On Gatheral's model: revision, reformulation and extensions

Marek Musiela and Thaleia Zariphopoulou

May 1, 2022

Abstract

We revisit Gatheral's model of optimal execution and provide new insights on key quantities and their properties, like the distribution and variance of the implementation slippage, the symmetry of the optimal policy, and the model's behavior in terms of the decay function.

We propose two modifications of the model. The first is introduced in order to correctly define the admissible strategies in Gatheral's model within the execution horizon. In the second modification, we reformulate the control strategies via probability measures whose cdfs model the cumulative fraction of the volume traded through time. For this modified model, we study the admissible and optimal policies in detail.

We examine the performance of the model in terms of the decay function and point out a problematic feature, namely, a possible degeneracy of the implementation slippage as the decay function dissipates. We discuss two possible ways to remedy this, namely, either by imposing additional constraints on the existing optimal solution or by introducing a new criterion that involves the variance of the implementation slippage.

1 Introduction

The aim herein is to present new results and discuss various issues related to Gatheral's model [4] for the optimal execution of a single order. We touch upon several topics, bring up new insights and introduce new concepts. While we focus only on the single order case, some of the results are stated more generally in order to prepare the ground for the sequential order case that we study in [5].

We provide a summary of the main contributions herein.

1. We first reformulate Gatheral's model to *correctly* define the execution strategies. Specifically, the policies in [4] are taken to be left continuous with right limits on $[0,T]$, with full execution to be completed at T. This, however, *excludes the possibility of a jump at* T. To remedy this, we assume that the order has to be fully executed by some $\hat{T} < T$. This, in turn,

generates the time lag $T - \hat{T}$ which is reflected on the impacted price and other quantities of interest.

2. Assuming a shorter than T execution horizon is also needed in order to properly define the sequential order model, introduced in [2] for two orders and extended in [5] for multi orders. We recall that a fundamental assumption in this extended setting is that each new order arrives *right after* the previous one is fully executed.

3. We propose an additional reformulation of Gatheral's model, replacing the strategies in [4] by *probability measures* whose cdfs represent the cumulativefraction of the volume traded through time. These measures become the defining element in the underlying optimization problem. The new parametrization yields quantities that have more natural continuity properties (e.g. right continuity with left limits, which is aligned with the standard assumptions in Ito's calculus) as well as convenient scaling properties with respect to the traded volume.

4. We provide a detailed analysis of the admissible control distributions. We analyze their densities and establish that the optimal one is symmetric around the middle point of the execution time, continuous in the interior and discontinuous at the end points with equal jumps, whose size we also compute. One of the key steps is a time-reversal property of the optimal measure.

5. So far, the analysis in both [4] and [2] has focused only on the mean of the implementation slippage. In the model we introduce (see point 3 above), we analyze other probabilistic properties like its distribution and variance, both for admissible and the optimal execution strategies. For general decay functions, we provide representation results for the variance and universal upper and lower bounds.

6. Starting with the tractable case of exponential decay functions, we observe a potentially *problematic* feature of Gatheral's model. Specifically, we note that, as the decay functions decrease, the model allows for *complete elimination* of the expected implementation slippage while the variance remains finite (or even becomes infinite), clearly an undesirable feature of the model.

   We show that similar issues arise for general decay functions as well; we do this by first establishing, and in turn using, the robustness of the implementation slippage with regards to these functions.

7. To remedy the problematic issue raised in point 6 we propose two possible directions. Firstly, still using the criterion of [4], we investigate how additional constraints on the optimized quantities - for example, requiring certain properties of the optimal variance (its size or asymptotic growth) - could be used to prevent pathological behavior. This approach leads

to *constraints* on model inputs, like the stock's volatility and the decay function.

Secondly, we propose to depart from the objective in [4] and work with a *new optimality criterion*. We propose a mean-variance type criterion. Its formulation gives rise to a two-dimensional constrained calculus of variations problem; the variance representations in point 5 above play a key role here.

8. Finally, we study the different kinds of market impacts proposed and developed by several authors. We attempt to reconcile the various notions within our model, noting however that some of the most popular ones have been developed in models that do *not* allow for discontinuous policies. In a related direction, we study the effects of the power law assumption for the average price change on the selection of model inputs within our model.

Some of the above points might at first look a bit redundant for the single order case examined here. However, as we argue in [5], they all become important in the sequential order framework. For example, in the general setting, in addition to the randomness generated by the Brownian motion, there is randomness coming from the unknown volumes of the incoming orders as well as the related random decay functions and volatility levels. As a result, the analysis of the aggregate implementation slippage naturally involves both conditional (to the information at order arrival) and unconditional arguments, and quantities like variances and covariances for both the stock and the implementation slippage processes play a key role. Furthermore, the time lags (see point 1) have aggregate effects on the impacted stock price, sequentially between each time an order is fully executed and the next one arrives.

2 Gatheral's model revisited and reformulated

We start with a brief review of the classical Gatheral's model [4], bring up several issues that we further discuss and develop in subsequent sections, and provide new insights and results for various quantities of interest. To facilitate the exposition, we abstract for now from most technical assumptions and only present the technical steps that are relevant for the new parts we develop.

We consider a single order of volume V, which arrives at time 0 and must be executed by (an a priori known) time $T < \infty$. Without loss of generality, we assume that this is a *buy* order, $V > 0$. In [4], it is assumed that execution of this order is done by *deterministic* policies that are *left continuous with right hand limits*.

The first step in our analysis is to correctly set up this problem in a *modified execution horizon*. Specifically, we impose that, instead of T, the order must be completed strictly before T, say at $\hat{T} < T$. There are two reasons for this:

i) Block trades are allowed even at the end point of the execution horizon; as a matter of fact, the optimal policy turns out to include a jump therein. As a result, the right hand continuity requirement in [4] forces us to properly define the policies also right after the execution time.

ii) A fundamental assumption in the extended model of sequential orders (say a total of N orders) is that each order must be *entirely completed* before the next one arrives. Thus, if the n^{th} order arrives at T_{n-1} and the $n+1$ order arrives at T_n, the $n-1$ order must be completed by $\hat{T}_{n-1} < T_{n-1}$ and the n^{th} order by $\hat{T}_n < T_n$. As we argue in [5], even though each interval $T_i - \hat{T}_i$, $i = 1, ..., N$, is very small, these infinitesimal sequential time lags have aggregate effects on various quantities, especially, on the variance and covariances of the individual implementation slippages and the movements of the impacted stock price in each interval $\left(T_i - \hat{T}_i\right]$.

2.1 First modification of Gatheral's model

To properly define the admissible policies and, also, set the framework for the upcoming sequential order case, we modify accordingly the set of admissible policies of [4] to

$$\mathcal{A}_{[0,T]} = \{X : X_t,\ t \in [0,T]\,,\ \text{deterministic, non-decreasing,}$$
$$\text{left continuous with right hand limits, } X_0 = 0 \text{ and } X_{\hat{T}+} = V,\ \hat{T} < T\Big\}. \quad (2)$$

The optimal execution model is now described on $[0,T]$ as follows:

- **Interval** $\left[0, \hat{T}\right]$: The impacted stock price process is modeled as

$$S_t = S_0 + \int_{[0,t)} G(t-s)\, dX_s + \sigma W_t, \quad 0 < t \leq \hat{T}, \quad (3)$$

  with $X \in \mathcal{A}_{[0,T]}$, W being a Brownian motion in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\sigma > 0$ a given constant. The decay function $G(|t|) > 0$, $t \in \mathbb{R}$, is also given and for, now, assumed to be smooth enough for $|t| \neq 0$ (we refine its properties later on).

For each $t \in \left[0, \hat{T}\right]$, the left continuous function X_t represents the *cumulative volume* bought just before time t and with excluding a possible block trade at t, while X_{t+} is the volume held at t, including a possible jump at t. Since $V > 0$, X_t and X_{t+} are non-negative and non-decreasing. Thus, X is (uniquely) decomposed as $X = X^c + X^d$ in terms of its continuous and singular components,

$$X_t^d = \sum_{0 \leq s \leq t} \Delta X_s \quad \text{and} \quad X_t^c = X_t - X_t^d, \quad (4)$$

where $\Delta X_t := X_{t+} - X_t$ denotes the jump at time t.

Jumps, may occur at any time $t \in \left[0, \hat{T}\right]$, including the end points 0 and $\hat{T}$. The requirement $X_{\hat{T}+} := V$ in (2) ensures that a possible jump may occur at the end of the effective execution period, $\left[0, \hat{T}\right]$. When there is a block trade, $X_{t+} - X_t > 0$, $0 \leq t \leq \hat{T}$, and thus the impacted price satisfies

$$S_{t+} = S_t + G(0)\, \Delta X_t > S_t. \tag{5}$$

- **Interval** $\left(\hat{T}, T\right]$: In order to define the right limit of X at $\hat{T}$, we set

  $$X_t := V, \quad \hat{T} < t \leq T.$$

  At times $t \in \left(\hat{T}, T\right]$, there is no execution, $\Delta X_t \equiv 0$. The impacted price is given by

  $$S_t = S_{\hat{T}+} + \sigma W_t = S_{\hat{T}} + G(0)\, \Delta X_{\hat{T}} + \sigma W_t, \quad t \in \left(\hat{T}, T\right], \tag{6}$$

  as X might have a jump at $\hat{T}$.

We note that every strategy $X \in \mathcal{A}_{[0,T]}$ determines uniquely a finite Borel measure. Specifically, this measure, denoted by χ, is defined for $0 \leq a < b \leq T$ as

$$\chi([a, b)) := X_b - X_a. \tag{7}$$

Conversely, for a finite Borel measure χ such that $\chi\left(\left[0, \hat{T}\right]\right) = V$, $\hat{T} < T$, the execution strategy is specified as

$$X_t = \begin{cases} 0, \quad t = 0 \\ \chi([0, t)), \ 0 < t \leq \hat{T} \\ V, \quad \hat{T} < t \leq T. \end{cases} \tag{8}$$

We will also consider the normalized by the volume V policy ψ_t, $t \in [0, T]$, defined as

$$d\psi_t := \frac{1}{V} dX_t \ \text{ with } \psi_0 = 0. \tag{9}$$

The *implementation cost* $IC(X)$ of an arbitrary strategy $X \in \mathcal{A}_{[0,T]}$ is defined in terms of the corresponding costs of the continuous and singular components $IC(X^c)$ and $IC(X^d)$ (cf. (4)).

