathcal{F}$  is superdifferentiable at any point  $s \in (0, \bar{s})$  because  $f$  is concave, and the existence of a superdifferential at 0 and  $\bar{s}$  follows from  $R_i f(s) \leq \gamma s$ , and the fact that  $f$  is non-decreasing. Moreover,  $\partial f(s)$  is a non-decreasing correspondence in the strong set order that is singleton-valued for almost all  $s$ . A transaction size  $\hat{s}_i$  is a global maximum of  $R_i f(s) - \gamma s$  if and only if  $0 \in \partial(R_i f(\hat{s}_i) - \gamma \hat{s}_i)$ , or simply  $\gamma/R_i \in \partial f(\hat{s}_i)$ . If  $f$  is actually differentiable at  $s$ , the condition for optimality boils down to the usual first-order condition  $R_i f'(s) = \gamma$ .

We can now characterize  $\Psi_f(\cdot)$  for any  $f \in \mathcal{F}$ . For some  $s \in [0, \bar{s}]$  and some manipulation

---

<sup>42</sup>See Rockafellar (1970) for the definitions of the subderivative and subdifferential of a convex function. The superderivative and superdifferential have the analogous definitions for a concave function.

threshold  $R$ ,

$$\begin{aligned}
\Psi_f(s) &= \mathbb{P}(\hat{s}_i \leq s \mid |R_i| \geq R) \\
&= \mathbb{P}(\partial f(\hat{s}_i) \geq \partial f(s) \mid R_i \geq R) \\
&= \mathbb{P}\left(\frac{\gamma}{R_i} \geq f'(s^+) \mid R_i \geq R\right) \\
&= \frac{H\left(\frac{\gamma}{f'(s^+)}\right) - H(R)}{1 - H(R)}.
\end{aligned}$$

Here,  $f'(s^+)$  denotes the right derivative of  $f$  at  $s$  and (when applied to sets) the inequality  $\geq$  is the strong set order.<sup>43</sup> Because the right derivative of a concave function is a right-continuous and non-increasing function,  $\Psi_f(\cdot)$  is a well defined cdf. Discontinuities in  $f'$  correspond to atoms in the distribution of manipulated transaction sizes.

The concavity of  $f$  implies that the derivative of  $f$ , whenever it exists, lies below  $\gamma/R$ . Indeed, the derivative is non-increasing and the constraint  $f(s) \leq (\gamma/R)s$  implies that  $f'(0^+) \leq \gamma/R$ . Because  $f$  is non-decreasing, we also know that  $f'(s) \geq 0$ . The inequality  $f(s) \leq (\gamma/R)s$  implies that  $f(0) = 0$ . Once these properties are imposed, the constraint  $f(s) \leq (\gamma/R)s$  is redundant. We will study a relaxed problem in which we do not impose concavity of  $f$ , and instead apply the weaker conditions listed above. We will then verify that the solution to the relaxed problem is concave, validating our approach.

The relaxed problem can be phrased as an optimal control problem in which the control variable is the derivative of  $f$ . This approach is valid because our assumptions and previous analysis imply that  $f$  is absolutely continuous. So, we have

$$\min_{u: u(s) \in [0, \gamma/R]} \int_0^{\bar{s}} f^2(s) \left[ \sigma_U^2 H(R) g(s) ds + \sigma_M^2 dH\left(\frac{\gamma}{u(s)}\right) \right] \quad (\text{D.1})$$

subject to

$$f(0) = 0, \quad f'(s) = u(s), \quad \int_0^{\bar{s}} f(s) \left[ H(R) dG(s) + dH\left(\frac{\gamma}{u(s)}\right) \right] = \frac{1}{n}. \quad (\text{D.2})$$

To solve this problem, we apply a theorem that gives sufficient conditions for a control variable and the associated state variable to be optimal. Because the objective function is quadratic in the state variable  $f$  and the constraint (D.2) is linear in  $f$ , the Hamiltonian is convex in the state variable, implying a unique minimizer.

**Lemma 5** *There exists a unique solution to problem (D.1)-(D.2). The solution is non-*

---

<sup>43</sup>That is, for subsets  $X$  and  $Y$  of the real line,  $X \geq Y$  if for any  $x$  in  $X$  and  $y$  in  $Y$ , we have  $\max\{x, y\} \in X$  and  $\min\{x, y\} \in Y$ .

decreasing, concave, continuously differentiable everywhere, and coincides with  $(\gamma/R)s$  for small  $s$ . Moreover, the solution is given by the function  $f^*$  described by Theorem 1b whenever such a function exists.

Because  $f^*$  solves the relaxed problem (D.1)-(D.2) and is feasible for the original problem  $\mathcal{P}(R)$ ,  $f^*$  is also optimal for the original problem. Thus, the proof of Lemma 5 concludes the proof of Theorem 1a-1b.

### D.1.1 Proof of Lemma 1

Let  $V(f)$  be the value of the problem for a feasible function  $f$ , and let  $V^*$  be the value of the infimum in  $\mathcal{P}(R)$  (it exists because the objective function is bounded by zero from below). By definition of an infimum, there exists a sequence of feasible functions  $f_n$  such that  $\lim_n V(f_n) \rightarrow V^*$ . We have to prove that a subsequence of  $f_n$  converges to a well-defined and feasible limit  $f$ , and that  $V(f) = V^*$ .

First, we define the limiting function  $f$ . Because each  $f_n \in \mathcal{C}^{K,M}$ , we can define  $0 = s_1^n < s_2^n < \dots < s_{K_n-1}^n < s_{K_n}^n = \bar{s}$  such that  $f_n$  is Lipschitz with constant  $M$  on each  $(s_i^n, s_{i+1}^n)$ . Because  $K_n \leq K$ , and  $K$  is finite, there exists a subsequence (which we still denote by  $f_n$ ) such that  $K_n = L \leq K$  for all  $n$ , and  $s_i^n \rightarrow s_i$ . For any  $i$ , take a compact subset  $A_i \subset (s_i, s_{i+1})$ . Then, for large enough  $n$ , the sequence is uniformly bounded and equi-continuous on  $A_i$ , by assumption. By the Arzelá-Ascoli Theorem, we can find a subsequence that converges pointwise to some function  $\tilde{f}^i$  on  $(s_i, s_{i+1})$ , and convergence is uniform on every compact subset. The limiting function  $\tilde{f}^i$  preserves the Lipschitz constant  $M$ . Because the function is bounded and Lipschitz continuous, we can extend the function to  $[s_i, s_{i+1}]$  in such a way that  $\tilde{f}^i$  is continuous. Because there are finitely many  $i$ , we can find a subsequence of  $f_n$  such that the above properties hold for every interval  $(s_i, s_{i+1})$ . Finally, we define  $f$  to be a function that coincides with  $\tilde{f}^i$  on every  $(s_i, s_{i+1})$ , and is equal to  $\max\{\tilde{f}^{i-1}(s_i), \tilde{f}^i(s_i)\}$  for each  $s_i$ . The definition guarantees that  $f$  is upper semi-continuous, and thus belongs to the class  $\mathcal{C}^{K,M}$ . The chosen subsequence of  $f_n$  (which we will again denote by  $f_n$ ) converges to  $f$  uniformly on every compact  $A \subset [0, \bar{s}] \setminus \{s_1, \dots, s_L\}$ .

Second, we look at the properties of  $d\Psi_f(s)$  – the distribution of manipulated trade sizes. In this paragraph, we use  $R_i$  to denote a generic positive exposure type. Let

$$S_f(R_i) = \operatorname{argmax}_{s \in [0, \bar{s}]} \{R_i f(s) - \gamma s\} \equiv \operatorname{argmax}_{s \in [0, \bar{s}]} \left\{ f(s) - \frac{\gamma}{R_i} s \right\}$$

be the set of maximizers in the manipulator's problem. The function  $f(s) - (\gamma/R_i)s$ , defined on a lattice  $[0, \bar{R}] \times [0, \bar{s}]$ , is quasi-supermodular in  $s$ , and has a strict single crossing property in  $(s, R_i)$ . It follows from Milgrom and Shannon (1994) that the set  $S_f(R_i)$  is a complete



sublattice, and any selection  $s_f(R_i) \in S_f(R_i)$  is non-decreasing in  $R_i$ . In particular, this means that  $S_f(R_i)$  is a singleton for almost all  $R_i$ . Define  $\bar{s}_f(R_i) = \max S_f(R_i)$ . Then,  $\Psi_f(s) = \mathbb{P}_{R_i \sim H}(\bar{s}_f(R_i) \leq s)$ .

Third, we argue that  $\lim_n V(f_n) = V(f)$ , and that  $f$  satisfies the constraints of the problem  $\mathcal{P}(R)$ . It is enough to prove that

$$\lim_n \int_0^{\bar{s}} f_n^k(s) g(s) ds = \int_0^{\bar{s}} f^k(s) g(s) ds, \quad k \in \{1, 2\}, \quad (\text{D.3})$$

and

$$\lim_n \int_0^{\bar{s}} f_n^k(s) d\Psi_{f_n}(s) = \int_0^{\bar{s}} f^k(s) d\Psi_f(s), \quad k \in \{1, 2\}, \quad (\text{D.4})$$

Showing (D.3) is straightforward – it follows from the Lebesgue dominated convergence theorem and the fact that  $f_n$  converges to  $f$  almost surely. We prove that (D.4) holds as well. Recalling that we are looking at the problem where agents with  $R_i \geq R$  manipulate, we can write

$$\int_0^{\bar{s}} f^k(s) d\Psi_f(s) = \int_R^{\bar{R}} (f(\bar{s}_f(R_i)))^k \frac{dH(R_i)}{1 - H(R)}.$$

It is therefore enough to prove that  $f_n(\bar{s}_{f_n}(R_i)) \rightarrow f(\bar{s}_f(R_i))$  for almost all  $R_i$ . Intuitively, we have to show that the weight chosen by manipulators changes continuously with the weighting function, for almost all  $R_i$ . For some  $R_i$ , it is clear that the optimal choice can be discontinuous when the manipulator is indifferent between two transaction sizes, but as we saw in the second step of the proof, such situations are non-generic. It is enough to prove that  $\bar{s}_{f_n}(R_i)$  converges to  $\bar{s}_f(R_i)$  for almost all  $R_i$ . Indeed, if this is true, then the only scenario in which  $f_n(\bar{s}_{f_n}(R_i))$  might fail to converge to  $f(\bar{s}_f(R_i))$  is when  $\bar{s}_{f_n}(R_i)$  approaches some  $s_i$  at which  $f$  has a jump, and convergence to  $s_i$  is from the side where  $f$  is lower – however, this would contradict the optimality of  $\bar{s}_{f_n}(R_i)$ . To show convergence of  $\bar{s}_{f_n}(R_i)$  to  $\bar{s}_f(R_i)$  for almost all  $R_i$ , it is enough to prove that the limit of  $\bar{s}_{f_n}(R_i)$  is a solution to the agent's problem at  $f$ . Then, the conclusion follows from the fact that, by step 2 of the proof, the set of solutions is a singleton for almost all  $R_i$ .

