

Optimal Trading with Linear and (small) Non-Linear Costs

A. Rej¹, R. Benichou¹, J. de Lataillade¹,
G. Zérah¹ & J.-Ph. Bouchaud^{1,2}

1: Capital Fund Management,
23 rue de l'Université, 75007 Paris, France

2: CFM-Imperial Institute of Mathematical Finance
Imperial College, London, SW7 2AZ, UK

January 12, 2016

Abstract

We reconsider the problem of optimal trading in the presence of linear and quadratic (market impact) costs for arbitrary linear costs but in the limit where quadratic costs are small. Using matched asymptotic expansion techniques, we find that the trading speed vanishes inside a band that is narrower in the presence of market impact by an amount that scales as a cube root of the market impact parameter. Outside the band we find three regimes: a small boundary layer where the velocity vanishes linearly with the distance to the band, an intermediate region where the velocity behaves as a square-root and an asymptotic region where it becomes linear again. Our solution is consistent with available numerical results. We determine the conditions under which our expansion is useful in practical applications and generalize our solution to other forms of non-linear costs.

1 Introduction

Determining the optimal trading strategy in the presence of a predictive signal and transaction costs is of utmost importance for quantitative asset managers, since too much trading (both in volume and frequency) can quickly deteriorate the performance of a strategy, or even make the strategy a money-losing machine. The detailed structure of these costs is actually quite complex. Some costs are called “linear” because they simply grow as ΓQ , where Q is the traded volume and Γ the linear cost parameter. These are due to various fees (market fees, brokerage fees, etc.) or the bid-ask spread and they usually represent a small fraction of the amount traded (typically 10^{-4} on liquid markets, but sometimes much more in OTC/illiquid markets). Much more subtle are impact-induced costs, which come from the fact that a large order must be split into a sequence of small trades that are executed gradually. But since each executed trade, on average, impacts the price in

the direction of the trade, the average execution price is higher (if one buys) than the decision price, leading to what is called “execution shortfall”. This cost clearly increases faster than Q , since the price impact itself increases with the size of the trade. There seems to be a wide consensus now that the impact-induced costs are on the order of $\sigma Q^{3/2}/V^{1/2}$, where σ is the daily volatility and V the daily turnover (see [1, 2, 3] for recent accounts).

From a theoretical point of view, however, the $Q^{3/2}$ dependence of the costs makes the analysis difficult. As a simplifying assumption, one often replaces the empirical $Q^{3/2}$ behaviour by a “quadratic cost” formula ηQ^2 , so that the price impact is proportionally to Q , see e.g. [4, 5]. In the absence of linear costs ($\Gamma = 0$), the optimal strategy may be found as a result of a simple quadratic optimisation problem, see for example [4, 6]. The optimal policy is to rebalance at finite speed towards the target portfolio. This results in a position that is an exponential moving average of the trading signal. The pure linear cost problem (i.e. $\eta = 0$) was independently solved, in slightly different contexts, in [7, 8]. It requires *instantaneous* rebalancing towards a finite band around the ideal position, and no action inside the band, also called the no-trade (NT) region. The case where both linear and quadratic costs are present is of course highly interesting and no exact solution is known at this stage. An approximate solution was proposed in [9]. A method for constructing the exact solution in the small cost limit where *both* Γ and η tend to zero can be found in [10]. The aim of our paper is to show that one can in fact relax the assumption that Γ is small and expand around the general solution for linear cost, the expansion parameter being $\eta \rightarrow 0$. We will see that the solution defines four different regions (*c.f.* Figure 1):

- a) the no-trade (NT) region inside a band around the ideal position is still present but the band shrinks by an amount $\sim \eta^{1/3}$;
- b) a small “boundary layer” of width $\eta^{1/3}$ surrounding the band; the trading speed is on the order $\eta^{-1/3}$ and takes a scaling form;
- c) further away from the band, but still within its zone of influence, the trading speed is on the order $\eta^{-1/2}$ and behaves as a square-root of the distance to the band;
- d) finally, far away from the band, the trading speed is a linear function of the distance to the ideal position and one recovers the exact $\Gamma = 0$ solution as expected.

Our method in fact readily generalizes to other non-linear cost structures and in particular to the $Q^{3/2}$ law alluded to above. We briefly discuss how our results extend to this case in the final section of this paper.