The cost of implementing X^c is the sum of the costs of each infinitesimal in time trade dX_t, $t \in \left[0, \hat{T}\right]$, executed at price S_t,

$$IC(X^c) = \int_{\left[0, \hat{T}\right]} S_t dX_t^c.$$

The cost of implementing X^d is the sum of the costs of implementing each individual jump ΔX_t. It is assumed that the cost of each block trade ΔX_t, generating jump ΔS_t, can be decomposed into the sum of the costs of infinitesimal in space orders $\frac{1}{G(0)}dy$, each executed at price y, for each $y \in [S_t, S_{t+}]$. This assumption may be linked to the model proposed in [4] in which execution takes place in a limit order book (LOB) and the price impact of the block trade ΔX_t is linear in the order size; this is also the case in Gatheral's model, where we have

$$S_{t+} = S_t + G(0)\,\Delta X_t,$$

or $\Delta S_t = S_{t+} - S_t = G(0)\,\Delta X_t$. Then, the infinitesimal in space order $\frac{1}{G(0)}dy$ costs $\frac{y}{G(0)}dy$ to be executed. Therefore, the total cost of implementing an individual block trade ΔX_t is given by

$$\begin{aligned} IC(\Delta X_t) &= \int_{[S_t,S_{t+}]} \frac{y}{G(0)}dy = \frac{1}{2G(0)}\left(S_{t+} - S_t\right)^2 \\ &= \frac{1}{2}\Delta X_t\left(S_{t+} + S_t\right) = \frac{G(0)}{2}\left(\Delta X_t\right)^2 + S_t \Delta X_t. \end{aligned} \tag{10}$$

Consequently, the cost of implementing X^d is given by

$$\begin{aligned} IC(X^d) &= \sum_{t\in[0,\hat{T}]}\left(\frac{G(0)}{2}\left(\Delta X_t\right)^2 + S_t\Delta X_t\right) \\ &= \int_{[0,\hat{T}]} S_t dX_t^d + \sum_{t\in[0,\hat{T}]}\frac{G(0)}{2}\left(\Delta X_t\right)^2. \end{aligned}$$

We summarize the above findings in the following proposition.

Proposition 1 *The total implementation cost $IC(X)$ for strategy $X \in \mathcal{A}_{[0,T]}$ is given by*

$$IC(X) = \int_{[0,\hat{T}]} S_t dX_t + \sum_{t\in[0,\hat{T}]}\frac{G(0)}{2}\left(\Delta X_t\right)^2. \tag{11}$$

The main quantity of interest in optimal execution is the *implementation slippage* (or shortfall) introduced by Gatheral (see, [4], for example).

Definition 2 *The implementation slippage $IS(X)$ of strategy $X \in \mathcal{A}_{[0,T]}$ is defined as the difference between the cost of implementing strategy X and the cost of executing the entire volume V at initial price S_0,*

$$IS(X) := \int_{[0,\hat{T}]} S_t dX_t + \frac{G(0)}{2}\sum_{t\in[0,\hat{T}]}\left(\Delta X_t\right)^2 - VS_0.$$

Following the calculations in [4], we deduce (12) below. Properties (13) and (14) follow from the deterministic assumption on X and the properties of the Brownian motion.

Proposition 3 *The implementation slippage $IS(X)$ generated by strategy $X \in \mathcal{A}_{[0,T]}$ for a buy order of volume V and executed in $\left[0,\hat{T}\right]$ is given by*

$$IS(X) = \frac{1}{2}\int_{\left[0,\hat{T}\right]}\int_{\left[0,\hat{T}\right]} G(|t-s|)\, dX_s dX_t + \int_{\left[0,\hat{T}\right]} \sigma W_t dX_t. \tag{12}$$

It is an $\mathcal{F}_{\hat{T}}$-measurable random variable and is normally distributed with mean

$$E_{\mathbb{P}}(IS(X)) = \frac{1}{2}\int_{\left[0,\hat{T}\right]}\int_{\left[0,\hat{T}\right]} G(|t-s|)\, dX_s dX_t, \tag{13}$$

and variance

$$Var_{\mathbb{P}}(IS(X)) = \sigma^2 \int_{\left[0,\hat{T}\right]}\int_{\left[0,\hat{T}\right]} (t \wedge s) dX_s dX_t. \tag{14}$$

The measure $\mathbb{P}$ is the one related to the Brownian motion; for convenience, we will drop the $\mathbb{P}$-notation from now on.

As proposed in [4], the aim is to *minimize the expected slippage* (13) over all admissible strategies,

$$E(IS(X^*)) = \min_{X \in \mathcal{A}_{[0,T]}} \frac{1}{2}\int_{\left[0,\hat{T}\right]}\int_{\left[0,\hat{T}\right]} G(|t-s|)\, dX_s dX_t. \tag{15}$$

This problem was solved in [4] and more recently revisited in [2] where an optimal transport setting was used to prove the uniqueness of the optimizer. In these two works, slightly different assumptions on the decay function were introduced but, herein, we do not focus on this. For now, we introduce the assumption in [4] and recall the main optimality result therein.

Assumption 1: *The decay function $G(|t|) > 0$, $t \in \mathbb{R}$, is represented as the Fourier transform of a positive finite Borel measure φ on $\mathbb{R}$,*

$$G(|t|) = \int_{\mathbb{R}} e^{ity} \varphi(dy). \tag{16}$$

Proposition 4 *Let the decay function G satisfy (16). Then, the minimization problem (15) has a unique optimal execution strategy, X^*. Its optimality is equivalent to the existence of a (unique) $\lambda^* > 0$, such that, for each $t \in \left[0,\hat{T}\right]$,*

$$\int_{\left[0,\hat{T}\right]} G(|t-s|)\, dX^*_s = \lambda^*. \tag{17}$$

The minimized expected implementation slippage is given by

$$E(IS(X^*)) = \frac{1}{2}\lambda^* V. \tag{18}$$

It follows that $E(IS(X^*))$ is decreasing in the execution time, i.e. for $\hat{T}' < \hat{T} < T$,

$$E(IS(X^*_{\hat{T}}); \hat{T}) \leq E\left(IS(X^*_{\hat{T}'}); \hat{T}'\right),$$

with $X^*_{\hat{T}} \in \mathcal{A}_{[0,\hat{T}]}$ and $X^*_{\hat{T}'} \in \mathcal{A}_{[0,\hat{T}']}$. This horizon monotonicity property is consistent with the intuition and, also, demonstrates that the most expensive strategy is to execute the entire volume V at initial time 0. It also suggests that in order to minimize the expected implementation cost, one should increase the execution time. Therefore, while we originally assumed that the order must be executed *by* $\hat{T}$, it is optimal to complete its execution *exactly at* $\hat{T}$ and not earlier.

Remark 5 *As demonstrated in (10), the cost $IC(\Delta X_t)$ to implement a block trade of size ΔX_t is proportional to it and to the average of the price before and right after the jump. In particular, the cost to buy the entire volume V at initial time 0 (in general, a suboptimal policy) is $IC(\Delta X_0) = \frac{1}{2}V(S_{0+} + S_0)$. In turn, the associated implementation slippage is given by*

$$IS(\Delta X_0) = \frac{1}{2}V(S_{0+} - S_0) = \frac{G(0)}{2}V^2.$$

2.2 Exponential decay function

Due to its tractability, a popular choice for the decay function is

$$G(|t|) = e^{-\kappa|t|}, \ t \in \mathbb{R} \text{ and } \kappa > 0. \tag{19}$$

Proposition 6 *If the decay function G is as in (19), the optimal strategy X^*_t satisfies*

$$dX^*_t = \frac{V}{2+\kappa\hat{T}}\left(\delta_0 + \delta_{\hat{T}} + \kappa dt\right), \quad 0 < t \leq \hat{T}, \tag{20}$$

and the optimality condition (17) holds for

$$\lambda^* = \frac{2V}{2+\kappa\hat{T}}. \tag{21}$$

The minimal expected implementation slippage is given by

$$E(IS(X^*)) = \frac{V^2}{2+\kappa\hat{T}} = \frac{1}{2}\lambda^* V \tag{22}$$

and its variance by

$$Var(IS(X^*)) = V^2\sigma^2\hat{T}\frac{1+\kappa\hat{T}+\frac{1}{3}\left(\kappa\hat{T}\right)^2}{\left(2+\kappa\hat{T}\right)^2}. \tag{23}$$

For fixed V and $\hat{T}$,

$$\lim_{\kappa\uparrow\infty} E(IS(X^*)) = 0 \quad \text{and} \quad \lim_{\kappa\uparrow\infty} Var(IS(X^*)) = \frac{1}{3}V^2\sigma^2\hat{T}. \tag{24}$$

Proof. For (20), (21) and (22) see [2] and [4]. To show (23) we work as follows. From (12), we have

$$IS\left(X^{*}\right)=\frac{1}{2}\int_{[0,\hat{T}]}\int_{[0,\hat{T}]}G\left(|t-s|\right)dX_{s}^{*}dX_{t}^{*}+\int_{[0,\hat{T}]}\sigma W_{t}dX_{t}^{*}.$$

Recalling ψ_t defined in (9), we have

$$\begin{aligned}Var\left(IS\left(X^{*}\right)\right)&=V^{2}\sigma^{2}\left(\int_{[0,\hat{T}]}\int_{[0,\hat{T}]}t\wedge s d\psi_{s}^{*}d\psi_{t}^{*}\right)\\&=V^{2}\sigma^{2}\int_{[0,\hat{T}]}\frac{1}{2+\kappa\hat{T}}\left(t+\int_{0}^{\hat{T}}\left(t\wedge s\right)\kappa ds\right)d\psi_{t}^{*}.\end{aligned}$$

Since $\int_0^{\hat{T}}\left(t\wedge s\right)\kappa ds=\kappa\left(\frac{1}{2}t^{2}+t\left(\hat{T}-t\right)\right)$, we deduce

$$\int_{[0,\hat{T}]}\int_{[0,\hat{T}]}t\wedge s d\psi_{s}^{*}d\psi_{t}^{*}$$

$$=\frac{1}{2+\kappa\hat{T}}\int_{[0,\hat{T}]}\left(t+\kappa\left(\frac{1}{2}t^{2}+t\left(\hat{T}-t\right)\right)\right)d\psi_{t}^{*}=\frac{1}{\left(2+\kappa\hat{T}\right)^{2}}\left(\hat{T}+\frac{1}{2}\kappa\hat{T}^{2}\right)$$

$$+\frac{1}{\left(2+\kappa\hat{T}\right)^{2}}\int_{0}^{\hat{T}}\left(t+\kappa\left(\frac{1}{2}t^{2}+t\left(\hat{T}-t\right)\right)\right)\kappa dt$$

$$=\frac{\hat{T}+\frac{1}{2}\kappa\hat{T}^{2}}{\left(2+\kappa\hat{T}\right)^{2}}+\frac{\frac{1}{2}\kappa\hat{T}^{2}+\kappa^{2}\left(\frac{1}{6}\hat{T}^{3}+\frac{1}{6}\hat{T}^{3}\right)}{\left(2+\kappa\hat{T}\right)^{2}}=\frac{1+\kappa\hat{T}+\frac{1}{3}\kappa^{2}\hat{T}^{2}}{\left(2+\kappa\hat{T}\right)^{2}}\hat{T}.$$

■

From the stock price equations (3) and (6) we derive the impacted price on $[0,T]$.