Let  $v(f) = \max_{s \in [0, \bar{s}]} (R_i f(s) - \gamma s)$ , for a fixed  $R_i$ . Because the function  $R_i f(s) - \gamma s$  is upper semi-continuous in  $s$ , it is enough to prove that  $v(f_n) \rightarrow v(f)$ . Let  $S_n$  be defined as  $[0, \bar{s}] \setminus \left( \bigcup_{i=1}^L [s_i - \frac{1}{n}, s_i + \frac{1}{n}] \right)$  – we removed each  $s_i$  with some small neighborhood from the domain. Then,  $f_n$  converges uniformly to  $f$  on  $S_n$ . We have, for large enough  $n$ ,

$$\begin{aligned} |v(f_n) - v(f)| &= \left| \max_{s \in [0, \bar{s}]} (R_i f_n(s) - \gamma s) - \max_{s \in [0, \bar{s}]} (R_i f(s) - \gamma s) \right| \\ &\leq \frac{O(1)}{n} + \left| \max_{s \in S_n} (R_i f_n(s) - \gamma s) - \max_{s \in S_n} (R_i f(s) - \gamma s) \right| \leq \frac{O(1)}{n} + R_i \max_{s \in S_n} |f_n(s) - f(s)|, \end{aligned}$$

and the last expression goes to 0 by uniform convergence. Here, the term  $O(1)$  denotes a constant, and the first inequality follows from the fact that all  $f_n$  are uniformly bounded and equi-continuous on each  $(s_i^n, s_{i+1}^n)$  (intuitively, removing a small part of the domain cannot change the value of the optimization problem too much). This concludes the proof.

### D.1.2 Proof of Lemma 2

Take a feasible function  $f$  and suppose it is not nondecreasing. We will prove the result by constructing a different feasible function  $\bar{f}$  that improves the objective function (hence,  $f$  cannot be optimal). By assumption, there exist  $s_0$  and  $s_1$  such that  $s_0 < s_1$ , but  $f(s_0) > f(s_1)$ . Without loss of generality we can assume (making the interval smaller if necessary and using the regularity conditions on  $f$ ) that either (i)  $f$  is strictly decreasing in  $[s_0, s_1]$  or (ii)  $f$  has a jump discontinuity at  $s_0$  and  $f(s)$  is lower than  $f(s_0)$  on  $(s_0, s_1]$ .<sup>44</sup>

Consider case (i). By the choice of  $s_0$  and  $s_1$ , there are no manipulations in  $(s_0, s_1)$ , and this will continue to be true for any function  $f$  that is non-increasing in this interval. We can construct a non-increasing, Lipschitz continuous function  $\hat{f}$  on  $[s_0, s_1]$  with the following properties:  $\hat{f}(s_0) = f(s_0)$ ,  $\hat{f}(s_1) = f(s_1)$ ,  $\int_{s_0}^{s_1} \hat{f}(s)g(s) ds = \int_{s_0}^{s_1} f(s)g(s) ds$  and there exists  $s_2 \in (s_0, s_1)$  such that  $\hat{f}(s) < f(s)$  for  $s \in (s_0, s_2)$  and  $\hat{f}(s) > f(s)$  for  $s \in (s_2, s_1)$ . We then define

$$\bar{f}(s) = \begin{cases} \hat{f}(s) & \text{if } s \in [s_0, s_1] \\ f(s) & \text{otherwise.} \end{cases}$$

By construction,  $\bar{f}$  is feasible (in particular it satisfies the constraint that guarantees an unbiased estimator). The difference in the value of the administrator's objective function  $\mathcal{P}$  under  $\bar{f}$  and  $f$  is (using the fact that there are no manipulations in  $[s_0, s_1]$  under  $\bar{f}$ ),

$$\int_{s_0}^{s_1} (\bar{f}^2(s) - f^2(s)) \sigma_U^2 g(s) ds = \int_{s_0}^{s_1} (\bar{f}(s) - f(s)) \phi(s) g(s) ds,$$

where  $\phi(s) \equiv (\bar{f}(s) + f(s)) \sigma_U^2$  is a strictly decreasing function. By the mean value theorem, there exists  $x \in (s_0, s_1)$  such that

$$\int_{s_0}^{s_1} (\bar{f}(s) - f(s)) \phi(s) g(s) ds = \phi(s_0) \int_{s_0}^x (\bar{f}(s) - f(s)) g(s) ds.$$

But  $\int_{s_0}^x (\bar{f}(s) - f(s)) g(s) ds < 0$  because the integrand is (strictly) negative on  $[s_0, s_2)$ , (strictly) positive on  $(s_2, s_1]$  and  $\int_{s_0}^{s_1} (\bar{f}(s) - f(s)) g(s) ds = 0$ .

Therefore,  $\bar{f}$  is feasible and yields a smaller value of the objective function than does  $f$ .

Now, consider case (ii). We can choose  $s_1$  so that  $f(s) < f(s_0)$  for all  $s \in (s_0, s_1)$

---

<sup>44</sup>We assume here that  $s_0 < \bar{s}$ . The opposite case is very easy to rule out.

but not on any larger interval. By the choice of  $s_1$ , there cannot be any manipulations in  $(s_0, s_1)$  – this is because a manipulator would strictly prefer to choose  $s_0$  over any  $s$  in that interval. Suppose that  $f$  is not (almost everywhere) constant on  $(s_0, s_1)$ . Then, there is a way to improve on  $f$  by replacing it in this interval by a constant  $\bar{f}(s) = \alpha$  with  $\alpha(G(s_1) - G(s_0)) = \int_{s_0}^{s_1} \bar{f}(s)g(s)ds = \int_{s_0}^{s_1} f(s)g(s)ds$ . Indeed, under both  $f$  and  $\bar{f}$ , there are only unmanipulated transactions in the interval  $(s_0, s_1)$ , so the objective function changes by

$$\sigma_U^2 \int_{s_0}^{s_1} (\bar{f}^2(s) - f^2(s))g(s)ds < \sigma_U^2 [G(s_1) - G(s_0)] \left[ \alpha^2 - \left( \frac{\int_{s_0}^{s_1} f(s)g(s)ds}{G(s_1) - G(s_0)} \right)^2 \right] = 0,$$

where the (strict) inequality follows from Jensen's Inequality and the fact that  $f$  is not (almost everywhere) constant. Thus,  $f$  could not be optimal.

Finally, consider the opposite case in which  $f$  is constant (almost everywhere) on  $(s_0, s_1)$ . By definition of  $s_1$ , we must in fact have  $s_1 = \bar{s}$ , and it is without loss of generality to assume that  $f(s) = \beta$  for all  $s > s_0$  for some  $\beta$  (in the opposite case there is a simple way to improve on  $f$ ). Because the construction of  $\bar{f}$  is similar to the previous cases, we only discuss it informally and omit a formal calculation. For  $\epsilon > 0$  small enough,  $\beta + \epsilon < f(s_0)$ , so if we raise  $f(s)$  from  $\beta$  to  $\beta + \epsilon$  on  $(s_0, \bar{s}]$ , this has no influence on the distribution of manipulated trades. To preserve constraint (4.3), we can now lower  $f$  by  $\delta$  on  $[s_0 - \Delta, s_0]$ . This might change the distribution of manipulated trades but only in the direction desired by the administrator – the manipulators are guaranteed to choose lower weights after the modification because, for small enough  $\epsilon$ , trade sizes above  $s_0$  are suboptimal. Define  $\bar{f}$  as a function obtained by modifying  $f$  in a way described above with  $\epsilon$ ,  $\delta$ , and  $\Delta$  such that constraint (4.3) is preserved. Then, for small enough  $\epsilon$  and  $\delta$ ,  $\bar{f}$  achieves a strictly lower value of the objective function than does  $f$ . Hence,  $f$  could not be optimal.

### D.1.3 Proof of Lemma 3

Take a feasible candidate solution  $f$  and suppose that it is not continuous. By the regularity condition and Lemma 2, it is enough to consider the case when  $f$  jumps up at some  $s_0 \in (0, \bar{s})$ . Consider lowering  $f$  by  $\epsilon > 0$  in the interval  $[s_0, \bar{s}]$ , where  $\epsilon$  is small. Note that after this modification, the distribution of manipulated transactions conditional on  $\hat{s} \in [s_0, \bar{s}]$  does not change, but it is possible that some manipulators switch to choosing a size  $\hat{s} < s_0$ . However, we can ignore this in the calculations because, by Lemma 2, the function  $f$  is lower on  $[0, s_0)$  than it is on  $[s_0, \bar{s}]$  (and continues to be lower if  $\epsilon$  is small enough) – hence, this can only improve the objective function. Next, notice that for small enough  $\Delta$  and  $\epsilon$ , there cannot be any manipulations in  $(s_0 - \Delta, s_0)$  because the choice of any  $s$  in this interval is dominated by

the choice of  $s_0$ . Let us define  $\bar{f}$  in the following way

$$\bar{f}(s) = \begin{cases} f(s) & s \leq s_0 - \Delta \\ f(s) + \delta & s \in (s_0 - \Delta, s_0), \\ f(s) - \epsilon & s \geq s_0 \end{cases}$$

where  $\delta$  is chosen so that the constraint (4.3) holds (as noted before, we can ignore the manipulated transactions):

$$\int_{s_0 - \Delta}^{s_0} \delta g(s) ds = \int_{s_0}^{\bar{s}} \epsilon g(s) ds. \quad (\text{D.5})$$

Because of (D.5), the function  $\bar{f}$  is feasible, so we only have to prove that  $\bar{f}$  achieves a lower value of the objective function. We have

$$\begin{aligned} & \int_0^{\bar{s}} \sigma_U^2 \bar{f}^2(s) g(s) ds - \int_0^{\bar{s}} \sigma_U^2 f^2(s) g(s) ds \\ &= \sigma_U^2 \left( \delta^2 \int_{s_0 - \Delta}^{s_0} g(s) ds + 2\delta \int_{s_0 - \Delta}^{s_0} f(s) g(s) ds + \epsilon^2 \int_{s_0}^{\bar{s}} g(s) ds - 2\epsilon \int_{s_0}^{\bar{s}} f(s) g(s) ds \right). \end{aligned}$$

The terms multiplied by  $\epsilon^2$  and  $\delta^2$  can be ignored because they are negligibly small compared to other terms once  $\epsilon$  and  $\delta$  are small enough (they cannot reverse a strict inequality). Using equality (D.5), it is enough to prove that

$$\epsilon \frac{1 - G(s_0)}{G(s_0) - G(s_0 - \Delta)} \int_{s_0 - \Delta}^{s_0} f(s) g(s) ds - \epsilon \int_{s_0}^{\bar{s}} f(s) g(s) ds < 0.$$

or equivalently,

$$\frac{\int_{s_0 - \Delta}^{s_0} f(s) g(s) ds}{G(s_0) - G(s_0 - \Delta)} < \frac{\int_{s_0}^{\bar{s}} f(s) g(s) ds}{1 - G(s_0)}.$$

This means that we are done because

$$\frac{\int_{s_0 - \Delta}^{s_0} f(s) g(s) ds}{G(s_0) - G(s_0 - \Delta)} < \frac{f(s_0^-) + f(s_0)}{2} < \frac{\int_{s_0}^{\bar{s}} f(s) g(s) ds}{1 - G(s_0)},$$

where  $f(s_0^-)$  is the left limit of  $f$  at  $s_0$ , and the inequality follows from the fact that  $f$  is globally non-decreasing, and that there is a jump at  $s_0$ .