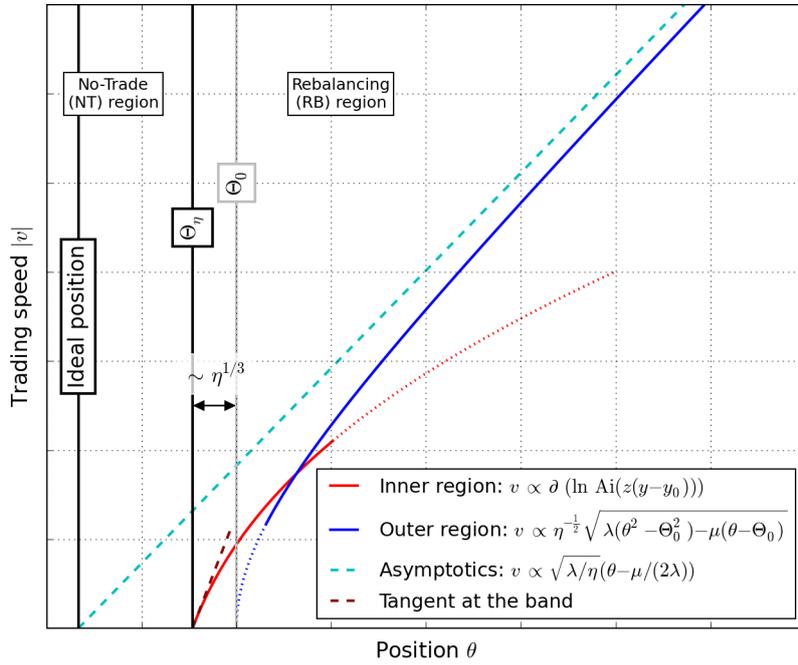


Figure 1: Trading speed as a function of the distance to the boundary of the NT region. The outer solution behaves as a square-root close to the band and linearly farther out. The square-root singularity is regularized in a boundary layer of width $\eta^{1/3}$ so that the velocity vanishes linearly at the (shifted) boundary of the band Θ_η .

2 Set up of the problem and the $\Gamma = 0$ solution

Following [7] we assume that the value X_t of the traded instrument has a mean-reverting dynamics governed by the following drift-diffusion equation¹

$$dX_t = \mu(X_t)dt + \sigma dW_t. \quad (1)$$

We will call \mathcal{L}_x the associated Itô differential operator

$$\mathcal{L}_x[f] = \mu(x)\frac{\partial f}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x^2}. \quad (2)$$

The position (number of shares/lots, etc.) of the manager at time t is denoted by θ_t . For a given rebalancing policy, the expected risk-adjusted P&L per unit time, conditional on $X_t = x$ and $\theta_t = \theta$, is given by

$$W(\theta, x|t) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^\infty dt' e^{-(t'-t)/T} \left[\mu(X_{t'}) \theta_{t'} - \lambda \theta_{t'}^2 - \Gamma |\dot{\theta}_{t'}| - \eta \dot{\theta}_{t'}^2 \right], \quad (3)$$

where λ is the cost of risk (that includes a factor σ^2). The first term is the average gain of the position, the last two terms are rebalancing costs. We now introduce the value function $V(\theta, x)$, defined as $W(\theta, x|t)$ for the optimal future rebalancing policy. Note that because we assume a stationary process for X_t , the value function is in fact independent of t ². As is well known, $V(\theta, x)$ then obeys an HJB equation that in the present case reads

$$0 = \mu(x)\theta - \lambda\theta^2 + \max_{\dot{\theta}} \left\{ -\Gamma |\dot{\theta}| - \eta \dot{\theta}^2 + \frac{\partial V}{\partial \theta} \dot{\theta} + \mathcal{L}_x[V] \right\}. \quad (4)$$

The maximisation with respect to $\dot{\theta}$ is very simple and leads to:

$$v = \dot{\theta}^* = \frac{1}{2\eta} \left[\frac{\partial V}{\partial \theta} - \Gamma \text{sign}(\dot{\theta}^*) \right] \quad \text{or} \quad v = 0, \quad (5)$$

where the NT region ($v = 0$) is defined by $|\frac{\partial V}{\partial \theta}| \leq \Gamma$. In this region the HJB equation simplifies to

$$\mathcal{L}_x[V_{\text{NT}}] = -\mu(x)\theta + \lambda\theta^2. \quad (6)$$

In the rebalancing (RB) region, on the other hand, the HJB equation becomes a non-linear PDE equation

$$\left(\frac{\partial V^\pm}{\partial \theta} \pm \Gamma \right)^2 = 4\eta^2 v^2 = 4\eta [\lambda\theta^2 - \mu(x)\theta - \mathcal{L}_x[V^\pm]], \quad (7)$$

where the \pm sign corresponds, respectively, to large enough positive θ s such that the optimal policy is to sell ($v < 0$), or to large enough negative θ s such that the optimal policy is to buy ($v > 0$).

¹The diffusion constant σ^2 can also depend on X_t , as in [7], without materially impacting the following results. For the sake of simplicity, we keep σ constant.

²Technically, it can be useful to keep T large but finite, which amounts to regularizing the differential operator $\mathcal{L}_x[f]$ with a term $-f/T$ and taking the limit $T \rightarrow \infty$.