Proposition 7 *Under the optimal execution policy (20), the impacted price process S_t^*, $t\in[0,T]$, is given by*

$$S_0,\ t=0\quad and\quad S_{0+}^{*}=S_{0}+\frac{V}{2+\kappa\hat{T}},$$

$$S_{t}^{*}=S_{0}+\frac{V}{2+\kappa\hat{T}}+\sigma W_{t}=S_{0+}^{*}+\sigma W_{t},\quad 0<t\leq\hat{T},$$

$$S_{\hat{T}+}^{*}=S_{0}+\frac{2V}{2+\kappa\hat{T}}+\sigma W_{\hat{T}}=S_{\hat{T}}^{*}+\frac{V}{2+\kappa\hat{T}},$$

and

$$S_{t}^{*}=S_{0}+\frac{2V}{2+\kappa\hat{T}}+\sigma\left(W_{t}-W_{\hat{T}}\right)=S_{\hat{T}+}^{*}+\sigma\left(W_{t}-W_{\hat{T}}\right),\quad\hat{T}<t<T.$$

Next, we comment on the above results.

1. The optimal strategy is to first execute the block trade $\frac{V}{2+\kappa\hat{T}}$ at time 0. Subsequently, by each $t \in \left(0, \hat{T}\right]$, the volume $\frac{V}{2+\kappa\hat{T}} + \frac{\kappa V}{2+\kappa\hat{T}}t$ is traded with smooth execution rate $\frac{\kappa V}{2+\kappa\hat{T}}$. The remaining volume $\frac{V}{2+\kappa\hat{T}}$ is executed as a block trade at $\hat{T}$.

2. The two block trades at 0 and $\hat{T}$ are *symmetric*, each having size $\frac{V}{2+\kappa\hat{T}}$. As $\kappa \uparrow \infty$, the optimal policy converges to smooth, uniform execution in $\left(0, \hat{T}\right)$ of constant rate $\frac{V}{\hat{T}}$.

3. Contrary to the standing assumptions in stochastic calculus (cadlag), the above impacted stock price S^* is left continuous with right hand limits.

4. For fixed V and κ, it holds that

$$\frac{\partial}{\partial \hat{T}} E(IS\left(X^*\right); \hat{T}) = -\frac{\kappa V^2}{\left(2+\kappa\hat{T}\right)^2} < 0$$

   and, hence, the optimal expected slippage decreases with the execution time. Furthermore, $\lim_{\hat{T}\uparrow\infty} E(IS\left(X^*\right); \hat{T}) = 0$, but the convergence is rather slow.

5. For fixed $\hat{T}$, $E(IS\left(X^*\right))$ depends *exclusively* on the volume V and the size of the equal at 0 and $\hat{T}$ jumps,

$$E(IS\left(X^*\right)) = \frac{V}{2+\kappa\hat{T}}V = \frac{1}{2}V\left(\Delta S_0^* + \Delta S_{\hat{T}}^*\right) = V\Delta S_0^* = V\Delta S_{\hat{T}}^*.$$

   In particular, the *smooth execution* component $\frac{\kappa V}{2+\kappa\hat{T}}t$, $t \in \left(0, \hat{T}\right)$ does *not generate* any cost.

6. For each $\hat{T}$ and $\kappa > 0$, $E(IS\left(X^*\right); V)$ is, as expected, strictly increasing in the volume,

$$\frac{\partial E(IS\left(X^*\right); V)}{\partial V} = \frac{2V}{2+\kappa\hat{T}} > 0.$$

   On the other hand, it is quadratic in V, which might at first look counter intuitive. However, one may allow κ to depend on volume V, which would give $E(IS\left(X^*\right); \kappa\left(V\right), V) = \frac{V}{2+\kappa(V)\hat{T}}$. In particular, if one chooses $\kappa\left(V\right) = V$, then

$$\frac{E(IS\left(X^*\right)}{V} = \frac{V}{2+V\hat{T}} < \frac{1}{\hat{T}}$$

   and

$$\lim_{\kappa\uparrow\infty} \frac{E\left(IS(X^*)\right)}{V} = \lim_{V\uparrow\infty} \frac{E\left(IS(X^*)\right)}{V} = \frac{1}{\hat{T}},$$

   yielding a linear dependence on V, for large V.

7. For fixed $\hat{T}$ and V, $E(IS\left(X^{*}\right);\kappa)$ is strictly decreasing in the decay parameter κ as
$$\frac{\partial E(IS\left(X^{*}\right);\kappa)}{\partial\kappa}=-\frac{V^{2}\hat{T}}{\left(2+\kappa\hat{T}\right)^{2}}<0.$$
This is intuitively pleasing as the decay function (19) decreases monotonically with κ. Note, however, that as κ increases, we have in the limit
$$\lim_{\kappa\uparrow\infty}E(IS\left(X^{*}\right);\kappa)=0\quad\text{and}\quad\lim_{\kappa\uparrow\infty}Var\left(IS(X^{*})\right)=\frac{1}{3}V^{2}\sigma^{2}\hat{T}.\tag{25}$$
In other words, the model allows for complete *elimination* of the expected slippage in the limit, while yielding a non-zero variance. This seems like an *unsuitable* modeling feature.

Observations in point 6 and, in particular, point 7 prompt us to examine how the parameter κ should be chosen given a targeted model behavior. More importantly, the limiting behavior as $\kappa\uparrow\infty$ indicates that, perhaps, only minimizing the expected slippage *might not be an adequate* modeling objective. We study these questions later on in section 4.

3 Second modification of Gatheral's model

We further modify Gatheral's model. We introduce a *probability measure* whose cumulative distribution function models the fraction of the traded volume cumulatively through time. This measure now becomes the defining element in the optimization problem we study. We make this precise next.

Consider the set of probability measures
$$\mathcal{P}_{[0,T]}=\left\{\nu:\nu\text{ probability measure on }\mathcal{B}\left([0,T]\right)\text{ with }\nu\left(\left[0,\hat{T}\right]\right)=1,\ \hat{T}<T\right\}\tag{26}$$
and the associated cumulative distribution functions by $\nu_t:[0,T]\rightarrow[0,1]$,
$$\nu_t:=\nu\left([0,t]\right),\ 0\leq t\leq T.\tag{27}$$
Note that ν_t is right continuous, nondecreasing with left limits (cadlag), $\nu_{0-}:=0$ and $\nu_t=1$, for $t\in\left[\hat{T},T\right]$.

We may then relate ν_t and ψ_t, defined in (9), through $\psi_t=\nu_{t-}$, $t\in[0,T]$. Recall that the execution strategies $X\in\mathcal{A}_{[0,T]}$ and the associated measures χ (cf. (7)) and ψ have "opposite" continuity properties, in that they are left continuous with right limits.

Next, we view ν_t, $0\leq t\leq T$, as the execution policies. Working as in section 1, we derive the implementation slippage,
$$IS(\nu)=\frac{1}{2}V^{2}\int_{[0,\hat{T}]}\int_{[0,\hat{T}]}G\left(|t-s|\right)d\nu_s d\nu_t+V\int_{[0,\hat{T}]}\sigma W_t d\nu_t,\tag{28}$$

and consider the related minimization problem

$$E(IS(\nu^{*})) = \min_{\nu\in\mathcal{P}_{[0,T]}} \frac{1}{2}V^{2}\int_{[0,\hat{T}]}\int_{[0,\hat{T}]} G\left(|t-s|\right)d\nu_s d\nu_t. \tag{29}$$

We will analyze this problem next. We stress that the probability measure ν now becomes the defining element as it gives rise to the cumulative distribution ν_t, and in turn to the strategies ψ_t and X_t (cf. (9) and (2)) in the formulation of the model in section 1.

3.1 Properties of admissible policies

We start with some general properties of the probability measure ν and its distribution function. Properties (33), (34) and (35) will be used to compute the variance of the implementation slippage while (36) and (37) to study the robustness of the model in terms of the decay function.

We choose to present these properties for an arbitrary time interval, $[a,b]$, instead of $\left[0,\hat{T}\right]$, in order to facilitate the adaptation of the results to the model of sequential orders, where a and b will represent arrival and execution times, respectively. For $0\leq a<b$, the set $\mathcal{P}_{[a,b]}$ is defined in analogy to (26).