#### D.1.4 Proof of Lemma 4

Take a feasible  $f$  and suppose it is not concave. By Lemma 2 and Lemma 3, we can assume that  $f$  is continuous and non-decreasing. This means that we can find an affine increasing function  $\varphi(s) = a + bs$  and an interval  $[s_0, s_1]$  such that  $\varphi(s_0) = f(s_0)$ ,  $\varphi(s_1) = f(s_1)$  and

$\varphi(s) \geq f(s)$  for all  $s \in (s_0, s_1)$ , with a strict inequality for at least some  $\tilde{s} \in (s_0, s_1)$ . We first prove that there can be no manipulations<sup>45</sup> in  $(s_0, s_1)$ . It's enough to show that for generic  $R$ , and for all  $s \in (s_0, s_1)$ ,

$$Rf(s) - \gamma s < \max \{Rf(s_0) - \gamma s_0, Rf(s_1) - \gamma s_1\}.$$

We have

$$\max \{Rf(s_0) - \gamma s_0, Rf(s_1) - \gamma s_1\} = \begin{cases} Ra + (Rb - \gamma) s_1 & \text{if } Rb > \gamma, \\ Ra + (Rb - \gamma) s_0 & \text{if } Rb < \gamma. \end{cases}$$

Take the case  $Rb > \gamma$ . Then we have, for all  $s \in (s_0, s_1)$ ,

$$Rf(s) - \gamma s \leq Ra + (Rb - \gamma) s < Ra + (Rb - \gamma) s_1.$$

Similarly, for  $Rb < \gamma$  and for all  $s \in (s_0, s_1)$ ,

$$Rf(s) - \gamma s \leq Ra + (Rb - \gamma) s < Ra + (Rb - \gamma) s_0.$$

This conclusion depended only on the fact that  $f$  lies below the affine function  $\varphi$ . Thus, if  $f$  cannot be improved upon by another feasible function  $\tilde{f}$ , it must be the case that  $f$  restricted to the interval  $[s_0, s_1]$  arises as a solution to the following optimal control problem:

$$\min_{u \geq 0} \int_{s_0}^{s_1} \tilde{f}^2(s) g(s) ds \tag{D.6}$$

subject to

$$\int_{s_0}^{s_1} \tilde{f}(s) g(s) ds = \int_{s_0}^{s_1} f(s) g(s) ds,$$

$$\tilde{f}'(s) = u(s),$$

$$\tilde{f}(s_0) = f(s_0),$$

$$\tilde{f}(s_1) \leq f(s_1),$$

$$\tilde{f}(s) \leq \varphi(s).$$

Here, the first derivative plays the role of the control variable, and the weighting function is the state variable. Notice that this is a problem mathematically equivalent to that considered in Proposition 2. A standard application of optimal control techniques (see, for example, the

---

<sup>45</sup>Strictly speaking, the measure of manipulations is zero.

Arrow's Theorem on page 107 of [Seierstad and Sydsaeter, 1987](#)) yields the conclusion that the optimal  $\tilde{f}(s)$  is equal to  $\varphi(s)$  up to some  $s_2 \in (s_0, s_1)$ , and is constant equal to  $\tilde{f}(s_2)$  on  $(s_2, s_1]$ , where  $s_2$  is chosen to satisfy the constraint  $\int_{s_0}^{s_1} \tilde{f}(s)g(s) ds = \int_{s_0}^{s_1} f(s)g(s) ds$ . Note that  $s_2 < s_1$  because, by assumption,  $f$  lies strictly below  $\varphi$  for at least some points. Define the function  $\bar{f}$  that coincides with  $f$  outside of the interval  $(s_0, s_1)$  and is equal to the optimal  $\tilde{f}$  otherwise. Then,  $\bar{f}$  achieves a weakly lower value of the objective function than  $f$ , and has a jump discontinuity at  $s_1$ . By Lemma 3,  $\bar{f}$  can be (strictly) improved upon, and hence  $f$  cannot be optimal either.

### D.1.5 Proof of Lemma 5

We will first find a solution to the relaxed problem (D.1) - (D.2), and then prove that it satisfies the properties listed in Theorem 1, a generalization of Theorem 1a-1b found in Appendix C.1. This will establish Theorem 1, and thus Lemma 5 and Theorem 1a-1b as a special case.

We fix an  $R \in (0, \hat{R})$  which guarantees that the set of functions  $f \in \mathcal{F}$  that satisfy the constraints of problem (D.1) - (D.2) is non-empty.

We can simplify the objective function (D.1): Applying integration by parts for the Riemann-Stieltjes Integral, and using the fact that  $f$  is absolutely continuous, we obtain

$$\begin{aligned} \int_0^{\bar{s}} f^2(s) dH\left(\frac{\gamma}{f'(s)}\right) &= f^2(\bar{s}) - 2 \int_0^{\bar{s}} f(s) f'(s) H\left(\frac{\gamma}{f'(s)}\right) ds \\ &= 2 \int_0^{\bar{s}} f(s) f'(s) \left(1 - H\left(\frac{\gamma}{f'(s)}\right)\right) ds. \end{aligned}$$

Therefore, the objective function (D.1) becomes

$$\int_0^{\bar{s}} \left[ f^2(s) \sigma_U^2 H(R) g(s) + 2f(s) f'(s) \sigma_M^2 \left(1 - H\left(\frac{\gamma}{f'(s)}\right)\right) \right] ds.$$

Applying the same method, we get

$$\int_0^{\bar{s}} f(s) dH\left(\frac{\gamma}{f'(s)}\right) = \int_0^{\bar{s}} f'(s) \left(1 - H\left(\frac{\gamma}{f'(s)}\right)\right) ds,$$

which allows us to express the constraint (D.2) as

$$\int_0^{\bar{s}} \left[ f(s) H(R) g(s) + f'(s) \left(1 - H\left(\frac{\gamma}{f'(s)}\right)\right) \right] ds = \frac{1}{n}.$$

Moreover, we can transform the problem into an unconstrained one by defining an auxiliary

state variable  $\Gamma$  by

$$\Gamma(s) = \int_0^s \left[ f(t)H(R)g(t) + f'(t) \left( 1 - H \left( \frac{\gamma}{f'(t)} \right) \right) \right] dt, \quad s \in [0, \bar{s}].$$

This means that

$$\Gamma'(s) = f(s)H(R)g(s) + f'(s) \left( 1 - H \left( \frac{\gamma}{f'(s)} \right) \right)$$

with  $\Gamma(0) = 0$  and  $\Gamma(\bar{s}) = 1/n$ .

We thus have the following optimal control problem:

$$\max_{u: u(s) \in [0, \gamma/R]} - \int_0^{\bar{s}} \left[ f^2(s)\sigma_U^2 H(R)g(s) + 2f(s)u(s)\sigma_M^2 \left( 1 - H \left( \frac{\gamma}{u(s)} \right) \right) \right] ds \quad (\text{D.7})$$

subject to

$$f'(s) = u(s), \quad f(0) = 0, \quad f(\bar{s}) \text{ - free}, \quad (\text{D.8})$$

$$\Gamma'(s) = f(s)H(R)g(s) + u(s) \left( 1 - H \left( \frac{\gamma}{u(s)} \right) \right), \quad \Gamma(0) = 0, \quad \Gamma(\bar{s}) = \frac{1}{n}. \quad (\text{D.9})$$

The Hamiltonian corresponding to the problem is

$$\begin{aligned} \mathcal{H}(f(s), u(s), s) = & - \left[ f^2(s)\sigma_U^2 H(R)g(s) + 2f(s)u(s)\sigma_M^2 \left( 1 - H \left( \frac{\gamma}{u(s)} \right) \right) \right] \\ & + p_1(s)u(s) + p_2(s) \left[ f(s)H(R)g(s) + u(s) \left( 1 - H \left( \frac{\gamma}{u(s)} \right) \right) \right], \quad (\text{D.10}) \end{aligned}$$

where  $p_i(s)$ , for  $i = 1, 2$ , are the multipliers on the two state variables  $f$  and  $\Gamma$ .

The lemma below gives sufficient conditions for optimality and uniqueness of a candidate solution.

**Lemma 6** *Let  $(f(s), u(s))$  be a feasible pair for the problem (D.7) - (D.9). If there exists a continuous and piecewise continuously differentiable function  $p(s) = (p_1(s), p_2(s))$  such that the following conditions are satisfied*

1.  $p_1'(s) = [2f(s)\sigma_U^2 - \eta] H(R)g(s) + 2u(s)\sigma_M^2 \left( 1 - H \left( \frac{\gamma}{u(s)} \right) \right)$ ;
2.  $p_2'(s) = 0$ ;
3.  $u(s)$  maximizes  $\mathcal{H}(f(s), u, s)$  over  $u \in [0, \gamma/R]$  for all  $s \in [0, \bar{s}]$ ;
4.  $p_1(\bar{s}) = 0$ ;
5.  $\hat{H}(f, s) = \max_{u \in [0, \gamma/R]} \mathcal{H}(f, u, s)$  exists and is concave in  $f$  for all  $s$ ,

then  $(f(s), u(s))$  solve the problem (D.7) - (D.9). If  $\hat{H}(f, s)$  is strictly concave in  $f$  for all  $s$ , then  $f$  is the unique solution.

*Proof:* By direct application of the Arrow Sufficiency Theorem (Theorem 5 on page 107 of Seierstad and Sydsaeter, 1987).  $\square$

Before we proceed, we state a simple lemma that will be used throughout.

**Lemma 7** *Suppose  $X$  is a nonnegative random variable with a finite variance and a continuously differentiable decreasing density  $h$  on  $(0, \infty)$ . Then  $\lim_{x \rightarrow \infty} h(x)x^2 = \lim_{x \rightarrow \infty} h'(x)x^3 = 0$ .*

*Proof:* The first claim follows directly from the definition of variance, and the second can be obtained by applying integration by parts.  $\square$

We will construct the functions  $p_1$ ,  $p_2$ , and show that the conditions of Lemma 6 all hold with  $(f, f')$  as described in Theorem 1. (We omit the superscript in  $f^*$  and write  $f$  throughout.) We let  $\eta = p_2(s)$  for all  $s$ , for some constant  $\eta > 0$ . We conjecture that  $\eta \in [2f(s_0)\sigma_U^2, 2f(s_0)\sigma_M^2]$  (we will verify that conjecture later). This definition of  $p_2(s)$  satisfies condition 2 of Lemma 6.