3 The $\eta = 0$ solution

For $\eta = 0$, the solution of the corresponding HJB equation has been worked out by Martin & Schöneborn (M&S) in [7], and will be denoted by $V_0(\theta, x)$. The NT region is parameterized by two functions $\Theta_0^-(x), \Theta_0^+(x)$, such that for a given x , the speed of trading v vanishes inside the interval $[-\Theta_0^-(x), \Theta_0^+(x)]$, hereafter referred to as a band. Outside the band, the $\eta = 0$ solution to Eq. (7) is given by

$$V_0^\pm(\theta, x) = V_{\text{NT},0}(\pm\Theta_0^\pm(x), x) - \Gamma |\theta \mp \Theta_0^\pm(x)|. \quad (8)$$

Inside the band, the general solution to the linear equation (6) can be constructed using the Green's function \mathcal{G}_x of the operator \mathcal{L}_x and the two independent solutions $\psi_{1,2}(x)$ of the homogeneous equation $\mathcal{L}_x f = 0$, see [7] for details. Schematically,

$$V_{\text{NT},0}(\theta, x) = \mathcal{G}_x[-\mu(x)\theta + \lambda\theta^2] + \alpha_1(\theta)\psi_1(x) + \alpha_2(\theta)\psi_2(x), \quad (9)$$

where $\alpha_{1,2}(\theta)$ are two yet-to-be-determined functions. The reference [7] proposes to fix these functions in two steps. First, one imposes that at the (still unknown) boundaries of the NT zone, the derivative of $V_0(\theta, x)$ are continuous, i.e.

$$\left. \frac{\partial V_{\text{NT},0}(\theta, x)}{\partial \theta} \right|_{\theta=\Theta_0^\pm(x)} = \mp \Gamma. \quad (10)$$

This allows one to solve for $\alpha_{1,2}(\theta)$ as functionals of the boundary positions $\Theta_0^\pm(x)$. Second, one determines these boundaries by invoking the variational argument, i.e that these boundaries should maximize the value function $V_{0,\text{NT}}(\theta, x)$ everywhere in the NT region. This second condition allows to fully determine $\Theta_0^\pm(x)$. While we fully agree with the final expressions obtained by M&S, we argue that their second condition does not generalize to the case $\eta \neq 0$. The general condition should rather be that the second derivative of the value function with respect to θ is continuous everywhere, including the boundaries between NT and RB regions – see Appendix A. In fact, we show in Appendix B that the M&S solution obeys

$$\left. \frac{\partial^2 V_{\text{NT},0}(\theta, x)}{\partial \theta^2} \right|_{\theta=\Theta_0^\pm(x)} \equiv 0, \quad (11)$$

a property that apparently went unnoticed in [7] and that is actually much simpler than the variational condition. In the next section we will attempt to construct a consistent solution to the HJB for arbitrary Γ but small quadratic costs $\eta \rightarrow 0$. We will make use of the continuity of the first and second derivative of the value function to determine the new location of the boundaries.

4 The small η matched asymptotic expansions

4.1 The outer region

Let us assume that, far enough from the new band positions $[-\Theta_\eta^-(x), \Theta_\eta^+(x)]$ (called the ‘‘outer region’’), the trading solution for small η reads:

$$V_\eta^\pm(\theta, x) = V_0^\pm(\theta, x) + \sqrt{\eta}V_1^\pm(\theta, x) + \dots \quad (\eta \rightarrow 0). \quad (12)$$

In the sequel we will confine ourselves to the ‘+’ sector where the trading speed is negative and we will drop the superscripts on the position of the band Θ_0 and on V_0, V_1 . Plugging our ansatz into Eq. (7) and retaining leading terms in η gives

$$\frac{\partial V_1}{\partial \theta} = -2\sqrt{\lambda\theta^2 - \mu(x)\theta - \mathcal{L}_x[V_0]}, \quad (13)$$

where we have used that $\frac{\partial V_0}{\partial \theta} = -\Gamma$ so that the zeroth-order term in the LHS vanishes. Using the solution outside the band (8) we write

$$\mathcal{L}_x[V_0] = \mathcal{L}_x[V_{\text{NT},0}(\Theta_0(x), x)] + \Gamma\mathcal{L}_x[\Theta_0(x)]. \quad (14)$$

Note that

$$\mathcal{L}_x[V_{\text{NT},0}(\theta, x)] = -\mu(x)\theta + \lambda\theta^2, \quad (15)$$

from which we deduce (the dependence of μ and Θ_0 on x is henceforth suppressed)

$$\begin{aligned} \mathcal{L}_x[V_{\text{NT},0}(\Theta_0, x)] &= -\mu\Theta_0 + \lambda\Theta_0^2 + \mu\Theta_0'(V_{\text{NT},0})_\theta \\ &+ \frac{\sigma^2}{2} [\Theta_0'^2(V_{\text{NT},0})_{\theta\theta} + \Theta_0''(V_{\text{NT},0})_\theta + 2\Theta_0'(V_{\text{NT},0})_{x\theta}], \end{aligned} \quad (16)$$

where the ‘‘prime’’ stands for derivatives wrt x and the subscripts indicate variables (other than x) with regard to which derivatives are taken. The boundary conditions at Θ_0 are

$$\forall x : \quad (V_{\text{NT},0})_\theta(\Theta_0(x), x) \equiv -\Gamma, \quad (V_{\text{NT},0})_{\theta\theta}(\Theta_0(x), x) \equiv 0, \quad (17)$$

implying

$$(V_{\text{NT},0})_{x\theta} = -\Theta_0'(V_{\text{NT},0})_{\theta\theta} \equiv 0, \quad (18)$$

so that $\mathcal{L}_x[V_0]$ finally simplifies to:

$$\mathcal{L}_x[V_0] \equiv -\mu\Theta_0 + \lambda\Theta_0^2. \quad (19)$$

Therefore, the equation for V_1 becomes:

$$\frac{\partial V_1}{\partial \theta} = -2\sqrt{\lambda(\theta^2 - \Theta_0^2) - \mu(\theta - \Theta_0)}. \quad (20)$$