Lemma 8 *If $\nu\in\mathcal{P}_{[a,b]}$, then ν_t, satisfies*

$$\int_{[a,b]}\int_{[a,b]} t\wedge s d\nu_s d\nu_t = a + \int_{(a,b]}\left(\nu\left([t,b]\right)\right)^{2}dt. \tag{30}$$

Proof. Let A and B be of finite variation, right continuous functions on $[a,b]$, with $A_{a-}=B_{a-}=0$, and f be a bounded measurable function on $[a,b]$. Then, for $a\leq s<t\leq b$,

$$\int_{[a,t]} f\left(s\right)dA_s = f\left(a\right)A_a + \int_{(a,t]} f\left(s\right)dA_s \tag{31}$$

and

$$A_tB_t - A_sB_s = \int_{(s,t]} A_{u-}dB_u + \int_{(s,t]} B_{u-}dA_u + \sum_{u\epsilon(s,t]}\left(\Delta A_u\right)\left(\Delta B_u\right), \tag{32}$$

where $\Delta C_u = C_u - C_{u-}$, $C=A,B$ (see, for example, [3]). In turn, (31) yields

$$\int_{[a,b]} t\wedge s d\nu_s = a\nu_a + \int_{(a,b]} t\wedge s d\nu_s = a\nu_a + \int_{(a,t]} t\wedge s d\nu_s + \int_{(t,b]} t\wedge s d\nu_s$$

$$= a\nu_a + \int_{(a,t]} s d\nu_s + t\nu\left((t,b]\right).$$

From (32) we deduce that, if $a<t$, $t\nu_t - a\nu_a = \int_{(a,t]} s d\nu_s + \int_{(a,t]}\nu_{s-}ds$. Hence,

$$\int_{[a,b]} t\wedge s d\nu_s = a\nu_a + t\nu_t - a\nu_a - \int_{(a,t]}\nu_{s-}ds + t\nu\left((t,b]\right)$$

$$= t\nu_t + t\nu\left((t,b]\right) - \int_{(a,t]} \nu_{s-} ds = t - \int_{(a,t]} \nu_{s-} ds$$

and, thus,

$$\int_{[a,b]} \int_{[a,b]} (t \wedge s)\, d\nu_s d\nu_t = \int_{[a,b]} \left(t - \int_{(a,t]} \nu_{s-} ds \right) d\nu_t$$

$$= a\nu_a + \int_{(a,b]} \left(t - \int_{(a,t]} \nu_{s-} ds \right) d\nu_t = a\nu_a + \int_{(a,b]} t d\nu_t - \int_{(a,b]} \int_{(a,t]} \nu_{s-} ds d\nu_t$$

$$= a\nu_a + b\nu_b - a\nu_a - \int_{(a,b]} \nu_{s-} ds - \int_{(a,b]} \int_{(a,t]} \nu_{s-} ds d\nu_t$$

$$= b\nu_b - \int_{(a,b]} \nu_{s-} ds - \int_{(a,b]} \int_{(a,t]} \nu_{s-} ds d\nu_t.$$

Next, using (32) with $A_t = \int_{(a,t]} \nu_{s-} ds$ and $B_t = \nu_t$, we deduce

$$\int_{(a,b]} \int_{(a,t]} \nu_{s-} ds d\nu_t = \int_{(a,b]} A_t dB_t = A_b B_b - A_a B_a - \int_{(a,b]} B_{t-} dA_t$$

$$= \left(\int_{(a,b]} \nu_{s-} ds \right) \nu_b - \int_{(a,b]} \nu_{t-} dA_t = \left(\int_{(a,b]} \nu_{s-} ds \right) \nu_b - \int_{(a,b]} \nu_{s-}^2 dt.$$

Recalling that $\nu_b = 1$, we then obtain

$$\int_{[a,b]} \int_{[a,b]} (t \wedge s) d\nu_s d\nu_t = b\nu_b - \int_{(a,b]} \nu_{s-} ds - \int_{(a,b]} \int_{]a,t]} \nu_{s-} ds d\nu_t$$

$$= b\nu_b - \int_{(a,b]} \nu_{s-} ds - \left(\left(\int_{(a,b]} \nu_{t-} ds \right) \nu_b - \int_{(a,b]} \nu_{t-}^2 dt \right)$$

$$= b - 2 \int_{(a,b]} \nu_{s-} ds + \int_{(a,b]} \nu_{t-}^2 dt = a + \int_{(a,b]} \left(1 - 2\nu_t + \nu_{t-}^2 \right) dt$$

$$= a + \int_{(a,b]} \left(1 - \nu_{t-}\right)^2 dt = a + \int_{(a,b]} \left(\nu\left([t,b]\right) \right)^2 dt,$$

and we conclude. ∎

Proposition 9 *Let* $\nu \in \mathcal{P}_{[a,b]}$. *Then, for* $t \in [a,b]$ *,*

$$\int_{[a,b]} (W_t - W_a)\, d\nu_t = \int_{(a,b]} \nu\left([t,b]\right) dW_t, \tag{33}$$

$$E\left(\int_{[a,b]} (W_t - W_a)\, d\nu_t \right)^2 = \int_{(a,b]} \left(\nu\left([t,b]\right) \right)^2 dt, \tag{34}$$

and

$$\frac{1}{b-a}\left(\int_{[a,b]}(t-a)\,d\nu_t\right)^2 \le \int_{(a,b]}(\nu([t,b]))^2\,dt \le \int_{[a,b]}(t-a)\,d\nu_t. \tag{35}$$

Furthermore,

$$\inf_{\nu}\int_{(a,b]}(\nu([t,b]))^2\,dt = 0 \tag{36}$$

and

$$\sup_{\nu}\int_{(a,b]}(\nu([t,b]))^2\,dt = b-a. \tag{37}$$

Proof. We have

$$\begin{aligned}\int_{[a,b]}(W_t-W_a)\,d\nu_t &= \int_{(a,b]}(W_t-W_a)\,d\nu_t\\ &= (W_b-W_a)\,\nu_b - \int_{(a,b]}\nu_{t-}dW_t\\ = \int_{(a,b]}(1-\nu_{t-})\,dW_t &= \int_{(a,b]}(1-\nu([a,t)))\,dW_t = \int_{(a,b]}\nu([t,b])\,dW_t\end{aligned}$$

and (33) follows.

To show (34), we recall that ν_t is deterministic and bounded. Therefore,

$$\begin{aligned}E\left(\int_{[a,b]}(W_t-W_a)\,d\nu_t\right)^2 &= E\int_{[a,b]}\int_{[a,b]}(W_t-W_a)(W_s-W_a)\,d\nu_t d\nu_s\\ &= \int_{[a,b]}\int_{[a,b]}E\,(W_t-W_a)(W_s-W_a)\,d\nu_t d\nu_s\\ &= \int_{[a,b]}\int_{[a,b]}(t\wedge s-a)\,d\nu_t d\nu_s = \int_{(a,b]}(\nu([t,b]))^2\,dt,\end{aligned}$$

where we used the previous Lemma. Next, we observe that

$$\begin{aligned}\int_{[a,b]}(t-a)\,d\nu_t &= \int_{(a,b]}(t-a)\,d\nu_t = (b-a)\,\nu_b - \int_{(a,b]}\nu_{t-}dt\\ &= \int_{(a,b]}(1-\nu_{t-})\,dt = \int_{(a,b]}\nu([t,b])\,dt.\end{aligned}$$

Hence,

$$\int_{(a,b]}(\nu([t,b]))^2\,dt \le \int_{(a,b]}\nu([t,b])\,dt = \int_{[a,b]}(t-a)\,d\nu_t,$$

and using that

$$\left(\int_{(a,b]}\nu([t,b])\,dt\right)^2 \le (b-a)\int_{(a,b]}(\nu([t,b]))^2\,dt,$$

we conclude. Assertions (36) and (37) follow directly using dominated convergence. ∎

3.2 Properties of optimal policies: smoothness, time-reversal and symmetry

We analyze properties of the candidate optimal policies represented via the cummulative distribution function ν_t. We state them in a slightly more general setting, in that we analyze properties of probability measures $\nu \in \mathcal{P}_{[a,b]}$ which satisfy for each $t \in [a,b]$,

$$\int_{[a,b]} G(|t-s|)\, d\nu_s = \mu, \tag{38}$$

for some $\mu > 0$ that is independent of t. The above condition is the direct analogue of (17) if $a = 0, b = T$ and $\mu = \frac{\lambda^*}{V}$.

Proposition 10 *Let $\nu \in \mathcal{P}_{[a,b]}$ satisfying (38). Assume that function G satisfies Assumption 1 and, for $t \geq 0$, $G(t)$ is strictly decreasing and strictly convex, three times continuously differentiable for $t > 0$ and with[1] $G'_+(0) < 0$. Then, the following assertions hold for $t \in (a,b)$:*

i) the density of ν_t is continuous on (a,b) and satisfies

$$\nu'_t = \frac{1}{2G'_+(0)} \left(-G''(b-t) - \int_{(a,t]} \nu_u G'''(t-u)\, du + \int_{(t,b]} \nu_u G'''(u-t)\, du \right).$$

ii) ν_t satisfies

$$2G'_+(0)\, \nu_{t-} - G'(b-t) + \int_{(a,b]} \nu_{u-} G''(|t-u|)\, du = 0,$$

and, hence, it has a jump at point a of size

$$\nu_a = \frac{1}{2G'_+(0)} \left(G'(b-a) - \int_{(a,b]} \nu_{u-} G''(u-a)\, du \right). \tag{39}$$

Proof. For $a < t < b$, we have

$$\int_{[a,b]} G(|t-s|)\, d\nu_s = G(t-a)\, \nu_a + \int_{(a,t]} G(t-s)\, d\nu_s + \int_{(t,b]} G(s-t)\, d\nu_s$$

with

$$\int_{(a,t]} G(t-s)\, d\nu_s = G(0)\, \nu_t - G(t-a)\, \nu_a + \int_{(a,t]} \nu_{s-} G'(t-s)\, ds$$

and

$$\int_{(t,b]} G(s-t)\, d\nu_s = G(b-t)\, \nu_b - G(0)\, \nu_t - \int_{(t,b]} \nu_{s-} G'(s-t)\, ds.$$

[1] We use the notation $G'_+(0)$ to denote the right derivative of G at point 0.

Therefore, for each $t \in (a,b)$,

$$G(b-t) + \int_{(a,t]} \nu_{s-} G'(t-s)\, ds - \int_{(t,b]} \nu_{s-} G'(s-t)\, ds = \mu,$$

and, in turn,

$$2G'_{+}(0)\, \nu_{t-} - G'(b-t) + \int_{(a,b]} \nu_{s-} G''(|t-s|)\, ds = 0. \tag{40}$$

Consequently, ν_{t-} is continuous in (a,b) and left continuous on $[a,b]$, with $\nu_{a-} = 0$ and $\lim_{t \downarrow a} \nu_{t-} = \lim_{t \downarrow a} \nu_t = \nu_a$, as ν_t is right continuous. Rearranging terms in (40) gives

$$\nu_a = \frac{1}{2G'_{+}(0)} \left(G'(b-a) - \int_{(a,b]} \nu_{s-} G''(s-a)\, ds \right). \tag{41}$$

Using that G is strictly decreasing and strictly convex we deduce that $\nu_a > 0$.