Consider the interval  $[0, s_0]$ , where  $f(s) = (\gamma/R)s$ , and  $u(s) = f'(s) = \gamma/R$ . We want to make sure that condition 3 of Lemma 6 is satisfied:

$$\frac{\gamma}{R} \in \operatorname{argmax}_{u \in [0, \frac{\gamma}{R}]} \left\{ - [2f(s)\sigma_M^2 - \eta] u \left( 1 - H\left(\frac{\gamma}{u}\right) \right) + p_1(s)u \right\}.$$

It is enough to show that the derivative of the objective function with respect to  $u$  is non-negative, for all  $u \in [0, \gamma/R]$ :

$$\left[ \eta - 2\frac{\gamma}{R}s\sigma_M^2 \right] \left[ 1 - H\left(\frac{\gamma}{u}\right) + \frac{\gamma}{u}h\left(\frac{\gamma}{u}\right) \right] + p_1(s) \geq 0. \quad (\text{D.11})$$

Notice that the second derivative with respect to  $u$  is given by

$$\left[ \eta - 2\frac{\gamma}{R}s\sigma_M^2 \right] \left[ \frac{\gamma^2}{u^3} \left( -h'\left(\frac{\gamma}{u}\right) \right) \right]$$

which, by the assumption that  $h$  is decreasing, is non-negative if and only if  $\eta \geq 2(\gamma/R)s\sigma_M^2$ . Thus, the Hamiltonian is convex in the control variable  $u$  (implying a boundary solution) for all  $s$  such that  $\eta \geq 2(\gamma/R)s\sigma_M^2$ , and is concave otherwise. In either case, it is enough to show that (D.11) holds for  $u = \gamma/R$ , i.e., that

$$\left[ \eta - 2\frac{\gamma}{R}s\sigma_M^2 \right] [1 - H(R) + Rh(R)] + p_1(s) \geq 0. \quad (\text{D.12})$$



To satisfy condition 2 of Lemma 6, we set

$$p_1'(s) = \underbrace{\left[2\frac{\gamma}{R}s\sigma_U^2 - \eta\right] H(R)g(s)}_{\leq 0} + 2\frac{\gamma}{R}\sigma_M^2(1 - H(R))$$

in the interval  $s \in [0, s_0]$ , using the assumption that  $\eta \geq 2f(s_0)\sigma_U^2$ . Thus, we can write

$$p_1(s) = p_1(0) + 2\frac{\gamma}{R}s\sigma_M^2(1 - H(R)) + \int_0^s \left[2\frac{\gamma}{R}\tau\sigma_U^2 - \eta\right] H(R)g(\tau)d\tau.$$

To show (D.12), we need to prove that

$$\left[\eta - 2\frac{\gamma}{R}s\sigma_M^2\right] [1 - H(R) + Rh(R)] + p_1(0) + 2\frac{\gamma}{R}s\sigma_M^2(1 - H(R)) + \int_0^s \left[2\frac{\gamma}{R}\tau\sigma_U^2 - \eta\right] H(R)g(\tau)d\tau \geq 0.$$

This is equivalent to

$$\eta[1 - H(R)] + \left[\eta - 2\frac{\gamma}{R}s\sigma_M^2\right] [Rh(R)] + p_1(0) + \int_0^s \left[2\frac{\gamma}{R}\tau\sigma_U^2 - \eta\right] H(R)g(\tau)d\tau \geq 0.$$

The derivative of the left hand side is equal to

$$-2\frac{\gamma}{R}\sigma_M^2Rh(R) + \left(2\frac{\gamma}{R}s\sigma_U^2 - \eta\right) H(R)g(s) \leq 0,$$

as long as  $2(\gamma/R)s\sigma_U^2 \leq \eta$  which is true by conjecture when  $s \leq s_0$ . Thus, we can choose  $p_1(0)$  to satisfy the inequality (D.12) at  $s = s_0$ , and then it will hold on the entire interval  $[0, s_0]$ . We can define  $p_1(0)$  so that the inequality binds at  $s_0$  which gives us

$$p_1(s) = \left[2\frac{\gamma}{R}s\sigma_M^2 - \eta\right] [1 - H(R)] - \left[\eta - 2\frac{\gamma}{R}s_0\sigma_M^2\right] [Rh(R)] - \int_s^{s_0} \left[2\frac{\gamma}{R}\tau\sigma_U^2 - \eta\right] H(R)g(\tau)d\tau.$$

We have thus shown that conditions 1-3 of Lemma 6 all hold in the interval  $[0, s_0]$ .

Next, consider the interval  $[s_0, \bar{s}]$ . In this interval, we want to have an interior maximizer  $u(s)$  of the Hamiltonian (D.10). Because  $\eta \leq 2f(s_0)\sigma_M^2$ , the Hamiltonian is concave in  $u$ , and thus it is enough that the first-order condition holds at  $u = u(s)$  :

$$- [2f(s)\sigma_M^2 - \eta] \left[1 - H\left(\frac{\gamma}{u(s)}\right) + \frac{\gamma}{u(s)}h\left(\frac{\gamma}{u(s)}\right)\right] + p_1(s) = 0. \quad (\text{D.13})$$

Taking the derivative over  $s$ , and using the fact that the equality holds at  $s = s_0$ , this is

equivalent to

$$-2f'(s)\sigma_M^2 \left[ 1 - H\left(\frac{\gamma}{u(s)}\right) + \frac{\gamma}{u(s)}h\left(\frac{\gamma}{u(s)}\right) \right] + [2f(s)\sigma_M^2 - \eta] \frac{\gamma^2}{u^3(s)} h'\left(\frac{\gamma}{u(s)}\right) u'(s) + p_1'(s) = 0. \quad (\text{D.14})$$

Using the fact that  $u(s) = f'(s)$ , so that  $u'(s) = f''(s)$ , and combining (D.14) with the differential equation from condition 1 of Lemma 6 for  $p_1$ , we obtain

$$[2f(s)\sigma_M^2 - \eta] \frac{\gamma^2}{u^3(s)} h'\left(\frac{\gamma}{u(s)}\right) f''(s) = [\eta - 2f(s)\sigma_U^2] H(R)g(s) + 2\gamma\sigma_M^2 h\left(\frac{\gamma}{u(s)}\right).$$

This means that it is enough to show that ODE (C.1) holds whenever  $u(s) > 0$ , and that  $\eta = 2f(s)\sigma_U^2$  whenever  $u(s) = 0$ .

Notice that from the first-order condition (D.13), we have

$$p_1(s) = [2f(s)\sigma_M^2 - \eta] \left[ 1 - H\left(\frac{\gamma}{u(s)}\right) + \frac{\gamma}{u(s)}h\left(\frac{\gamma}{u(s)}\right) \right],$$

and thus condition 4 of Lemma 6 will hold as long as  $u(\bar{s}) = 0$ . Moreover, to show that  $p_1(s)$  is continuous and piecewise continuously differentiable, it is enough to prove that  $u(s)$  is continuous. All of that is accomplished by the following lemma.

**Lemma 8** *There exists  $\eta \in [2f(s_0)\sigma_U^2, 2f(s_0)\sigma_M^2]$ , and a non-decreasing, concave solution  $f$  of class  $C^1$  to the ODE*

$$f''(s) = \phi(s, f, (s) f'(s)) \equiv \begin{cases} -\frac{[\eta - 2f(s)\sigma_U^2]H(R)g(s) + 2\gamma\sigma_M^2 h(\frac{\gamma}{f'(s)})}{[2f(s)\sigma_M^2 - \eta] \left(-h'(\frac{\gamma}{f'(s)})\right) \frac{\gamma^2}{[f'(s)]^3}} & \text{if } f'(s) > 0 \\ 0 & \text{if } f'(s) \leq 0 \end{cases} \quad (\text{D.15})$$

on an interval  $[s_0, \bar{s}]$  with boundary conditions  $f(s_0) = \frac{\gamma}{R}s_0$  and  $f'(s_0) = \frac{\gamma}{R}$  such that  $f'(\bar{s}) = 0$ . Moreover, if  $f'(s_1) = 0$  for some  $s_1 < \bar{s}$ , then  $\eta = 2f(s_1)\sigma_U^2$  (in the opposite case,  $\eta \geq 2f(\bar{s})\sigma_U^2$ ).

*Proof:* In the proof, we will rely on Lemma 7 which implies that the denominator of the ODE (D.15) goes to 0 as  $f'(s) \rightarrow 0$ . Fix a small  $\epsilon > 0$ . We will work with a modified ODE

$$f''(s) = \phi_\epsilon(s, f, (s), f'(s)) \equiv \min \{0, \phi(s, f, (s), \min\{\epsilon, f'(s)\})\}. \quad (\text{D.16})$$

With this modification, the function  $\phi_\epsilon$  is uniformly Lipschitz continuous in  $f$  and  $f'$  (using the assumption that the density  $h$  is twice continuously differentiable). By the Picard-Lindelöf Theorem, there exists a unique solution of class  $C^1$  which we will denote by  $f_{\eta, \epsilon}(s)$ ; moreover,

the solution depends on  $\eta$  in a continuous way. Because the second derivative of  $f_{\eta,\epsilon}(s)$  is non-positive by definition of  $\phi_\epsilon$ , we know that  $f_{\eta,\epsilon}(s)$  is concave.

Next, we will choose  $\eta$  such that  $f'_{\eta,\epsilon}(\bar{s}) = 0$ . When  $\epsilon$  is small enough, and we take  $\eta$  to be close enough to  $2(\gamma/R)s_0\sigma_M^2$ , we have  $f''_{\eta,\epsilon}(s_0) \rightarrow -\infty$ , so the function  $f'_{\eta,\epsilon}(s)$  will hit zero for some  $s < \bar{s}$ , and we will have  $f'_{\eta,\epsilon}(\bar{s}) < 0$ . On the other hand, if we take  $\eta$  low enough, in particular  $\eta < 2(\gamma/R)s_0\sigma_U^2$ , then  $\phi_\epsilon(s, f_{\eta,\epsilon}(s), f'_{\eta,\epsilon}(s)) = 0$ , and hence  $f_{\eta,\epsilon}(s)$  will coincide with  $(\gamma/R)s$ . In this case  $f'_{\eta,\epsilon}(\bar{s}) > 0$ . Thus, there exists an intermediate value  $\eta$  such that  $f'_{\eta,\epsilon}(\bar{s}) = 0$ : Let  $f_\epsilon = f_{\eta,\epsilon}$  for this  $\eta$ . Thus, we have found a solution  $f_\epsilon$  to the modified ODE (D.16) with the property that  $f'_\epsilon(\bar{s}) = 0$ .