Now, the velocity in the trading zone is simply

$$v = \frac{1}{2\sqrt{\eta}} \frac{\partial V_1}{\partial \theta}, \quad (21)$$

We thus find that: a) v diverges as $\eta^{-1/2}$ when $\eta \rightarrow 0$ recovering the instantaneous rebalancing in this limit; b) v behaves linearly for large $|\theta|$; c) v behaves as a square-root close to (see discussion below) the unperturbed band Θ_0

$$v \approx -\sqrt{\frac{2\lambda\Theta_0 - \mu}{\eta}} \sqrt{\theta - \Theta_0}. \quad (22)$$

This square root singularity is interesting because it means that there must be a region very close to the band where this naive perturbative solution breaks down. Indeed, the second derivative of $\sqrt{\eta}V_1$ wrt θ diverges and thus may not be neglected. One has to analyse this region by zooming in on the immediate proximity of the band conventionally dubbed the ‘‘inner region’’ or the boundary layer, see [11].

4.2 The inner region

To make a start on the analysis, we take the derivative of Eq. (7) with respect to θ and introduce $f := \partial V^+ / \partial \theta + \Gamma \equiv 2\eta v$. This leads to the exact equation

$$f f_\theta = -2\eta(\mu - 2\lambda\theta + \mathcal{L}_x[f]). \quad (23)$$

We postulate that close to the (new) band $\Theta_\eta(x)$ the function f exhibits the following scaling

$$f = -\eta^\alpha F(y), \quad \text{where} \quad y := \frac{\theta - \Theta_\eta(x)}{\eta^\beta} > 0, \quad \alpha, \beta > 0. \quad (24)$$

The parameters α and β are two exponents that need to be determined and $F(y)$ is a positive function (note indeed that $f \propto v < 0$ in the ‘+’ sector that we are considering here). A first condition comes from the fact that when $y \rightarrow \infty$, this scaling form must reproduce the above square-root solution, i.e., $F(y) \simeq A\sqrt{y}$ for large y . This requires

$$\eta^{\alpha-\beta/2} A = 2\sqrt{\eta} \sqrt{2\lambda\Theta_0 - \mu}, \quad (25)$$

leading to

$$\alpha - \beta/2 = 1/2, \quad A = 2\sqrt{2\lambda\Theta_0 - \mu}. \quad (26)$$

Injecting the scaling form into Eq. (23) and noting that derivatives with respect to x supply factors of $\eta^{-\beta}$, we find that the *leading* terms are

$$\eta^{2\alpha-\beta} F F_y = 2\eta \left(-\mu + 2\lambda\Theta_0 + \frac{\sigma^2}{2} \eta^{\alpha-2\beta} \Theta_0'^2 F_{yy} \right). \quad (27)$$

Matching the η powers of the two sides of the equation leads to

$$2\alpha - \beta = 1 \quad \text{or} \quad 2\alpha - \beta = 1 + \alpha - 2\beta, \quad (28)$$

where the first case only holds if $\alpha \geq 2\beta$ so that the last term in the RHS is negligible. This however leads to

$$(F^2)_y = 4(2\lambda\Theta_0 - \mu) \equiv A^2, \quad (29)$$

which still has a square-root singularity where F goes to zero, so that close enough to the singularity the last term of the RHS diverges and cannot be neglected contradicting our original assumption.

The only other possibility is $2\alpha - \beta = 1 + \alpha - 2\beta$, which, together with $\alpha - \beta/2 = 1/2$, leads to $\alpha = 2/3$ and $\beta = 1/3$. The leading ODE for F takes the following form

$$(F^2)_y - BF_{yy} = A^2, \quad B \equiv 2\sigma^2\Theta_0'^2. \quad (30)$$

Upon integration

$$F^2 - BF_y = A^2(y - y_0). \quad (31)$$

The solution to the above may be expressed in terms of Airy functions. Writing $F = -B\Psi_y/\Psi$, the above equation reads

$$\Psi_{yy} = \frac{A^2}{B^2}(y - y_0)\Psi. \quad (32)$$

We want a solution to this equation such that $\Psi_y(y = 0) = 0$ (i.e. $F(y = 0) = 0$) such that $F(y > 0) > 0$. The general solution is