Differentiating (40) with respect to $t \in (a,b)$ and using that in this interval $\nu_{t-} = \nu_t$ together with the smoothness assumptions on G, we obtain

$$2G'_{+}(0)\, \nu'_t + G''(b-t) + \int_{(a,b]} \nu_s \frac{\partial}{\partial t} G''(|t-s|)\, ds$$

$$= 2G'_{+}(0)\, \nu'_t + G''(b-t) + \int_{(a,b]} \nu_s G'''(|t-s|) \frac{\partial}{\partial t} |t-s|\, ds$$

$$= 2G'_{+}(0)\, \nu'_t + G''(b-t) + \int_{(a,t]} \nu_s G'''(t-s)\, ds - \int_{(t,b]} \nu_s G'''(s-t)\, ds = 0.$$

Consequently, the density ν'_t of the cumulative distribution ν_t is continuous for all $t \in (a,b)$ and satisfies

$$\nu'_t = \frac{1}{2G'_{+}(0)} \left(-G''(b-t) - \int_{(a,t]} \nu_s G'''(t-s)\, ds + \int_{(t,b]} \nu_s G'''(s-t)\, ds \right). \tag{42}$$

Using the above we rewrite (38) as

$$\int_{[a,b]} G(|t-s|)\, d\nu_s$$

$$= G(t-a)\, \nu_a + \int_{(a,b)} G(|t-s|)\, \nu'_s ds + G(b-t)(1-\nu_{b-}) = \mu.$$

■

Corollary 11 *Equality (38) may be written, for every $t \in [a,b]$, as*

$$\int_{[a,b]} G(|t-s|)\, d\nu_s$$

$$= G(t-a)\, \nu_a + \int_{(a,b)} G(|t-s|)\, \nu'_s ds + G(b-t)(1-\nu_{b-}) = \mu. \tag{43}$$

Let $\nu \in \mathcal{P}_{[a,b]}$ and define, for $t \in [a,b]$, the non-decreasing function

$$\widetilde{\nu}_t := 1 - \nu_{a+b-t} \quad \text{and} \quad \widetilde{\nu}_{b+} := 1. \tag{44}$$

Lemma 12 *The function $\widetilde{\nu}_t$ satisfies $\widetilde{\nu}_a = 0$, $\widetilde{\nu}_{a+} = \nu_a$, $\widetilde{\nu}_b = 1 - \nu_a$ and $\widetilde{\nu}_{b+} - \widetilde{\nu}_b = \nu_a$. Thus, the measure $\tilde{\nu}$ generated by $\widetilde{\nu}_t$ has jumps at both a and b, and of the same size equal to ν_a in (41). Furthermore, for $t \in (a,b)$, the corresponding densities satify*

$$\widetilde{\nu}'_t = \nu'_{a+b-t}.$$

Proposition 13 *i) Let ν_t satisfy (38) and $\widetilde{\nu}_t$ as in (44). Then, $\nu_t = \tilde{\nu}_t$, $t \in [a,b]$, and thus the corresponding measures coincide,*

$$\nu = \tilde{\nu}. \tag{45}$$

ii) The function ν_t has equal jumps at the end points a and b, with jump size equal to ν_a, given in (39),

$$\nu_{a+} - \nu_a = \nu_{b+} - \nu_b = \frac{1}{2G'(0_+)} \left(G'(b-a) - \int_{(a,b]} \nu_{s-} G''(s-a)\, ds \right).$$

Proof. Using $\widetilde{\nu}$ in (43) yields that for each $t \in [a,b]$,

$$\int_{(a,b)} G(|t-s|)\, \widetilde{\nu}'_s ds = \int_{(a,b)} G(|t-s|)\, \nu'_{a+b-s} ds$$

$$= \int_{(a,b)} G(|a+b-t-s|)\, \nu'_s ds,$$

$$G(a+b-t-a)\, \nu_a = G(b-t)\, \nu_a,$$

and

$$G(b-(a+b-t))\,(1-\nu_{b-}) = G(t-a)\, \nu_a.$$

Consequently, $\widetilde{\nu}$ also satisfies (43), and (45) follows from the uniqueness of the optimizer. ∎

An alternative way to describe property (45) is to refer to the *symmetry* of the (optimal) measure ν, or of its cumulative distribution function ν_t. We discuss this next.

Proposition 14 *Let ν satisfy (38). The following assertions hold:*

i) The cumulative distribution ν_t satisfies, for all $t \in [a,b]$,

$$\nu'_{\frac{a+b}{2}+t} = -\nu'_{\frac{a+b}{2}-t} \quad \text{and} \quad \nu''_{\frac{a+b}{2}+t} = -\, \nu''_{\frac{a+b}{2}-t} \tag{46}$$

and

$$G(t-a)\, \nu_a + \int_{(a,b)} G(|t-s|)\, \nu'_s ds + G(b-t)\, \nu_a = \mu. \tag{47}$$

ii) The density ν'_t has a maximum or a minimum at the middle point $t = \frac{a+b}{2}$.
iii) The following inequalities hold

$$\frac{b-a}{4} \leq \int_{(a,b]} (\nu([t,b]))^2 \, dt \leq \frac{b-a}{2}. \tag{48}$$

Proof. Let $\mathcal{V}$ be a random variable such that $P(\mathcal{V} \leq t) = \nu([a,t]) = \nu_t$. Recall that $\mathcal{V}$ is symmetric over the interval $[a,b]$ if the random variables $\mathcal{V} - \frac{a+b}{2}$ and $-\left(\mathcal{V} - \frac{a+b}{2}\right)$ have the same distribution,

$$P\left(\mathcal{V} - \frac{a+b}{2} \leq t\right) = \nu_{\frac{a+b}{2}+t} \quad \text{and} \quad P\left(-\mathcal{V} + \frac{a+b}{2} \leq t\right) = 1 - \nu_{\left(\frac{a+b}{2}-t\right)-}.$$

Hence, for all $t \in [a,b]$, $\nu_{\frac{a+b}{2}+t} = 1 - \nu_{\left(\frac{a+b}{2}-t\right)-}$ or, alternatively,

$$\nu_{t-} = 1 - \nu_{a+b-t}. \tag{49}$$

It, then, follows that $\nu_{\frac{a+b}{2}} = 1 - \nu_{a+b-\frac{a+b}{2}}$, and thus $\nu_{\frac{a+b}{2}} = \frac{1}{2}$. Moreover, because the density ν'_t is bounded, given the regularity assumptions on G, we also get that $E\left(\mathcal{V} - \frac{a+b}{2}\right) = E\left(-\left(\mathcal{V} - \frac{a+b}{2}\right)\right)$. Hence,

$$E(\mathcal{V}) = \int_{[a,b]} t d\nu_t = \frac{a+b}{2}.$$

Using the symmetry condition (49) we get that, for all $t \in (a,b)$, $\nu'_t = \nu'_{a+b-t}$.

Furthermore, $\nu''_{\frac{a+b}{2}+t} = -\nu''_{\frac{a+b}{2}-t}$ and, thus, $\nu''_{\frac{a+b}{2}} = -\nu''_{\frac{a+b}{2}}$ which implies that $\nu''_{\frac{a+b}{2}} = 0$ and (ii) holds. The rest of the proof follows easily. ■

Proposition 15 *For each $t \in [a,b]$,*

$$G\left(\int_{[a,b]} |t-s| \, d\nu_s\right) \leq \int_{[a,b]} G(|t-s|) \, d\nu_s = \mu \leq G(0). \tag{50}$$

Furthermore,

$$G\left(\frac{b-a}{2}\right) \leq G\left(2\int_{\left(a,\frac{a+b}{2}\right]} \nu_s ds\right) \leq \mu \leq G(0). \tag{51}$$

Proof. The first inequality in (50) follows from the convexity of G and the second from its monotonicity. To show (51), we work as follows. For $t \in [a,b]$, let

$$f(t) := \int_{[a,b]} |t-s| \, d\nu_s.$$

Then,

$$f(a) = \int_{[a,b]} (s-a) \, d\nu_s = \frac{b-a}{2} \quad \text{and} \quad f(b) = \int_{[a,b]} (b-s) \, d\nu_s = \frac{b-a}{2}.$$

Moreover, for $a < t < b$,

$$f(t) = \int_{[a,t]} (t-s)\, d\nu_s + \int_{(t,b]} (s-t)\, d\nu_s = t\nu_t - \int_{[a,t]} s d\nu_s + \int_{(t,b]} s d\nu_s - t\nu((t,b])$$

$$= 2t\nu_t - t - 2\int_{[a,t]} s d\nu_s + \int_{[a,b]} s d\nu_s = 2t\nu_t - t - 2\int_{[a,t]} s d\nu_s + \frac{a+b}{2}$$

$$= 2t\nu_t - t - 2\left(t\nu_t - \int_{(a,t]} \nu_s ds\right) + \frac{a+b}{2} = -t + 2\int_{(a,t]} \nu_s ds + \frac{a+b}{2}.$$

Therefore, $f'(t) = -1 + 2\nu_t$ and $f''(t) = 2\nu_t'$. It then follows that f is convex, decreasing on $\left[a, \frac{a+b}{2}\right]$ and increasing on $\left[\frac{a+b}{2}, b\right]$. Furthermore,

$$\nu_{t_0} = \frac{1}{2} \quad \text{if and only if } \; t_0 = \frac{a+b}{2}.$$

Consequently, $f(t)$ attains its minimum at $t_0 = \frac{a+b}{2}$ and, in addition,

$$f\left(\frac{a+b}{2}\right) = 2\int_{\left(a, \frac{a+b}{2}\right]} \nu_s ds$$

$$\le \left(\frac{a+b}{2} - a\right) = \frac{b-a}{2} = f(a) = f(b).$$

Using (50), the monotonicity of G and that

$$2\int_{\left(a, \frac{a+b}{2}\right]} \nu_s ds \le 2\nu_{\frac{a+b}{2}} \left(\frac{a+b}{2} - a\right) = \frac{b-a}{2}$$

we conclude. ■

4 Rethinking of optimality criteria in Gatheral's model

We discuss the behavior of the model with respect to the decay function and bring up some problematic features. For completeness, we first state the optimality results for problem (29). They are directly analogous to the ones in Proposition 4 so the proof is ommited. The representation for $Var(IS(\nu^*))$ is new and follows from (34).