Moreover, by the boundary conditions, we can write  $f_\epsilon(s) = \frac{\gamma}{R}s_0 + \int_{s_0}^s f'_\epsilon(t)dt$ , and we know that  $\eta \geq 2f_\epsilon(s)\sigma_U^2$  for all  $s \geq s_0$ , for  $\epsilon$  small enough. Indeed, if this last claim was not true, then by the properties of the function  $\phi_\epsilon$ , we could show that as  $f'_\epsilon$  goes to 0,  $\phi$  becomes positive, and thus  $\phi_\epsilon$  becomes 0. This, however, contradicts the fact that  $f'_\epsilon(\bar{s}) = 0$ . When  $\eta \geq 2f_\epsilon(s)\sigma_U^2$  for all  $s \geq s_0$ , and  $\eta < 2f_\epsilon(s_0)\sigma_M^2$ , then  $\phi(s, f_\epsilon(s), f'_\epsilon(s)) \leq 0$ , so  $f_\epsilon$  is a solution to the ODE

$$f''(s) = \phi(s, f_\epsilon(s), \min\{\epsilon, f'_\epsilon(s)\}). \quad (\text{D.17})$$

This means that we can write  $f'_\epsilon(s)$  as a fixed point of the following operator

$$f'_\epsilon(s) = \Lambda_\epsilon(f'_\epsilon(s)) \equiv \max \left\{ 0, \frac{\gamma}{R} - \int_{s_0}^s \phi_\epsilon \left( t, \frac{\gamma}{R}s_0 + \int_{s_0}^t f'_\epsilon(\tau)d\tau, f'_\epsilon(t) \right) dt \right\}.$$

We want to prove that  $f'(s) = \lim_{\epsilon \rightarrow 0} f'_\epsilon(s)$  exists, and that  $f'$  is a fixed point of the limit operator  $\Lambda = \lim_{\epsilon \rightarrow 0} \Lambda_\epsilon$ . By Tychonoff's Theorem, we can obtain  $f'(s)$  which is a pointwise limit of a subsequence of  $f'_\epsilon(s)$  because  $f'_\epsilon \in [0, \gamma/R]$ . We prove that the limiting function  $f'$  is in fact continuous. The only point at which continuity of  $f'$  might fail is a point  $s_1$  at which  $f''$  diverges to  $-\infty$  (at such a point,  $f'$  could have a jump discontinuity from a positive level to 0). Because  $h \in C^2$ , we can find a number  $B > 0$  such that  $f''_\epsilon(s) \geq -B/f'_\epsilon(s)$  uniformly in  $\epsilon$  and  $s$ . Intuitively,  $f''_\epsilon(s)$  cannot be highly negative unless  $f'_\epsilon(s)$  is close to 0. But this means that  $f'_\epsilon(s) \leq -B/f''_\epsilon(s)$  and in particular  $f'_\epsilon(s) \rightarrow 0$  when  $f''_\epsilon(s) \rightarrow -\infty$ . Therefore,  $f'(s_1) = 0$  if  $f''$  diverges to  $-\infty$  at  $s_1$ , and hence  $f'$  is continuous at  $s_1$ . When a sequence of non-decreasing continuous functions converges pointwise to a continuous (non-decreasing) function, the convergence is uniform. Therefore, we have proven that  $f'_\epsilon \rightrightarrows f'$ . Because the convergence is uniform,  $f'$  is also a fixed point of the limiting functional  $\Lambda$ . Thus, we have obtained a continuous  $f'$  such that

$$f'(s) \equiv \max \left\{ 0, \frac{\gamma}{R} - \int_{s_0}^s \phi \left( t, \frac{\gamma}{R}s_0 + \int_{s_0}^t f'(\tau)d\tau, f'(t) \right) dt \right\},$$

and  $f'(\bar{s}) = 0$ . In particular, this means that whenever  $f' > 0$ ,  $f$  is a solution to the ODE (D.15) (and hence (C.1)).

To finish the proof, we argue that  $\eta \geq 2f(\bar{s})\sigma_U^2$ , and if  $f'(s_1) = 0$  for some  $s_1 < \bar{s}$ , then  $2f(s_1)\sigma_U^2 = \eta$ . The first claim is a consequence of the inequality  $\eta \geq 2f_\epsilon(s)\sigma_U^2$  for every  $\epsilon > 0$  (proven earlier). Suppose that the second claim is not true. Then, because we proved uniform convergence of  $f'_\epsilon$  to  $f'$ , for any  $\delta > 0$ , we can find  $\bar{\epsilon} > 0$  such that for all  $\epsilon < \bar{\epsilon}$ ,

$$\max_{s \in [s_1, \bar{s}]} |f'_\epsilon(s)| < \delta. \quad (\text{D.18})$$

However, when  $2f(s_1)\sigma_U^2 > \eta$ , (so that  $2f_\epsilon(s)\sigma_U^2$  is bounded away from  $\eta$  on  $[s_1, \bar{s}]$ ), this implies that  $-f''_\epsilon$  gets arbitrarily large as  $\delta$  gets small. This is a contradiction with  $f'_\epsilon$  being a solution to ODE (D.17) that at the same time satisfies (D.18).  $\square$

Given Lemma 8, the proof of Lemma 5 is immediate. By taking  $\eta$  whose existence is guaranteed by Lemma 8, we satisfy conditions 1-4 of Lemma 6. The functions  $p_1(s)$  and  $p_2(s)$  are continuous and continuously differentiable by construction (and Lemma 8 which guarantees that  $u(s)$  is continuous everywhere). The function  $\hat{H}(f, s)$  is strictly concave in  $f$  for all  $s$  because the Hamiltonian  $\mathcal{H}$  is a quadratic (strictly concave) function of  $f$ . Finally, we can choose  $s_0$  such that the corresponding  $f$  is feasible, that is, satisfies constraint (D.9), or equivalently, (4.3). Indeed, (i)  $f$  depends on  $s_0$  in a continuous way, (ii) choosing  $s_0 = \bar{s}$  yields  $f(s) = (\gamma/R)s$  which gives  $\Gamma(\bar{s}) > 1/n$  because  $R < \hat{R}$ , and (iii) when  $s_0 \rightarrow 0$ , the corresponding  $f(s)$  also converges to zero pointwise, so  $\Gamma(\bar{s}) < 1/n$ . By the intermediate value theorem, there exists  $s_0 \in (0, \bar{s})$  such that  $\Gamma(\bar{s}) = 1/n$ , that is, constraint (4.3) holds.

This implies that the constructed  $f$  is the unique solution to the problem (D.7) - (D.9). Because this function is feasible for the original problem  $\mathcal{P}(R)$ , it is also the unique solution to  $\mathcal{P}(R)$ .

## D.2 Proof of Proposition 3

We will show that the optimal benchmark fixing with  $R \in \{0, \hat{R}\}$  is dominated by choosing a weighting function of the form

$$f_\beta(s) = \frac{\gamma}{R(\beta)} \max\{s, \beta\},$$

for some  $\beta \in [0, \bar{s}]$ , where  $R(\beta)$  is chosen to make  $f_\beta$  feasible, that is, to satisfy (4.3):

$$\frac{H(R(\beta))}{R(\beta)} \left( \int_0^\beta \gamma \tau g(\tau) d\tau + \gamma \beta (1 - G(\beta)) \right) + \frac{1 - H(R(\beta))}{R(\beta)} \gamma \beta = \frac{1}{n}.$$

As noted in the discussion of Theorem 1a-1b, the optimal weighting function for  $R = 0$  is  $f(s) = 1/n$ , and the optimal weighting function for  $R = \hat{R}$  is  $f(s) = (\gamma/\hat{R})s$ . Importantly, these two functions are the limits of the family  $f_\beta$  as  $\beta$  varies from 0 to  $\bar{s}$ .<sup>46</sup> Moreover,  $R(\beta) \in (0, \hat{R})$  for all  $\beta \in (0, \bar{s})$ . Let

$$V(\beta) = \frac{H(R(\beta))}{R^2(\beta)} \left[ \int_0^\beta \gamma^2 \tau^2 g(\tau) d\tau + \gamma^2 \beta^2 (1 - G(\beta)) \right] \sigma_U^2 + \frac{1 - H(R(\beta))}{R^2(\beta)} \gamma^2 \beta^2 \sigma_M^2$$

denote the value of the objective function (D.1) at  $f_\beta$ . Then,  $V(0)$  corresponds to the value attained by the optimal weighting function with  $R = 0$ , and  $V(\bar{s})$  corresponds to the value attained by the optimal weighting function with  $R = \hat{R}$ . Because  $V$  is continuous and differentiable, to prove Proposition 3, it is enough to show that  $V'(0) < 0$ , and  $V'(\bar{s}) > 0$ .

Using the implicit function theorem, we can write  $R(\beta)$  as a function of  $\beta$  with

$$\lim_{\beta \rightarrow 0} \frac{R(\beta)}{\beta} = \gamma n,$$

and

$$R'(\beta) = \frac{R(1 - H(R(\beta))G(\beta))}{\beta - (H(R(\beta)) - h(R(\beta))R(\beta)) \left( \beta G(\beta) - \int_0^\beta \tau g(\tau) d\tau \right)}.$$

We can calculate the derivative of  $V(\beta)$  at  $\beta = \bar{s}$  directly:

$$\begin{aligned} V'(\bar{s}) &= \frac{h(\hat{R})\hat{R}^2 - 2H(\hat{R})\hat{R}}{\hat{R}^3} \frac{1 - H(\hat{R})}{\bar{s} - (H(\hat{R}) - h(\hat{R})\hat{R}) \left( \bar{s} - \int_0^{\bar{s}} \tau g(\tau) d\tau \right)} \left[ \int_0^{\bar{s}} \tau^2 g(\tau) d\tau \sigma_U^2 - \bar{s}^2 \sigma_M^2 \right] \\ &\quad - \left[ \frac{2H(\hat{R})}{\hat{R}^2} \bar{s} - \frac{2}{\hat{R}^2} \bar{s}^2 \frac{1 - H(\hat{R})}{\bar{s} - (H(\hat{R}) - h(\hat{R})\hat{R}) \left( \bar{s} - \int_0^{\bar{s}} \tau g(\tau) d\tau \right)} + \frac{2}{\hat{R}^2} \bar{s} \right] \sigma_M^2. \end{aligned}$$

If we let  $\lambda = \sigma_M^2/\sigma_U^2$ , then  $V'(\bar{s}) > 0$  is equivalent, after some simplifications, to

$$\left[ 2H(\hat{R}) - h(\hat{R})\hat{R} \right] \left[ \int_0^{\bar{s}} \left( \frac{\tau}{\bar{s}} \right)^2 g(\tau) d\tau \right] < \lambda \left[ h(\hat{R})\hat{R} + 2 \left( H(\hat{R}) - h(\hat{R})\hat{R} \right) \int_0^{\bar{s}} \left( \frac{\tau}{\bar{s}} \right) g(\tau) d\tau \right].$$

We know that the density  $h$  is decreasing, and because a density is integrable, we must have  $\lim_{R \rightarrow 0} h(R)R = 0$ . It follows that  $H(R) > h(R)R$  for all  $R > 0$  because

$$\frac{d}{dR} [H(R) - h(R)R] = h(R) - h(R) - h'(R)R > 0.$$

<sup>46</sup>Formally, we have  $f_{\bar{s}}(s) = (\gamma/\hat{R})s$  and  $\lim_{\beta \rightarrow 0} f_\beta(s) \rightarrow 1/n$  for all  $s > 0$ .