$$\Psi(y) = c_1\text{Ai}(z(y - y_0)) + c_2\text{Bi}(z(y - y_0)), \quad z = \left(\frac{A}{B}\right)^{2/3}, \quad (33)$$

but the asymptotic behaviour of the Airy functions imposes $c_2 = 0$. The condition $\text{Ai}'(-zy_0) = 0$ selects the first maximum of Ai that occurs for $-zy_0 \approx -1.018\dots$, thereby fixing y_0 . Finally, the sought-after solution is

$$F(y) = -Bz \frac{\text{Ai}'(z(y - y_0))}{\text{Ai}(z(y - y_0))}. \quad (34)$$

For large y , one uses the asymptotic behaviour $\ln \text{Ai}(u) \approx -2u^{3/2}/3$ to obtain

$$F(y) \approx Bz^{3/2}\sqrt{y} \equiv A\sqrt{y}, \quad (35)$$

as desired. We have found a solution that goes smoothly to zero when $y \rightarrow 0$, i.e. in a region of width $\eta^{1/3}$ immediately outside the boundary of the band $-\Theta_\eta$, see Figure 1. Note that $F_y(0) = 1.018\dots A^{4/3}B^{-1/3}$.

4.3 Boundaries shift inwards

Next, we need to find the shifted band position Θ_η . We will use the fact that the second derivative of the value function should be continuous at the boundary. Integrating f leads to the following equality for the value function in the trading zone

$$V^+(\theta, x) = \int_{\Theta_\eta}^{\theta} d\theta' f(\theta', x) - \Gamma(\theta - \Theta_\eta) + V_{\text{NT},\eta}(\Theta_\eta, x), \quad (36)$$

that by construction coincides with $V_{\text{NT},\eta}$ at the boundaries of the band. Observe that $V_{\text{NT},\eta}$ differs from the original M&S solution by the change of boundary conditions. Because of the optimality of the $\eta = 0$ boundaries, one immediately infers that $|\Theta_\eta - \Theta_0| = O(\eta^a)$ translates to $|V_{\text{NT},\eta}(\theta) - V_{\text{NT},0}(\theta)| = O(\eta^{2a})$.

We will now show that $a = 1/3$, i.e. that the shift of the band is of the same order as the width of the boundary layer. The continuity of the second derivative at $\theta = \Theta_\eta$ imposes, to leading order

$$-\eta^{1/3} F_y(0) = (V_{\text{NT},0})_{\theta\theta}(\Theta_\eta) \approx (V_{\text{NT},0})_{\theta\theta\theta}(\Theta_0)(\Theta_\eta - \Theta_0), \quad (37)$$

where we used the condition $(V_{\text{NT},0})_{\theta\theta}(\Theta_0) = 0$ derived in Appendix B. This immediately yields

$$\Theta_\eta = \Theta_0 - \frac{F_y(0)}{(V_{\text{NT},0})_{\theta\theta\theta}(\Theta_0)} \eta^{1/3}, \quad (38)$$

i.e. the inward shift of the band is on the order $\eta^{1/3}$ exactly as the width of the boundary layer. This is reasonable as it means that both effects of the quadratic term on the immediate proximity of the band are of the same order. As we show in Appendix B, the third derivative of $V_{\text{NT},0}$ turns out to be positive at the unperturbed band, hence the above expression implies that the boundary shifts inwards, i.e. the NT region is *reduced* by the presence of quadratic costs.

The alert reader might wonder how the solution given by Eq. (36) above still has a continuous first derivative at Θ_η . In fact, from the vanishing of the second derivative of the M&S solution at the unperturbed band, the Taylor expansion gives:

$$(V_{\text{NT},0})_\theta(\Theta_\eta) \approx (V_{\text{NT},0})_\theta(\Theta_0) + 0 + \frac{1}{2}(V_{\text{NT},0})_{\theta\theta\theta}(-\Theta_0)(\Theta_0 - \Theta_\eta)^2, \quad (39)$$

with $(V_{\text{NT},0})_\theta(\Theta_0) \equiv -\Gamma$. Now, since the HJB equation itself is independent of η in the NT region, its solution can only depend on η through boundary conditions. But because of the optimality of the position of the $\eta = 0$ boundary, we have everywhere inside the band

$$V_{\text{NT},\eta}(\theta) = V_{\text{NT},0}(\theta) + \frac{1}{2}(\Theta_0 - \Theta_\eta)^2 g(\theta) + \dots \quad (40)$$

Here $g(\theta)$ is the second derivative of the M&S solution with respect to the position of the boundary.