Proposition 16 *i) The optimal measure ν^* for (29) is unique and, for each $t \in \left[0, \hat{T}\right]$,*

$$\int_{\left[0, \hat{T}\right]} G(|t-s|)\, d\nu_s^* = \mu^* = \frac{\lambda^*}{V}, \tag{52}$$

with λ^* as in Proposition 4.

ii) The optimal implementation slippage $IS(\nu^*)$ is an $\mathcal{F}_{\hat{T}}-$measurable random variable, given by

$$IS(\nu^*) = \frac{1}{2}V^2 \int_{[0,\hat{T}]} \int_{[0,\hat{T}]} G(|t-s|)\, d\nu_s^* d\nu_t^* + V \int_{[0,\hat{T}]} \sigma W_t d\nu_t^*. \tag{53}$$

It is normally distributed with mean and variance given, respectively, by

$$E(IS(\nu^*)) = \frac{1}{2}V^2\mu^* \quad \text{and} \quad Var(IS(X^*)) = V^2\sigma^2 \int_{(0,\hat{T}]} \left(\nu^*\left(\left[t,\hat{T}\right]\right)\right)^2 dt. \tag{54}$$

In the discussion after Proposition 6 (point 7), we remarked that Gatheral's criterion to solely minimize the expected implementation slippage appears inadequate, for it might lead to arbitrarily small, or even completely eliminated, execution costs without any penalty in doing so. This appears to be a *deficiency of the model* and prompts us to investigate how this problematic behavior could be possibly remedied.

As a first step, one could investigate how the model inputs could be chosen so that the underlying model offers a satisfactory solution if we impose constraints on the *already optimized* quantities. We recall that if the execution horizon is known, there are only two such inputs: the decay function and the volatility parameter. One choice - by no means the only one - is to require that the optimal $Var(IS(\nu^*))$ stays at a targeted level. We stress, however, that at this point we do not replace criterion (29) by a mean-variance one. Rather, we aim at controlling possible pathological behavior of the model through an implied model input choice. We further elaborate by first considering the tractable exponential case.

4.1 Exponential decay functions

If $G(|t|) = e^{-\kappa|t|},\ t \in \mathbb{R}$, for some $\kappa > 0$, the optimal policy ν_t^* satisfies

$$d\nu_t^* = \frac{1}{2+\kappa\hat{T}}\left(\delta_0 + \delta_{\hat{T}} + \kappa dt\right) \quad \text{and} \quad \int_{[0,\hat{T}]} e^{-\kappa|t-s|} d\nu_s^* = \mu^* = \frac{2}{2+\kappa\hat{T}}.$$

From (54),

$$E(IS(\nu^*)) = \frac{V^2}{2+\kappa\hat{T}} \quad \text{and} \quad Var(IS(\nu^*)) = V^2\sigma^2\hat{T}\frac{1+\kappa\hat{T}+\frac{1}{3}\left(\kappa\hat{T}\right)^2}{\left(2+\kappa\hat{T}\right)^2}, \tag{55}$$

and, as previously observed,

$$\lim_{\kappa\uparrow\infty} E(IS(\nu^*)) = 0 \quad \text{while} \quad \lim_{\kappa\uparrow\infty} Var(IS(\nu^*)) = \frac{1}{3}\sigma^2V^2\hat{T} < \infty.$$

i) Choosing the decay function: Assume that, for a given pair (σ, V) and execution horizon $\hat{T}$, we *require that the optimized variance $Var\left(IS\left(\nu^*\right)\right)$* stays at targeted level, say $M > 0$,

$$Var\left(IS\left(\nu^*\right)\right) = M.$$

Notice that the function

$$g\left(x\right) := \frac{1 + x + \frac{1}{3}x^2}{\left(2 + x\right)^2}, \ x > 0,$$

appearing in (55) with $x = \kappa\hat{T}$, satisfies

$$\lim_{x \downarrow 0} g\left(x\right) = \frac{1}{4} \quad \text{and} \quad \lim_{x \uparrow \infty} g\left(x\right) = \frac{1}{3},$$

and is strictly increasing,

$$g'\left(x\right) = \frac{\frac{1}{3}x^2 + \frac{2}{3}x}{\left(2 + x\right)^4} > 0, \quad \text{for } x > 0.$$

Therefore, *if* $K := \frac{M}{\sigma^2 V^2 \hat{T}}$ satisfies $K \in \left(\frac{1}{4}, \frac{1}{3}\right)$, the "implied" model choice for the *decay parameter* κ^* is *unique,*

$$\kappa^* = \frac{1}{\hat{T}} g^{(-1)} \left(\frac{M}{\sigma^2 V^2 \hat{T}}\right). \tag{56}$$

More generally, one may require that $Var\left(IS\left(\nu^*\right)\right)$ belongs to an *acceptable range*, say

$$M_1 \leq Var\left(IS\left(\nu^*\right)\right) \leq M_2 \tag{57}$$

which would yield

$$\frac{M_1}{\sigma^2 V^2 \hat{T}} < g(x) < \frac{M_2}{\sigma^2 V^2 \hat{T}}.$$

As long as $\frac{\sigma^2 V^2 \hat{T}}{4} < M_1 < M_2 < \frac{\sigma^2 V^2 \hat{T}}{3}$, the targeted variance range (57) would be attained for any κ^* satisfying

$$\frac{1}{\hat{T}} g^{(-1)} \left(\frac{M_1}{\sigma^2 V^2 \hat{T}}\right) < \kappa^* < \frac{1}{\hat{T}} g^{(-1)} \left(\frac{M_2}{\sigma^2 V^2 \hat{T}}\right). \tag{58}$$

The above observations may be somewhat generalized if we consider decay functions given by the extended exponential,

$$\tilde{G}\left(|t|\right) = \exp\left(-\sum_{i=1}^{n} \kappa_i \left|t\right|\right), \ t \in \mathbb{R}, \tag{59}$$

or, more generally, by completely monotonic functions, namely,

$$\bar{G}\left(|t|\right) = \int_{\alpha}^{\beta} e^{-y|t|} \varkappa\left(dy\right), \ t \in \mathbb{R}, \tag{60}$$

for $0 < \alpha \leq \beta < \infty$ and some positive Borel measure $\varkappa$ of finite mass.

We note that there is no linearity property at the optimum, in that $\nu^* \left(\tilde{G}\right) \neq \sum_{i=1}^{n} \nu^{*,i} \left(e^{-\kappa_i |t|}\right)$; similarly, for the optimal policies corresponding to $\bar{G}$.

ii) Choosing the volatility parameter: In a related direction, one may instead want to control the *growth rate* of the optimized variance $Var\left(IS\left(X^*\right)\right)$ as the decay parameter $\kappa \uparrow \infty$. For fixed $\left(V, \hat{T}\right)$, this may be done by letting the *volatility parameter* to depend on κ, for some $\sigma = \sigma(\kappa)$ with $\lim_{\kappa\uparrow\infty} \sigma(\kappa) = \infty$. One may then make $\lim_{\kappa\uparrow\infty} Var\left(IS\left(\nu^*\right)\right)$ to converge to infinity with an arbitrary rate. For example, if we choose $\sigma(\kappa) = \sqrt{2 + \kappa \hat{T}}$,

$$Var\left(IS\left(\nu^*\right)\right) = \frac{\hat{T} + \kappa \hat{T}^2 + \frac{1}{3}\kappa^2 \hat{T}^3}{2 + \kappa \hat{T}} \hat{T}$$

and the growth rate becomes linear. The rate may become quadratic, if $\sigma(\kappa) = 2 + \kappa \hat{T}$,

$$Var\left(IS\left(\nu^*\right)\right) = \left(\hat{T} + \kappa \hat{T}^2 + \frac{1}{3}\kappa^2 \hat{T}^3\right) \hat{T}.$$

4.2 General decay functions

For arbitrary decay functions we naturally loose the tractability of the exponential case, as explicit expressions cannot be obtained. However, as we show next, we may still obtain various monotonicity, robustness and limiting results for sequences of decay functions. These results highlight the possibly problematic behavior of the model we mentioned ealier.

Proposition 17 *Let $(G_m)_{m\geq 1}$ be a sequence of decreasing decay functions,*

$$0 < G_{m+1}(|t|) \leq G_m(|t|), \ t \in \mathbb{R}, \ m \geq 1.$$

Let also, for $m \geq 1$, $\nu_m^* \in \mathcal{P}_{[0,T]}$ and μ_m^* satisfy the related optimality conditions

$$\int_{[0,\hat{T}]} G_m\left(|t-s|\right) d\nu_{m,s}^* = \mu_m^*, \quad \textit{for each} \quad t \in \left[0, \hat{T}\right],$$

with $\mu_m^* = \frac{\lambda_m^*}{V}$ with λ_m^* as in Proposition 4. The following assertions hold:

i) The sequence $(\mu_m^*)_{m\geq 1}$ is decreasing,

$$\mu_{m+1}^* \leq \mu_m^*, \ m \geq 1. \tag{61}$$

ii) If $\lim_{m\uparrow\infty} G_m(|t|) = 0$ uniformly in t, then

$$\lim_{m\uparrow\infty} \mu_m^* = \lim_{m\uparrow\infty} \int_{[0,\hat{T}]} G_m\left(|t-s|\right) d\nu_{m,s}^* = 0. \tag{62}$$

Proof. Inequality (61) follows directly since

$$\mu^*_{m+1} = \int_{[0,\hat{T}]} G_{m+1}(|t-s|)\, d\nu^*_{m+1,s} \leq \int_{[0,\hat{T}]} G_m(|t-s|)\, d\nu^*_{m+1,s}$$

$$\leq \int_{[0,\hat{T}]} G_m(|t-s|)\, d\nu^*_{m,s} = \mu^*_m,$$

where we used that $\nu^*_{m+1,t}$ is an admissible but, in general suboptimal policy under decay function G_m. If $\lim_{m\uparrow\infty} G_m(|t|) = 0$ uniformly in t, then

$$\lim_{m\uparrow\infty} \mu^*_m = \lim_{m\uparrow\infty} \int_{[0,\hat{T}]} G_m(|t-s|)\, d\nu^*_{m,s}$$

$$= \lim_{m\uparrow\infty} \int_{[0,\hat{T}]} \int_{[0,\hat{T}]} G_m(|t-s|)\, d\nu^*_{m,s} d\nu^*_{m,t} \leq \lim_{m\uparrow\infty} \int_{[0,\hat{T}]} \int_{[0,\hat{T}]} G_m(|t-s|)\, d\nu_s d\nu_t,$$

for an arbitrary $\nu \in \mathcal{P}_{[0,\hat{T}]}$. Choosing, for example, ν to be the uniform distribution with respect to the Lebesgue measure, we easily conclude. ■