Therefore,  $V'(\bar{s}) > 0$  is equivalent to

$$\lambda > \frac{\left[2H(\hat{R}) - h(\hat{R})\hat{R}\right] \frac{\mathbb{E}s_1^2}{\bar{s}^2}}{h(\hat{R})\hat{R} + 2 \left(H(\hat{R}) - h(\hat{R})\hat{R}\right) \frac{\mathbb{E}s_1}{\bar{s}}}. \quad (\text{D.19})$$

Next, we have

$$\frac{\left[2H(\hat{R}) - h(\hat{R})\hat{R}\right] \frac{\mathbb{E}s_1^2}{\bar{s}^2}}{h(\hat{R})\hat{R} + 2 \left(H(\hat{R}) - h(\hat{R})\hat{R}\right) \frac{\mathbb{E}s_1}{\bar{s}}} \leq \frac{\left[2H(\hat{R}) - h(\hat{R})\hat{R}\right] \frac{\mathbb{E}s_1}{\bar{s}}}{h(\hat{R})\hat{R} + 2 \left(H(\hat{R}) - h(\hat{R})\hat{R}\right) \frac{\mathbb{E}s_1}{\bar{s}}} \leq 1,$$

where the last inequality follows from the fact that, by direct calculation of the derivative, the middle expression is increasing in  $\mathbb{E}s_1/\bar{s}$ . This proves that (D.19) always holds because its left hand side is strictly greater than 1, while the right hand side is less than 1.

Now, we will show that  $V'(0) < 0$ . We have

$$V'(0) = \lim_{\beta \rightarrow 0} V'(\beta) = [\sigma_U^2 - \sigma_M^2] \frac{\gamma^2 h(0)}{n} < 0.$$

This ends the proof.

### D.3 Proof of Theorem 2

We will first show that  $f^*(s) = (\gamma/\hat{R})s$  solves problem  $\mathcal{P}$ . It follows that  $f^*$  solves problem  $\mathcal{P}(R)$  for any  $R \leq \hat{R}$ , because  $f^*$  is feasible for  $\mathcal{P}(R)$  for any  $R \leq \hat{R}$ .

By arguments analogous to the ones used in the proof of Lemma 2 and Lemma 3, the optimal function  $f$  is continuous and non-decreasing. By Bruckner and Ostrow (1962), Assumption 1 is equivalent to the following condition when  $f(0) = 0$ , and  $f$  is non-decreasing and continuous:

$$f'(s^-) \geq \frac{f(s)}{s}, \text{ for all } s \in (0, \bar{s}],$$

where  $f'(s^-)$  denotes the left Dini derivative of  $f$  at  $s$ . Because  $f \in \mathcal{C}^{K,M}$  together with continuity of  $f$  implies that  $f$  is absolutely continuous, we can write that condition as

$$f'(s) \geq \frac{f(s)}{s}, \text{ for a.e. } s \in (0, \bar{s}). \quad (\text{D.20})$$

We first prove that under Assumption 1, all manipulators ( $R_i \geq R$ ) choose  $\hat{s}_i = \bar{s}$ . We have

$$\frac{d}{ds} (Rf(s) - \gamma s) = Rf'(s) - \gamma \geq R \frac{f(s)}{s} - \gamma = s(Rf(s) - \gamma s).$$

This implies that if there exists any  $s > 0$  at which a manipulator can make positive profits,

then that manipulator maximizes profits by choosing  $\hat{s}_i = \bar{s}$ . This implies that the problem to solve is

$$\inf_{f \in \mathcal{C}^{K,M}} \int_0^{\bar{s}} f^2(s) \sigma_U^2 H(R_f) g(s) ds + f^2(\bar{s}) \sigma_M^2 (1 - H(R_f)) \quad (\text{D.21})$$

subject to (D.20), and

$$f(s) \leq \frac{\gamma}{R_f} s, \forall s \in [0, \bar{s}], \quad (\text{D.22})$$

$$\int_0^{\bar{s}} f(s) H(R_f) g(s) ds + f(\bar{s}) (1 - H(R_f)) = \frac{1}{n}. \quad (\text{D.23})$$

Similarly as for the baseline model, we will parameterize the above problem by  $R = R_f$ , and solve it first for any fixed  $R \leq \hat{R}$ .

To simplify the objective function, note that

$$f^2(\bar{s}) = 2 \int_0^{\bar{s}} f(s) f'(s) ds.$$

Similarly, we can express condition (D.23) with  $R = R_f$  as

$$H(R) \int_0^{\bar{s}} f(s) g(s) ds + (1 - H(R)) \int_0^{\bar{s}} f'(s) ds = \frac{1}{n}.$$

To incorporate condition (D.20) into the problem, we will redefine the control variable  $u(s)$  relative to the baseline model. Instead of  $f'(s) = u(s)$ , we let  $u(s) = f'(s) - f(s)/s$ . Constraint (D.20) can now be expressed as  $u(s) \geq 0$ . Thus, the full problem can be written as

$$\min_{u \geq 0} \int_0^{\bar{s}} \left[ f^2(s) \sigma_U^2 g(s) + 2f(s) \sigma_M^2 \left( u(s) + \frac{f(s)}{s} \right) \right] ds \quad (\text{D.24})$$

subject to

$$f'(s) = u(s) + \frac{f(s)}{s}, f(0) = 0, f(\bar{s}) - \text{free} \quad (\text{D.25})$$

$$\Gamma'(s) = H(R) f(s) g(s) + (1 - H(R)) \left( u(s) + \frac{f(s)}{s} \right), \Gamma(0) = 0, \Gamma(\bar{s}) = \frac{1}{n} \quad (\text{D.26})$$

$$f(s) \leq \frac{\gamma}{R} s. \quad (\text{D.27})$$

We conjecture that the constraint  $f(s) \leq (\gamma/R)s$  is slack. We want to prove that the optimal  $f$  is linear:  $f(s) = \alpha s$  for some  $\alpha \leq \gamma/R$ . There exists a unique  $\alpha$  under which a linear  $f$  satisfies the constraint (D.23) (or D.26), namely,  $\alpha = \gamma/\hat{R}$ . Such  $f$  satisfies constraint (D.27), and thus if it solves the relaxed problem, it also solves the original problem.

The Hamiltonian corresponding to the relaxed problem (D.24) - (D.26) is

$$\begin{aligned} \mathcal{H}(f(s), u(s), s) = & - \left[ f^2(s) \sigma_U^2 g(s) + 2\sigma_M^2 f(s) \left( u(s) + \frac{f(s)}{s} \right) \right] \\ & + p_1(s) \left( u(s) + \frac{f(s)}{s} \right) + p_2(s) \left( H(R) f(s) g(s) + (1 - H(R)) \left( u(s) + \frac{f(s)}{s} \right) \right). \end{aligned} \quad (\text{D.28})$$

We state sufficient conditions for a function  $f$  to be optimal, using Arrow's Theorem.

**Lemma 9** *Let  $(f(s), u(s))$  be a feasible pair for the problem (D.24) - (D.26). If there exists a continuous and piecewise continuously differentiable function  $p(s) = (p_1(s), p_2(s))$  such that the following conditions are satisfied*

1.  $p'_1(s) = 2f(s)\sigma_U^2 g(s) + 2\sigma_M^2 \left( u(s) + 2\frac{f(s)}{s} \right) - p_1(s)\frac{1}{s} - p_2(s) \left( H(R)g(s) + (1 - H(R))\frac{1}{s} \right)$ ;
2.  $p'_2(s) = 0$ ;
3.  $u(s)$  maximizes  $\mathcal{H}(f(s), u, s)$  over  $u \geq 0$  for all  $s \in [0, \bar{s}]$ ;
4.  $p_1(\bar{s}) = 0$ ;
5.  $\hat{H}(f, s) = \max_{u \in [0, \gamma/R]} \mathcal{H}(f, u, s)$  exists and is concave in  $f$  for all  $s$ ,

then  $(f(s), u(s))$  solve the problem (D.24) - (D.26). If  $\hat{H}(f, s)$  is strictly concave in  $f$  for all  $s$ , then  $f$  is the unique solution.

*Proof:* By direct application of the Arrow Sufficiency Theorem (Theorem 5 on page 107 of Seierstad and Sydsaeter, 1987).  $\square$

Since we want to prove that  $f(s) = \alpha s$  is optimal, we have  $u(s) = 0$  for all  $s \in [0, \bar{s}]$ . The Hamiltonian is maximized at  $u = 0$  across feasible  $u \geq 0$  when

$$-2\sigma_M^2 f(s) + p_1(s) + p_2(s)(1 - H(R)) \leq 0.$$

We can set  $p_2(s) = \eta$  for some constant  $\eta$  for all  $s$  (this will satisfy condition 2 of Lemma 9). The Hamiltonian is strictly concave in  $f$ . Thus, to satisfy all conditions of Lemma 9, it is enough to prove that there exists a continuously differentiable  $p(s)$  (we abuse notation slightly by dropping the subscript from  $p_1(s)$ ) and a constant  $\eta$  such that

$$p(s) + \eta(1 - H(R)) \leq 2\sigma_M^2 \alpha s, \quad (\text{D.29})$$

$$p(\bar{s}) = 0, \quad (\text{D.30})$$

$$p'(s) + p(s)\frac{1}{s} = 2\alpha s \sigma_U^2 g(s) + 4\sigma_M^2 \alpha - \eta H(R) g(s) - \eta(1 - H(R))\frac{1}{s}, \quad (\text{D.31})$$



for all  $s \in [0, \bar{s}]$ . Solving the ODE (D.31), we obtain

$$p(s) = \frac{1}{s} \left( \kappa + \int_0^s [2\alpha\tau^2\sigma_U^2g(\tau) + 4\sigma_M^2\alpha\tau - \eta H(R)g(\tau)\tau - \eta(1 - H(R))] d\tau \right),$$

for all  $s > 0$ , and some constant  $\kappa$ . With the final condition (D.30), we obtain

$$p(s) = -\frac{1}{s} \int_s^{\bar{s}} [2\alpha\tau^2\sigma_U^2g(\tau) + 4\sigma_M^2\alpha\tau - \eta H(R)g(\tau)\tau - \eta(1 - H(R))] d\tau.$$

This means in particular that  $p(s)$  is well defined and continuously differentiable for all  $s \in (0, \bar{s}]$ . To guarantee that we can define  $p(0)$  so that  $p(s)$  is continuous at  $s = 0$ , we need

$$\int_0^{\bar{s}} [2\alpha\tau^2\sigma_U^2g(\tau) + 4\sigma_M^2\alpha\tau - \eta H(R)g(\tau)\tau - \eta(1 - H(R))] d\tau = 0. \quad (\text{D.32})$$