In view of (36) the continuity of the first derivative of the perturbed solution across the band is guaranteed by

$$-\Gamma \equiv (V_{\text{NT},\eta})_{\theta}(\Theta_{\eta}), \quad (41)$$

or, using the two previous equations,

$$\begin{aligned} -\Gamma &= (V_{\text{NT},0})_{\theta}(\Theta_{\eta}) + \frac{1}{2}(\Theta_0 - \Theta_{\eta})^2 g'(\Theta_0) + \dots \\ &= (V_{\text{NT},0})_{\theta}(\Theta_0) + \frac{1}{2}(\Theta_0 - \Theta_{\eta})^2 [(V_{\text{NT},0})_{\theta\theta\theta}(\Theta_0) + g'(\Theta_0)] + \dots \end{aligned}$$

However, remembering that $(V_{\text{NT},0})_{\theta}(\Theta_0) = -\Gamma$, one finally arrives at the following consistency condition

$$g'(\Theta_0) = -(V_{\text{NT},0})_{\theta\theta\theta}(\Theta_0). \quad (42)$$

As we show in Appendix B this is indeed a property of the M&S solution. This establishes that our boundary layer solution (36) is C^2 across the NT-RB boundary.

5 Discussion & extensions

All the results above are compatible with the numerical results of Ref. [10], which exhibit a square-root-like trading speed close to the band and a band that shrinks when η increases (*c.f.* Figure 1 therein; the parameters λ, ϵ are to be identified with our η, Γ , respectively). We have ourselves solved the HJB equation numerically and found a band position compatible with our prediction above $\Theta_0 - \Theta_{\eta} \propto \eta^{1/3}$.

In the above we did not elaborate on the domain of validity of the small η expansion. For perturbation theory to make sense, the shift in the position of the band must remain small compared to the width of the band itself, i.e.:

$$\left| \frac{\Theta_0^+ - \Theta_{\eta}^+}{\Theta_0^+ + \Theta_0} \right| \ll 1. \quad (43)$$

Assuming an Ornstein-Uhlenbeck process for the prediction³, $\mu(x) = -\Omega x$, we obtain in the small Γ limit

$$(V_{\text{NT},0})_{\theta\theta\theta}(\Theta_0) \approx \frac{8\lambda^2}{\sigma^2\Omega} \left(\frac{3\Gamma\sigma^2}{2\Omega} \right)^{1/3}. \quad (44)$$

³ Ω^{-1} is the mean-reversion time

This, together with the results of [7]

$$\Theta_0^+ + \Theta_0^- \sim \frac{\sigma^2}{\lambda} \left(\frac{\Gamma \Omega^2}{\sigma^4} \right)^{1/3}, \quad \Theta_0' \sim \frac{\Omega}{\lambda}, \quad (45)$$

where we have dropped all $O(1)$ numerical constants, allows us to recast the above as⁴

$$\frac{\eta}{\Gamma^{4/3}} \ll \frac{\lambda}{(\sigma \Omega)^{4/3}}, \quad (46)$$

so that the relevant combination is indeed $\eta \Gamma^{-4/3}$, as anticipated in [10] in the $\Gamma, \eta \rightarrow 0$ limit. In order to make sense of the above inequality, it is useful to substitute the price of risk λ with a risk target \mathcal{R} corresponding to the volatility of the ideal position. In the absence of costs this ideal position reads

$$\theta^* = \frac{\mu}{2\lambda}, \quad (47)$$

leading to a typical risk $\mathcal{R} \sim \sqrt{\Omega} \sigma^2 / 2\lambda$. The above inequality can thus be rewritten as

$$\frac{\eta}{\Gamma^{4/3}} \ll \frac{\sigma^{2/3}}{\mathcal{R} \Omega^{5/6}}. \quad (48)$$

Suppose that the target risk \mathcal{R} is a fraction φ of the daily volume V , i.e. $\mathcal{R} = \varphi V \sigma$. The quadratic cost parameter may be expressed using dimensionful quantities $\eta = \gamma \sigma T^{3/2} / V$, where γ is a number and $T = 1$ day. The final dimensionless condition is, interestingly enough, independent of the volume V

$$\gamma \varphi \ll \left(\frac{\Gamma}{\sigma \sqrt{T}} \right)^{4/3} (\Omega T)^{-5/6}. \quad (49)$$

Taking for example $\Gamma = 2$ bp, $\sigma \sqrt{T} = 2\%$ and $\Omega T = 0.1$, we find $\gamma \varphi \ll 10^{-2}$. In conclusion, our small η expansion makes sense, for a given quadratic cost coefficient γ , if the portfolio's typical positions represent a small fraction of the daily traded volume.

Let us conclude by pointing out that the above matched asymptotic expansion around the M&S solution can be undertaken for arbitrary non-linear costs. If instead of the quadratic cost term $\eta \dot{\theta}^2$ considered above one worked with the more realistic $\zeta \dot{\theta}^{3/2}$ term, one would find the following HJB equation in the trading region

$$\left(\frac{\partial V^\pm}{\partial \theta} \pm \Gamma \right)^3 \propto \zeta^2 [\lambda \theta^2 - \mu(x) \theta - \mathcal{L}_x[V^\pm]], \quad (50)$$

requiring a perturbative expansion of the form

$$V_\zeta^\pm(\theta, x) = V_0^\pm(\theta, x) + \zeta^{2/3} V_1^\pm(\theta, x) + \dots \quad (51)$$

⁴In the following discussion we assume that $\lambda \Theta_0 \ll \mu$, i.e. that the price of risk for a position at the edge of the band is small compared to the expected gains.