Next, we recall that, for each $m \geq 1$,

$$E(IS(\nu^*_m)) = \frac{1}{2} V^2 \mu^*_m \quad \text{and} \quad Var(IS(\nu^*_m)) = V^2 \sigma^2_m \int_{(0,\hat{T}]} \left(\nu^*_m\left(\left[t,\hat{T}\right]\right)\right)^2 dt.$$

Therefore,

$$E\left(IS\left(\nu^*_{m+1}\right)\right) \leq E(IS(\nu^*_m)) \quad \text{and} \quad \lim_{m\uparrow\infty} E(IS(\nu^*_m)) = \frac{1}{2} \lim_{m\uparrow\infty} V^2 \mu^*_m = 0.$$

On the other hand,

$$\lim_{m\uparrow\infty} Var(IS(\nu^*_m)) = \lim_{m\uparrow\infty} V^2 \sigma^2_m \int_{(0,\hat{T}]} \left(\nu^*_m\left(\left[t,\hat{T}\right]\right)\right)^2 dt.$$

Thus, depending on appropriate conditions on the sequence

$$\left(\sigma^2_m \int_{(0,\hat{T}]} \left(\nu^*_m\left(\left[t,\hat{T}\right]\right)\right)^2 dt\right)_{m\geq 1}$$

we may have

$$\lim_{m\uparrow\infty} Var(IS(\nu^*_m)) < \infty \quad \text{or} \quad \lim_{m\uparrow\infty} Var(IS(\nu^*_m)) = \infty$$

while, *however*, $\lim_{m\uparrow\infty} E(IS(\nu^*_m)) = 0$.

This motivates us to work towards introducing alternative criteria to (29) (and, in turn, to (15)).

4.3 An extended criterion balancing the mean and variance of the implementation slippage

In order to remedy the aforementioned shortcoming of Gatheral's model we could consider an optimality criterion that includes, in addition to its mean, the variance of the implementation slippage. Recall that, for $\nu \in \mathcal{P}_{[0,T]}$, the variance of the associated implementation slippage is given by

$$Var\left(IS\left(\nu\right)\right) = V^2\sigma^2 \int_{\left(0,\hat{T}\right]} \left(\nu\left(\left[t,\hat{T}\right]\right)\right)^2 dt.$$

We may then consider the *constrained optimization problem*

$$\begin{cases} \min_{\nu\in\mathcal{P}_{[0,T]}} \left(\frac{1}{2}V^2 \int_{\left[0,\hat{T}\right]} \int_{\left[0,\hat{T}\right]} G\left(\left|t-s\right|\right) d\nu_s d\nu_t\right) \\ \text{with} \;\; V^2\sigma^2 \int_{\left(0,\hat{T}\right]} \left(\nu\left(\left[t,\hat{T}\right]\right)\right)^2 dt > \chi > 0 \end{cases}, \tag{63}$$

where χ represents a variance threshold. The above problem can be simplified, setting $k := \frac{\chi}{\sigma^2 V^2}$, to

$$\begin{cases} \min_{\nu\in\mathcal{P}_{[0,T]}} \left(\frac{1}{2} \int_{\left[0,\hat{T}\right]} \int_{\left[0,\hat{T}\right]} G\left(\left|t-s\right|\right) d\nu_s d\nu_t\right) \\ \text{with} \quad \int_{\left(0,\hat{T}\right]} \left(\nu\left(\left[t,\hat{T}\right]\right)\right)^2 dt > k > 0 \end{cases}. \tag{64}$$

More generally, we may introduce constrained optimization criteria of the form

$$\begin{cases} \min_{\nu\in\mathcal{P}_{[0,T]}} \left(\frac{1}{2}V^2 \int_{\left[0,\hat{T}\right]} \int_{\left[0,\hat{T}\right]} G\left(\left|t-s\right|\right) d\nu_s d\nu_t\right) \\ \text{with} \;\; 0 < \chi_1 < V^2\sigma^2 \int_{\left(0,\hat{T}\right]} \left(\nu\left(\left[t,\hat{T}\right]\right)\right)^2 dt < \chi_2 \end{cases}, \tag{65}$$

for given thresholds χ_1, χ_2.

From the optimization point of view, (63) and (65) are two-dimensional constrained calculus of variations problems. They are currently being examined by the authors.

An important observation here is that because ν is a probability measure, and thus its total mass is constrained to 1, the variance constraints in (63) and (65) might *not always be viable* for a given triplet $(V, \sigma, \hat{T})$. This, in turn, raises questions about a possible interplay between variance thresholds and choices for the *volatility parameter.* The choice of volatility is more critical in the sequential order case and is discussed in detail in [5].

To our knowledge, problems (63) and (65) have not been considered before within the framework developed in [4] and [2].

5 Realized, permanent and temporary impacts

Our focus, so far, has been on the analysis of the implementation slippage. Next, we revert our attention to the *impact* that execution strategies have on the stock price process. This is widely discussed in the literature and, not surprisingly, there are differences in the way this impact is measured. Following the various approaches in the literature, we analyze the *realized, permanent* and *temporary price impacts* as defined in [1]. For this, we first recall the quantities:

- S_0 is the stock price before the order execution begins
- S_{post} is the market price after the order is completed, and
- $\overline{S}$ is the volume-weighted average realized price on the order.

The post trade price S_{post} should capture permanent effects of the order execution on the market prices. The analysis in [1] provides information on reasonable times after the completion of the order at which that S_{post} should be measured. Among others, empirical analysis suggests that one half-hour is adequate to achieve this. Consequently, if execution stops at time 1, then $S_{post} = S_{1+\delta}$, where δ is this extra time. We thus have the following definition.

Definition 18 *i) The permanent impact on the stock price caused by execution strategy* $\nu \in \mathcal{P}_{[0,T]}$ *is defined as*

$$I(\nu) := \frac{S_{post} - S_0}{S_0},$$

while the realized impact is defined as

$$J(\nu) := \frac{\overline{S} - S_0}{S_0},$$

where $\overline{S}$ *is the volume-weighted average realized price on the order.*

ii) The temporary impact, denoted by $TI(\nu)$, *is defined as the difference between the realized and the permanent impact,*

$$TI(\nu) = J(\nu) - I(\nu).$$

We stress that all computations in [1] are performed in *volume time,* denoted by τ_t, which represents the *fraction of an average market volume that has been executed up to and including clock time* t. Note that market volume time corresponds to our cdf ν_t which represents the cumulative fraction of the volume, for a given order, executed up to time t. This representation of a strategy is convenient because it does *not depend* on the specific volume to be executed in the market.

Until now, we did not take into consideration in our model the volume traded in the market. Clearly, an order whose volume exceeds what is traded cannot be executed, an assumption that has been so far missed in the single order models.

5.1 Impacts for admissible execution strategies

To align the notation of [1] within our framework, we assume from now on that the permanent impact of execution policy $\nu \in \mathcal{P}_{[0,T]}$ is defined as

$$I(\nu) := \frac{S_T - S_0}{S_0}.$$

Note that we choose the post time to be T, while the execution of the order is completed at time $\hat{T}$; hence, the *time gap* is given by $\delta := T - \hat{T}$. The order is executed over the interval $\left[0, \hat{T}\right]$ with $\hat{T} < T$, and we recall that the process S is left continuous with right limits on $\left[0, \hat{T}\right]$ and continuous on $\left(\hat{T}, T\right]$. Therefore,

$$S_T - S_0 = V \int_{[0,T]} G(T - s)\, d\nu_s + \sigma (W_T - W_0). \tag{66}$$

Consequently,

$$I(\nu) = \frac{V \int_{[0,T]} G(T-s)\, d\nu_s + \sigma (W_T - W_0)}{S_0}.$$

Renormalizing the decay function: In order to compare our results with the ones in [1] we need to reconcile the notation and the quantities used in each of these models. Recall that process S represents price, say, in dollars, while, in the original formulation, the strategy X and volume V represent number of shares. As a consequence, *unit inconsistency* arises in both the original and the modified (involving the measure ν) formulation. To remedy this, we work as follows.

Observe that the performance of any admissible policy ν_t for a given function G does not change if we multiply G by a constant, say C. It then turns out that to compare the model in [1] with (66), one simply needs to take $C = S_0$. The new function, denoted by $\tilde{G}$, is then given by

$$\tilde{G}(|t|) := S_0 G(|t|), \ t \in \mathbb{R}. \tag{67}$$

Renormalizing the volatility parameter: In (66), the volatility represents the so-called normal volatility of the Bachelier model while in [1] the volatility represents the log-normal volatility of the Black and Scholes model. In order to compare both models we only need to convert the normal volatility to the log-normal one, which is done by putting in our model

$$\tilde{\sigma} = S_0 \sigma, \tag{68}$$

where now σ represents the log-normal volatility.

This change in notation brings model (66) closer to the framework of [1]. Still, a difference remains, namely, how *time* is treated in [1]. Model (66) runs in *clock time* while the model in [1] runs in *volume time*. We will focus on this difference later on. For now, we only introduce a new parametrization to our model, which still runs in clock time, namely

$$S_t = S_0 + V \int_{[0,t)} \tilde{G}(t-s)\, d\nu_s + \tilde{\sigma}(W_t - W_0), \quad t \in [0,T), \tag{69}$$

where $\tilde{G}$ and $\tilde{\sigma}$ are given by (67) and (68), respectively. Under these adjustments, the permanent price impact takes the form

$$I(\nu) = V \int_{[0,T]} G(T-s)\, d\nu_s + \sigma(W_T - W_0). \tag{70}$$

Proposition 19 *The permanent price impact $I(\nu)$ is an $\mathcal{F}_T-$measurable random variable with mean*

$$E(I(\nu)) = \frac{V\int_{[0,T]} \tilde{G}(T-s)\, d\nu_s}{S_0} = V \int_{[0,T]} G(T-s)\, d\nu_s$$

and variance

$$Var(I(\nu)) = \frac{\tilde{\sigma}^2}{S_0^2} T = \sigma^2 T,$$

where σ denotes the log-normal volatility.