Condition (D.32) is also sufficient: By d'Hospital rule, if (D.32) holds, then the limit  $\lim_{s \searrow 0} p(s)$  exists and is finite. Condition (D.32) pins down a unique candidate for  $\eta$ :

$$\eta = \frac{\int_0^{\bar{s}} [2\alpha\tau^2\sigma_U^2g(\tau) + 4\sigma_M^2\alpha\tau] d\tau}{\int_0^{\bar{s}} [H(R)g(\tau)\tau + (1 - H(R))] d\tau}.$$

With  $\eta$  defined this way, and after simplifying the expressions, (D.29) becomes equivalent to

$$\begin{aligned} \frac{\int_0^{\bar{s}} [\tau^2\sigma_U^2g(\tau) + 2\sigma_M^2\tau] d\tau}{\int_0^{\bar{s}} [H(R)g(\tau)\tau + (1 - H(R))] d\tau} \left[ \int_s^{\bar{s}} \tau H(R)g(\tau)d\tau + (1 - H(R))\bar{s} \right] \\ \leq \sigma_M^2\bar{s}^2 + \int_s^{\bar{s}} \tau^2\sigma_U^2g(\tau)d\tau, \end{aligned}$$

for all  $s \in [0, \bar{s}]$ . Equivalently, after some simplifications,

$$\begin{aligned} H(R) \left( \int_0^{\bar{s}} \tau g(\tau)d\tau \right) \left( \int_0^s \tau^2\sigma_U^2g(\tau)d\tau \right) + (1 - H(R))\bar{s} \int_0^s \tau^2\sigma_U^2g(\tau)d\tau \\ \leq \sigma_M^2(1 - H(R))\bar{s}^2 \int_0^s \tau g(\tau)d\tau + H(R) \left( \int_0^{\bar{s}} \tau^2\sigma_U^2g(\tau)d\tau \right) \left( \int_0^s \tau g(\tau)d\tau \right), \end{aligned}$$

for all  $s \in [0, \bar{s}]$ . Because the above expression is linear in  $H(R)$ , and  $R$  does not appear anywhere else in the expression, it is enough to show that it holds for both  $H(R) = 0$  and  $H(R) = 1$ . That is, it is enough to show that

$$\frac{\int_0^{\bar{s}} \tau g(\tau)d\tau}{\int_0^{\bar{s}} \tau^2\sigma_U^2g(\tau)d\tau} \leq \frac{\int_0^s \tau g(\tau)d\tau}{\int_0^s \tau^2\sigma_U^2g(\tau)d\tau}, \quad (\text{D.33})$$

and

$$\int_0^s \tau^2 \sigma_U^2 g(\tau) d\tau \leq \bar{s} \int_0^s \tau \sigma_M^2 g(\tau) d\tau, \quad (\text{D.34})$$

for all  $s \in [0, \bar{s}]$ . Inequality (D.34) is clearly true because  $\sigma_U^2 < \sigma_M^2$  by assumption. To prove inequality (D.33), it is enough to show that

$$\frac{\int_0^s \tau g(\tau) d\tau}{\int_0^s \tau^2 \sigma_U^2 g(\tau) d\tau}$$

is decreasing in  $s$ . By calculating the derivative, we can show that a sufficient condition is  $\int_0^s [\tau - s] \tau g(\tau) d\tau \leq 0$  for all  $s$ , which is clearly satisfied. This ends the proof that conditions (D.29) – (D.31) all hold.

Therefore, all conditions of Lemma 9 also hold, and thus we have proven that  $f(s) = (\gamma/\hat{R})s$  is the unique solution to the relaxed problem (D.24) – (D.26) for any  $R \leq \hat{R}$ , and hence also the problem (D.24) – (D.27). It follows that the same  $f$  solves the problem  $\mathcal{P}$  and  $\mathcal{P}(R)$  for any  $R \leq \hat{R}$ .

The equality  $(\gamma/\hat{R})s = s/(n\mathbb{E}[\hat{s}_i])$  holds by the definition of  $\hat{R}$ .

## D.4 Proof of Theorem 3

It is immediate to observe that a VWAP benchmark prevents pairwise order-splitting and assigns no weight to transactions with zero size. It remains to prove the opposite direction of the equivalence. We fix an arbitrary fractional-weight fixing  $(f_i)_{i=1}^n$  that neutralizes pairwise order-splitting and assigns no weight to transactions with zero size.

The total weight function  $f_K$  for a subset  $K \subset N$  is defined as

$$f_K(\mathbf{s}) = \sum_{i \in K} f_i(\mathbf{s}).$$

The assumption that pairwise order splitting is neutralized implies that the total weight function for a binary set  $K$  only depends on the sum of the corresponding trade sizes. That is, for any  $i$  and  $j$  in  $N$ ,

$$f_{\{i,j\}}(\mathbf{s}) = \tilde{f}_{\{i,j\}}(s_i + s_j, \mathbf{s}_{-ij}),$$

for some  $\tilde{f}_{\{i,j\}} : \mathbb{R}_+^{n-1} \rightarrow \mathbb{R}_+$ . We will prove that this property can be generalized to an arbitrary group  $K$ .

**Lemma 10** *For any  $K \subseteq N$ , there exists some  $\tilde{f}_K : \mathbb{R}_+^{n-|K|+1} \rightarrow \mathbb{R}_+$  such that*

$$f_K(\mathbf{s}) = \tilde{f}_K \left( \sum_{i \in K} s_i, \mathbf{s}_{-K} \right).$$

*Proof:* We prove the lemma by induction over the cardinality of  $K$ . By assumption, the claim is true for  $|K| = 2$ . Suppose that we have proven this property whenever  $K$  has cardinality  $k - 1$ . We now prove it for any  $K$  with  $|K| = k$ .

We have

$$f_K(\mathbf{s}) = \sum_{i \in K} f_i(\mathbf{s}) = \frac{1}{k-1} \sum_{L \subset K, |L|=k-1} f_L(\mathbf{s}) = \frac{1}{k-1} \sum_{L \subset K, |L|=k-1} \tilde{f}_L \left( \sum_{i \in L} s_i, \mathbf{s}_{-L} \right),$$

for some  $\tilde{f}_L : \mathbb{R}_+^{n-|K|+2} \rightarrow \mathbb{R}_+$ , which exists by the inductive hypothesis.

Next, for each  $L \subset K$  with  $|L| = k - 1$ , there exists some  $\hat{f}_L : \mathbb{R}_+^{n-|K|+2} \rightarrow \mathbb{R}_+$  such that

$$\tilde{f}_L \left( \sum_{i \in L} s_i, \mathbf{s}_{-L} \right) = (k-1) \hat{f}_L \left( \sum_{i \in K} s_i, \mathbf{s}_{-L} \right).$$

This is true because there exists a one-to-one mapping between  $(\sum_{i \in L} s_i, \mathbf{s}_{-L})$  and  $(\sum_{i \in K} s_i, \mathbf{s}_{-L})$  for any such  $L$ .

Fixing  $\sum_{i \in K} s_i = S$ , we must show that

$$f_K(\mathbf{s}) = \sum_{L \subset K, |L|=k-1} \hat{f}_L(S, \mathbf{s}_{-L})$$

is invariant to  $\mathbf{s}_K$ . Formally, we can fix  $\mathbf{s}_{-K}$  (the same argument can be made contingent on any fixed profile  $\mathbf{s}_{-K}$ ), and consider any function  $\hat{f}_L$ , whose dependence on  $\mathbf{s}_{-K}$  and  $S$  is suppressed in the notation, on the linear subspace  $M = \{\mathbf{s} : \sum_{i \in K} s_i = S\}$ ; we must prove that

$$f_K(\mathbf{s}) = \sum_{i \in K} f_i(\mathbf{s}) = \sum_{i \in K} \hat{f}_{-i}(s_i)$$

is constant on  $M$ , where  $\hat{f}_{-i}(s_i)$  is the suppressed notation for

$$\hat{f}_{-(K \setminus \{i\})}(S, (s_i, \mathbf{s}_{-K})),$$

using the fact that  $L = K \setminus \{i\}$  for some  $i \in K$ .

Consider any two indices  $i, j \in K$ . Because  $|K| \geq 3$ , a restriction that  $s_i + s_j$  is a constant is a restriction of  $\mathbf{s}$  to an  $l$ -dimensional linear subspace of  $M$ , for some  $l > 0$ . This restriction implies that the total weight  $f_i(\mathbf{s}) + f_j(\mathbf{s})$ , and hence  $f_K(\mathbf{s})$ , must be invariant. But this implies that  $\hat{f}_{-i}(s_i) + \hat{f}_{-j}(s_j)$  can depend only on the sum  $s_i + s_j$ , which is only possible when

$$\hat{f}_{-i}(s_i) = \alpha s_i + \beta_{-i},$$

for some constants  $\alpha$  and  $\beta_{-i}$ . Applying this argument to any pair of transactions, we conclude that

$$f_K(\mathbf{s}) = \sum_{i \in K} f_i(\mathbf{s}) = \alpha \sum_{i \in K} s_i + \sum_{i \in K} \beta_{-i} = \alpha S + \sum_{i \in K} \beta_{-i},$$

which proves that indeed the function  $f_K$  is constant on the subspace  $M$ . This in turn proves that we can write

$$f_K(\mathbf{s}) = \tilde{f}_K \left( \sum_{i \in K} s_i, \mathbf{s}_{-K} \right),$$

which finishes the proof of the induction step and the lemma.  $\square$

We can now apply Lemma 10 with  $K = N \setminus \{i\}$ , for an arbitrary  $i \in N$ . Since the benchmark is fractional, we have

$$f_i(\mathbf{s}) + f_{N \setminus \{i\}}(\mathbf{s}) = 1. \quad (\text{D.35})$$

By the lemma, we have

$$f_{N \setminus \{i\}}(\mathbf{s}) = \tilde{f}_{-i} \left( \sum_{j \neq i} s_j, s_i \right),$$

for some  $\tilde{f}_{-i} : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ , and moreover, by the same argument that we have used in the proof of the inductive step,

$$\tilde{f}_{-i} \left( \sum_{j \neq i} s_j, s_i \right) = \alpha_i \left( s_i, \sum_{j \neq i} s_j \right) \sum_{j \neq i} s_j + \beta_i \left( s_i, \sum_{j \neq i} s_j \right),$$

for some functions  $\alpha_i$  and  $\beta_i$ . Using equality (D.35) and abusing notation slightly by redefining  $\beta_i$ , we obtain

$$f_i(\mathbf{s}) = \alpha_i \left( s_i, \sum_{j \neq i} s_j \right) \sum_{j \neq i} s_j + \beta_i \left( s_i, \sum_{j \neq i} s_j \right).$$

In particular, this implies that there exists a function  $\gamma_i : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$  such that

$$f_i(\mathbf{s}) = \gamma_i \left( s_i, \sum_{j \in N} s_j \right).$$

Fixing  $\sum_{j \in N} s_j = S$  and using once more the condition that the benchmark neutralizes pairwise order-splitting, we obtain that for any  $k \neq i$ ,

$$f_i(\mathbf{s}) + f_k(\mathbf{s}) = \gamma_i(s_i, S) + \gamma_k(s_k, S)$$

can depend only on  $s_i + s_k$ . But this is only possible when, abusing notation slightly,

$$\gamma_i(s_i, S) = \alpha(S)s_i + \beta_i(S),$$

and

$$\gamma_k(s_k, S) = \alpha(S)s_k + \beta_k(S).$$

That is, the functions  $\alpha_i$  and  $\beta_i$  defined earlier depend only on the sum  $S$  of transaction sizes, and moreover  $\alpha_i$  does not depend on  $i$  and can be expressed as some function  $\alpha$ .