This leads to trading speed behaving as $(\theta - \Theta_\zeta)^{1/3}$ in the outer region close enough to the band, but crossing over to the following boundary layer solution

$$f(\theta) = \zeta^{4/5} G \left(\frac{\theta - \Theta_\zeta(x)}{\zeta^{2/5}} \right), \quad (52)$$

where $G(y)$ now obeys the following Abel equation (see e.g. [12])

$$G^3 - B'G_y = A'^2(y - y_0). \quad (53)$$

One should impose $G(y = 0) = 0$ and $G(y) \sim y^{1/3}$ for large y . Since our asymptotic expansion in the quadratic case relies chiefly on the properties of the M&S solution in the NT region, all the results obtained readily transpose to the present case as well.

Acknowledgements

We thank J. Donier, C. A. Lehalle, J. Muhle-Karbe and M. Potters for enlightening discussions on these subjects.

Appendix A: Continuity of the second derivative of the value function

In this appendix we wish to show that the rebalancing (RB) and no-trade (NT) value functions match smoothly at the boundary of the RB-NT regions up to and including second order derivatives. We will suppose that this boundary is smooth, that the value function is continuous and differentiable in each region. We rewrite the HJB equation as

$$\mathcal{L}_x[V] = -\eta v^2(x, \theta) + H(x, \theta),$$

with the velocity $v(x, \theta)$ given by

$$v(x, \theta) = \begin{cases} (\frac{\partial V}{\partial \theta} + \Gamma)/2\eta & \text{inside the RB zone,} \\ 0 & \text{inside the NT zone.} \end{cases}$$

We write the remainder of the right hand side as $H(x, \theta)$ as it is a regular function and does not play any role in matching the two regions.

Along the x direction $v(x, \theta)$ is at most discontinuous and, since \mathcal{L}_x is a second-order linear operator, this implies that $\frac{\partial V}{\partial x}$ is continuous at the boundary. As we suppose that V is continuous across the boundary, this implies that its derivatives across the boundary are continuous if we assume that the boundary is not collinear with x . Therefore $v(x, \theta)$ must be continuous. By the same token, the continuity of $v(x, \theta)$ implies continuity of $\frac{\partial^2 V}{\partial x^2}$.

Differentiating the HJB equation with respect to θ gives

$$\mathcal{L}_x[V_\theta] = -2\eta v_\theta(x, \theta)v(x, \theta) + H_\theta(x, \theta).$$

In the right hand side, $v_\theta(x, \theta)$ is discontinuous when crossing the boundary, but the product $v_\theta(x, \theta)v(x, \theta)$ remains continuous since $v(x, \theta)$ passes continuously through zero at the boundary. This implies that the mixed derivative $\frac{\partial^2 V}{\partial x \partial \theta}$ is continuous, otherwise the term $\frac{\partial^3 V}{\partial x^2 \partial \theta}$ in the left hand side would have to contain a delta function. Continuity of V , its first derivatives and that of $\partial_x \partial_\theta V$ guarantees continuity of $\partial_\theta^2 V$.

Appendix B: Some properties of the M&S solution

The M&S solution has a vanishing second derivative at the boundary

We would like to compute the second derivative of the M&S solution at the optimal boundary

$$\left. \frac{\partial^2 V_{\text{NT},0}}{\partial \theta^2} \right|_{x=\hat{h}_\pm}, \quad (54)$$

where \hat{h}_\pm is such that $\Theta_0^\pm(\hat{h}_\pm) = \theta$.

To set the notation we use the decomposition (9)

$$\partial_\theta V_{\text{NT},0} = I(x, \theta) + \alpha'_1(\theta)\psi_1(x) + \alpha'_2(\theta)\psi_2(x). \quad (55)$$

The first term represents the theta derivative of the particular solution

$$I(x, \theta) = \int_{-\infty}^{\infty} (\mu(\xi) - 2\lambda\theta)\mathcal{G}(x, \xi)d\xi. \quad (56)$$

The derivatives functions α'_1, α'_2 are given by (see [7])

$$D\alpha'_1(\theta) = -\Gamma\psi_2(h_-) - \Gamma\psi_2(h_+) + I(h_-, \theta)\psi_2(h_+) - I(h_+, \theta)\psi_2(h_-), \quad (57)$$

$$D\alpha'_2(\theta) = \Gamma\psi_1(h_-) + \Gamma\psi_1(h_+) + I(h_+, \theta)\psi_1(h_-) - I(h_-, \theta)\psi_1(h_+), \quad (58)$$

where the determinant $D \equiv D(h_+, h_-) = \psi_1(h_+)\psi_2(h_-) - \psi_1(h_-)\psi_2(h_+)$. The optimal boundaries are defined by the variational equality

$$\left. \frac{\partial \alpha'_{1,2}}{\partial h_\pm} \right|_{h_\pm=\hat{h}_\pm} = 0. \quad (59)$$