The realized impact $J(\nu)$ defined in [1] corresponds, using the modified price dynamics (69), to

$$J(\nu) = \frac{\int_{[0,T]} S_s d\nu_s - S_0}{S_0},$$

where $\int_{[0,T]} S_s d\nu_s$ represents the *realized* price. Using (69) gives

$$\int_{[0,T]} S_s d\nu_s = S_0 \left(1 + V \int_{[0,T]} \int_{[0,t)} G(t-s)\, d\nu_s d\nu_t + \int_{[0,T]} \sigma W_t d\nu_t\right)$$

and, hence,

$$J(\nu) = V \int_{[0,T]} \int_{[0,t)} G(t-s)\, d\nu_s d\nu_t + \int_{[0,T]} \sigma W_t d\nu_t. \tag{71}$$

Proposition 20 *The realized impact $J(\nu)$ is an $\mathcal{F}_T-$measurable random variable with mean*

$$E(J(\nu)) = V \int_{[0,T]} \int_{[0,t)} G(t-s)\, d\nu_s d\nu_t$$

and variance

$$Var(J(\nu)) = \sigma^2 \int_{[0,T]} \int_{[0,T]} t \wedge s d\nu_s d\nu_t = \sigma^2 \int_{(0,T]} (\nu([t,T]))^2\, dt.$$

Next, we consider the *relative to the value of volume traded*, (i.e. relative to S_0V) *implementation slippage*, defined as

$$R(\nu) := \frac{IS(\nu)}{S_0V}.$$

Proposition 21 *The relative to the value of volume traded implementation slippage is given by*

$$R(\nu) = \frac{1}{2}V\int_{[0,T]}\int_{[0,T]} G(t-s)\,d\nu_s d\nu_t + \int_{[0,T]} \sigma W_t d\nu_t. \tag{72}$$

It is an $\mathcal{F}_T$-random variable, normally distributed with mean

$$E(R(\nu)) = \frac{1}{2}V\int_{[0,T]}\int_{[0,T]} G(|t-s|)\,d\nu_s d\nu_t$$

and variance

$$Var(R(\nu)) = \sigma^2\int_{(0,T]} (\nu([t,T]))^2\,dt.$$

Remark 22 *Concepts of permanent, realized and temporary price impacts defined in [1] are derived within models referring to the rate of trading. Such models have continuous trajectories. In our framework, the price process may have jumps and the jumps are priced using space and not time infinitesimal arguments. When the execution strategy is infinitesimal in time, we refer to the rate of trading. However, when the execution strategy involves a jump (which can be associated with a block trade) then, in a fixed time, this jump is priced through infinitesimal in space arguments. Here, the order book dynamics are used to develop the price concept. It turns out, as we have demonstrated before, that the block trades are implemented at the average price of S calculated before and after the jump they generate.*

We further deduce that

$$R(\nu) - J(\nu) = \frac{1}{2}G(0)\,V\sum_{0\le t\le T}(\Delta\nu_t)^2.$$

Moreover,

$$\begin{aligned} J(\nu) - I(\nu) &= V\int_{[0,T]}\int_{[0,t)} G(t-s)\,d\nu_s d\nu_t + \int_{[0,T]}\sigma W_t d\nu_t \\ &\quad - \left(V\int_{[0,T]} G(T-s)\,d\nu_s + \sigma(W_T - W_0)\right) \\ &= V\int_{[0,T]}\int_{[0,t)} G(t-s)\,d\nu_s d\nu_t - V\int_{[0,T]} G(T-s)\,d\nu_s \end{aligned}$$

$$+\int_{[0,T]}\sigma W_t d\nu_t-\sigma\left(W_T-W_0\right).$$

In accordance with the definition given in [1], the quantity

$$K\left(\nu\right):=V\int_{[0,T]}\int_{[0,t)}G\left(t-s\right)d\nu_s d\nu_t-V\int_{[0,T]}G\left(T-s\right)d\nu_s$$

represents the *temporary impact caused by strategy* ν_t. We also have

$$R\left(\nu\right)-J\left(\nu\right)+I\left(\nu\right)=\frac{1}{2}G\left(0\right)V\sum_{0\leq t\leq T}\left(\Delta\nu_t\right)^2$$

$$+V\int_{[0,T]}G\left(T-s\right)d\nu_s+\sigma\left(W_T-W_0\right).$$

Introducing

$$L\left(\nu\right):=\frac{1}{2}G\left(0\right)V\sum_{0\leq t\leq T}\left(\Delta\nu_t\right)^2+V\int_{[0,T]}G\left(T-s\right)d\nu_s,$$

we have the following decomposition

$$R\left(\nu\right)-J\left(\nu\right)+I\left(\nu\right)=L\left(\nu\right)+\sigma\left(W_T-W_0\right).\tag{73}$$

Remark 23 *We note that the term* $V\int_{[0,T]}G\left(T-s\right)d\nu_s$ *represents the mean of the permanent impact in the terminology of [1]. The term* $\frac{1}{2}G\left(0\right)V\sum_{0\leq t\leq T}\left(\Delta\nu_t\right)^2$ *represents the impact of jumps which are, however, excluded in the model used in [1].*

5.2 Optimal strategy and the power law

The optimal measure ν^* satisfies for, all $t\in\left[0,\hat{T}\right]$,

$$\int_{[0,\hat{T}]}\tilde{G}\left(t-s\right)d\nu_s^*=\tilde{\mu}^*\tag{74}$$

with $\tilde{G}$ as in (67) and $\tilde{\mu}^*:=S_0\mu^*$.

Proposition 24 *The random variables* $I\left(\nu^*\right),\ J\left(\nu^*\right)$ *and* $R\left(\nu^*\right)$ *satisfy*

$$E\left(I\left(\nu^*\right)\right)=V\mu^*\quad and\quad E\left(J\left(\nu^*\right)\right)=\frac{1}{2}V\left(\mu^*-2G\left(0\right)\left(\Delta\nu_0^*\right)^2\right),$$

and

$$E\left(R\left(\nu^*\right)\right)=\frac{1}{2}V\mu^*.$$

Moreover,

$$E\left(J\left(\nu^*\right)\right)<E\left(R\left(\nu^*\right)\right)\qquad and\quad Var\left(J\left(\nu^*\right)\right)=Var\left(R\left(\nu^*\right)\right).$$

Extensive empirical analysis (see [1] and references within) validates the so-called *power law,* which states that the *average relative price change* is adequately described by the quantity

$$\text{Average relative price change} \quad \sim \quad Y\sigma\left(\frac{Q}{V}\right)^{\delta}, \tag{75}$$

where σ is the daily volatility of the asset, Q represents the volume of a metaorder, and V is the daily traded volume.

The numerical constant Y is of order unity. The daily volatility σ and the daily volume V are measured contemporaneously to the trade. As indicated in [6], the power law holds for the levels of the ratio $\frac{Q}{V}$ ranging from a few 10^{-4} to a few 10^{-2}. Depending on the market and, also, on the contract types, the power δ varies between 0.5 and 0.7.

Applied to the optimal strategy ν^*, the power law states that

$$V\mu^* = \lambda^* = Y\sigma_D\left(\frac{V}{V}\right)^{\delta}, \tag{76}$$

where σ_D is the asset volatility on day D and V_D is the volume traded on this day, both measured contemporaneously to the trade.

Note, however, that the quantities σ_D and V_D might *not be measurable* with respect to $\mathcal{F}_0$, but λ^* is and, hence, the above equality *cannot hold true.* This poses a problem to the modified Gatheral's model (66) in which the order characteristics, the decay function and the volatility must be measurable with respect to $\mathcal{F}_0$.

One way to address this is to replace the power law by its estimate based on the data which are not contemporaneous to the trade but use the daily volatilities and volumes collected on the days *proceeding the trade.* Then, the power law becomes

$$V\mu^* = \lambda^* = YV^{\delta}E\left(\frac{\sigma_D}{V_D^{\delta}}\,|\mathcal{F}_0\right),$$

where $E\left(\frac{\sigma_D}{V_D^{\delta}}\,|\mathcal{F}_0\right)$ *represents an* $\mathcal{F}_0-$*measurable estimate* of the quantity $\frac{\sigma_D}{V_D^{\delta}}$.

5.3 Effects on the choice of model components

We conclude examining the relation between the average permanent impact and the choice of the decay function for which (75) holds. Once more, we discuss this in the realm of the exponential case where we seek to specify the value of the parameter κ for which (76) holds. Recall that, in this case, $\mu^* = \frac{\lambda^*}{V} = \frac{2}{2+\kappa\Delta}$ and, hence, we must have

$$\frac{2}{2+\kappa T} = \frac{1}{V}Y\sigma_D\left(\frac{V}{V_D}\right)^{\delta}.$$

Therefore, the parameter κ must satisfy

$$\kappa = \frac{2}{T}\left(\frac{V^{1-\delta}V_D^{\delta}}{Y\sigma_D} - 1\right).$$

Note, however, that because $\kappa > 0$, we must have

$$V^{1-\delta}V_D^{\delta} > Y\sigma_D,$$

which, in turn, imposes viability *constraints* for the various model inputs.

References

[1] R. Almgren, C. Thum, E. Hauptmann, and H. Li, Direct estimation of equity market impact. Preprint.

[2] G. Bordigoni, A. Figalli, A. Ledford, P. Ustinov, Strategic execution Strategies, Preprint, November 2020

[3] S. N. Cohen, R. J. Elliott, Stochastic Calculus and Applications, Second Edition, *Probability and Its Applications*, Birkhauser, 2015

[4] J. Gatheral, A. Schied, A. Slynko, Transient linear price impact and Fredholm integral equations, Mathematical Finance: An International Journal of Mathematics, *Statistics and Financial Economics*, 22 (3), 445-474, 2012

[5] M. Musiela and T. Zariphopoulou, Optimal execution of sequential orders, preprint.

[6] B. Toth, Y. Lemperiere, C. Deremble, J. de Lataillade, J. Kockelkoren, J. -P. Bouchaud, Anomalous price impact and the critical nature of liquidity in financial markets, June 29, 2021 arXiv: 1105.1694v1