Since the benchmark is fractional, we must have, whenever  $\sum_{j \in N} s_j = S$ ,

$$1 = \sum_{i \in N} f_i(\mathbf{s}) = \sum_{i \in N} (\alpha(S)s_i + \beta_i(S)) = \alpha(S)S + \sum_{i \in N} \beta_i(S),$$

from which it follows that

$$\alpha(S) = \frac{1 - \sum_{i \in N} \beta_i(S)}{S}.$$

Therefore, redefining  $\beta_i$  as  $n\beta_i$  to simplify notation, we conclude that robustness to pairwise order splitting implies that the benchmark fixing weight must take the form (5.1).

Finally, we apply the condition that  $f_i(0, \mathbf{s}_{-i}) = 0$  for all  $\mathbf{s}_{-i}$ . In particular, it must be that for all  $S$ ,

$$\left(1 - \frac{1}{n} \sum_{i \in N} \beta_i(S)\right) \left(\frac{0}{S}\right) + (\beta_i(S)) \left(\frac{1}{n}\right) = \beta_i(S) \left(\frac{1}{n}\right) = 0,$$

and hence that function  $\beta_i(\cdot)$  is identically zero. We conclude that

$$f_i(\mathbf{s}) = \frac{s_i}{\sum_{j=1}^n s_j},$$

which finishes the proof.

## D.5 Proof of Proposition 4

Let  $\hat{Y}$  be a fractional-weight fixing with weights  $(f_i)_{i=1}^n$ . We can then calculate the mean squared error as

$$V \equiv \mathbb{E}[(Y - \hat{Y})^2] = \mathbb{E} \left[ \left( Y \left( 1 - \sum_{i=1}^n f_i(\hat{\mathbf{s}}) \right) - \sum_{i=1}^n f_i(\hat{\mathbf{s}}) (\epsilon_i + \hat{z}_i) \right)^2 \right],$$

where expectation is taken with respect to the distributions of the primitive random variables and equilibrium strategies of agents. Because weights add up to 1, the first term in brackets

drops out, and we obtain

$$V = \sum_{i=1}^n \mathbb{E} [f_i^2(\hat{\mathbf{s}})(\epsilon_i + \hat{z}_i)^2] + \sum_{i \neq j} \mathbb{E} [f_i(\hat{\mathbf{s}})f_j(\hat{\mathbf{s}})(\epsilon_i + \hat{z}_i)(\epsilon_j + \hat{z}_j)]$$

The second term drops out because even conditional on  $\hat{\mathbf{s}}$  and  $(\epsilon_i + \hat{z}_i)$ ,  $(\epsilon_j + \hat{z}_j)$  for  $j \neq i$  has a distribution that is symmetric around 0.

Observe that  $(\epsilon_i + \hat{z}_i)$  is independent of  $\hat{\mathbf{s}}_{-i}$ ; therefore, by the law of iterated expectations and Jensen's inequality,

$$\begin{aligned} V &= \sum_{i=1}^n \mathbb{E} [\mathbb{E} [f_i^2(\hat{\mathbf{s}}) | \hat{s}_i] (\epsilon_i + \hat{z}_i)^2] > \sum_{i=1}^n \mathbb{E} [(\mathbb{E} [f_i(\hat{\mathbf{s}}) | \hat{s}_i])^2 (\epsilon_i + \hat{z}_i)^2] \\ &= \sum_{i=1}^n \mathbb{E} [\bar{f}_i^2(\hat{s}_i)(\sigma_\epsilon^2 + \sigma_z^2)] \equiv \underline{V}, \end{aligned}$$

where  $\bar{f}_i(\hat{s}_i)$  is the interim expected weight faced by agent  $i$ . Jensen's inequality holds in its strict version because weights can add up to 1 only if they depend on all transaction sizes, and because the quadratic function is strictly convex.

Now consider an absolute-size-weight fixing  $\tilde{Y}$  defined by the interim expected weights associated with the original fractional-weight fixing, that is, define  $\bar{f}_i(\hat{s}_i) = \mathbb{E} [f_i(\hat{s}_i, \hat{\mathbf{s}}_{-i}) | \hat{s}_i]$ . When these absolute-size weights are used instead of the fractional-weight fixing, the BNE between the agents remains the same, since the best response of any agent  $i$  in this equilibrium only depends on the interim expected weight on  $i$ 's transaction (see Section 3). Moreover, this equilibrium is now in dominant strategies, since the expected weight that agent  $i$  receives no longer depends on the behavior of other agents. And because the original weights add up to 1 ex-post, the expected weights in  $\tilde{Y}$  must lie between zero and one and add up to 1 in expectation; in particular, the absolute-size-weight fixing is unbiased.

We can thus calculate the mean squared error associated with  $\tilde{Y}$ :

$$\begin{aligned} \mathbb{E} [(\tilde{Y} - Y)^2] &= \left( \mathbb{E} \left[ \sum_{i=1}^n \bar{f}_i(\hat{s}_i) \right]^2 - 1 \right) \sigma_Y^2 + \sum_{i=1}^n \mathbb{E} [\bar{f}_i^2(\hat{s}_i)(\sigma_\epsilon^2 + \sigma_z^2)] \\ &= \left( \sum_{i=1}^n \mathbb{E} \bar{f}_i^2(\hat{s}_i) - \frac{1}{n} \right) \sigma_Y^2 + \sum_{i=1}^n \mathbb{E} [\bar{f}_i^2(\hat{s}_i)(\sigma_\epsilon^2 + \sigma_z^2)] \end{aligned}$$

Because the expected weight lies between zero and one, we can estimate

$$\mathbb{E} \bar{f}_i^2(\hat{s}_i) \leq \mathbb{E} \bar{f}_i(\hat{s}_i) = 1/n.$$

Therefore,

$$\mathbb{E} \left[ (\tilde{Y} - Y)^2 \right] \leq \frac{n-1}{n} \sigma_Y^2 + \underline{V}.$$

We have thus found an unbiased absolute-size-weight benchmark whose mean squared error does not exceed the mean squared error  $V$  of the fractional-weight fixing by more than  $\delta_n \sigma_Y^2$ , where  $\delta_n < 1$ .

If moreover  $\hat{Y}$  is asymptotically efficient, then it must be that  $\underline{V} \rightarrow 0$  as  $n \rightarrow \infty$ , so in particular  $\mathbb{E} \bar{f}_i^2(\hat{s}_i)$  must also converge to 0. Thus, when  $\hat{Y}$  is asymptotically efficient, we have that  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ .

## D.6 Proof of Proposition 5

Let  $\Lambda(\hat{s}_i)$  be the cdf of the transaction size  $\hat{s}_i$  (potentially manipulated) under the optimal A-VWAP benchmark from Section 4.2, that is:

$$\Lambda(s) = H(\hat{R})G(s) + (1 - H(\hat{R}))\mathbf{1}_{\{s \geq \bar{s}\}}.$$

By the Matthews-Border condition (see [Matthews, 1984](#), and [Border, 1991](#)), there exist weights  $f_i(\hat{s}_1, \dots, \hat{s}_n)$  such that  $\sum_{i=1}^n f_i(\hat{s}_1, \dots, \hat{s}_n) = 1$  for almost all  $\hat{s}_1, \dots, \hat{s}_n$  and  $\bar{f}_i(\hat{s}_i) = f^*(\hat{s}_i)$  if and only if (we use the fact that  $f^*(s)$  is strictly increasing)

$$\int_{\tau}^{\bar{s}} f^*(s) d\Lambda(s) \leq \frac{1 - \Lambda^n(\tau)}{n}, \forall \tau \in [0, \bar{s}]$$

with equality for  $\tau = 0$ . For  $\tau = 0$ , we have

$$\int_0^{\bar{s}} \frac{s}{n\mathbb{E}[\hat{s}_i]} d\Lambda(s) = \frac{1}{n},$$

so the equality indeed holds. For any  $\tau > 0$ , we want to prove that

$$L(\tau) \equiv \frac{H(\hat{R}) \int_{\tau}^{\bar{s}} s dG(s) + (1 - H(\hat{R}))\bar{s}}{H(\hat{R}) \int_0^{\bar{s}} s dG(s) + (1 - H(\hat{R}))\bar{s}} + H^n(\hat{R})G^n(\tau) \leq 1.$$

First, we argue that the inequality holds for  $\tau \in (0, \epsilon)$  for some  $\epsilon > 0$ . Taking the derivative of the left hand side with respect to  $\tau$  yields (assuming  $n \geq 3$ )

$$L'(\tau)|_{\tau=0} = 0$$

and

$$L''(\tau)|_{\tau=0} \leq 0.$$

Therefore,  $L(\tau) = 1 - L''(0)\tau^2 + o(\tau^2)$ , and in particular  $L(\tau) \leq 1$  for  $\tau \in (0, \epsilon)$  for some  $\epsilon > 0$ . Now, we show the inequality for  $\tau \geq \epsilon > 0$ . Denote

$$\delta = \frac{H(\hat{R}) \int_{\epsilon}^{\bar{s}} sdG(s) + (1 - H(\hat{R}))\bar{s}}{H(\hat{R}) \int_0^{\bar{s}} sdG(s) + (1 - H(\hat{R}))\bar{s}} < 1,$$

where the inequality follows from the fact that  $G$  has a positive density. Take  $n$  high enough so that  $H^n(\hat{R}) < 1 - \delta$ . Then we have

$$\frac{H(\hat{R}) \int_{\tau}^{\bar{s}} sdG(s) + (1 - H(\hat{R}))\bar{s}}{H(\hat{R}) \int_0^{\bar{s}} sdG(s) + (1 - H(\hat{R}))\bar{s}} + H^n(\hat{R})G^n(\tau) \leq \delta + 1 - \delta \leq 1$$

which finishes the proof of the first part.

To prove the conclusion for the uniform distribution  $g$  of unmanipulated transaction sizes and any  $n \geq 2$ , note that in this case

$$L'(\tau) = -\frac{2H(\hat{R})\frac{\tau}{\bar{s}}}{2 - H(\hat{R})} + H^n(\hat{R})n \left(\frac{\tau}{\bar{s}}\right)^{n-1}.$$

When the Matthews-Border inequality holds for  $n = 2$ , then it holds for any  $n > 2$  as well, so it is enough to show that  $L'(\tau) \leq 0$  for all  $\tau$  which is equivalent to

$$H(\hat{R}) \leq \frac{1}{2 - H(\hat{R})}$$

which holds for any  $H(\hat{R}) \in [0, 1]$ .