Consequently,

$$\left. \frac{d\alpha'_{1,2}}{d\theta} \right|_{h_\pm=\hat{h}_\pm} = \left. \frac{\partial \alpha'_{1,2}}{\partial \theta} \right|_{h_\pm=\hat{h}_\pm}. \quad (60)$$

We now have all the ingredients to compute the second derivative at the optimal boundaries

$$\partial_{\theta}^2 V_{\text{NT},0}(\hat{h}_+, \theta) = \partial_{\theta} I(\hat{h}_+, \theta) + \frac{d\alpha'_1}{d\theta} \Big|_{h_{\pm}=\hat{h}_{\pm}} \psi_1(\hat{h}_+) + \frac{d\alpha'_2}{d\theta} \Big|_{h_{\pm}=\hat{h}_{\pm}} \psi_2(\hat{h}_+). \quad (61)$$

Plugging the formulae (57)-(58) and using (60) we indeed find:

$$\partial_{\theta}^2 V_{\text{NT},0}(\hat{h}_+, \theta) = 0, \quad (62)$$

as expected from the continuity of the second derivative of the value function of the boundary. Remember that the M&S solution is a linear function of θ in the trading zone, therefore $\partial_{\theta}^2 V$ is identically zero there. A similar result may be found at $x = \hat{h}_-$.

A useful identity

Consider again the derivative of the M&S solution at the band (55). The derivative $\partial_{\theta} V_{\text{NT},0}$ must be equal to $\mp \Gamma$ at $\theta = \pm \Theta_0^{\pm}$. A change to any of the parameters of the problem will result in the boundaries being shifted by a $\delta\Theta$. This induces a small change of the functions $\alpha_{1,2}$ that we denote $\delta\alpha_{1,2}$. Since $I(x, \theta)$ does not depend explicitly on the boundary and is linear in θ , maintaining the correct boundary condition to second order in $\delta\Theta$ leads to the following identity

$$\delta\alpha'_1 \psi_1(x) + \delta\alpha'_2 \psi_2(x) + \frac{1}{2} (\delta\Theta)^2 [\alpha'''_1 \psi_1(x) + \alpha'''_2 \psi_2(x)] = 0, \quad (63)$$

where all quantities are evaluated at one of the boundaries. This is equivalent to (42).

Sign of the third derivative

Our strategy will be to take the derivative of the boundary conditions

$$\mp \Gamma = I(x, \pm \Theta_0^{\pm}) + \alpha'_1(\pm \Theta_0^{\pm}) \psi_1(x) + \alpha'_2(\pm \Theta_0^{\pm}) \psi_2(x), \quad (64)$$

$$0 = I'(x, \pm \Theta_0^{\pm}) + \alpha''_1(\pm \Theta_0^{\pm}) \psi_1(x) + \alpha''_2(\pm \Theta_0^{\pm}) \psi_2(x), \quad (65)$$

with respect to the linear cost parameter Γ . We will present the argument for the upper boundary. Introducing

$$\rho_x = \frac{\partial \Theta_0^+}{\partial x} < 0, \quad \rho_{\Gamma} = \frac{\partial \Theta_0^+}{\partial \Gamma} > 0, \quad (66)$$

the derivative of the first equation may be written as

$$\begin{aligned} 1 &= \rho_{\Gamma} [I'(x, \Theta_0^+) + \alpha''_1(\Theta_0^+) \psi_1(x) + \alpha''_2(\Theta_0^+) \psi_2(x)] + \\ &+ \frac{\partial \alpha'_1}{\partial \Gamma}(\Theta_0^+) \psi_1(x) + \frac{\partial \alpha'_2}{\partial \Gamma}(\Theta_0^+) \psi_2(x). \end{aligned} \quad (67)$$

Note that because of the second boundary condition the term inside the brackets vanishes. Taking the derivative of the resulting equation wrt x now leads to

$$0 = \rho_x \left[\frac{\partial \alpha_1''}{\partial \Gamma}(\Theta_0^+) \psi_1(x) + \frac{\partial \alpha_2''}{\partial \Gamma}(\Theta_0^+) \psi_2(x) \right] + \quad (68)$$

$$+ \frac{\partial \alpha_1'}{\partial \Gamma}(\Theta_0^+) \psi_1'(x) + \frac{\partial \alpha_2'}{\partial \Gamma}(\Theta_0^+) \psi_2'(x) \equiv J_1 + J_2. \quad (69)$$

One can use the explicit solutions for $\alpha_{1,2}$ in (57)-(58) to compute the partial derivatives with respect to Γ by observing that the implicit dependence through the boundaries cancels out because of the M&S optimality condition. We have

$$\frac{\partial \alpha_1'}{\partial \Gamma} = -\frac{1}{D} (\psi_2(h_-) + \psi_2(h_+)), \quad \frac{\partial \alpha_2'}{\partial \Gamma} = \frac{1}{D} (\psi_1(h_-) + \psi_1(h_+)), \quad (70)$$

so that the second term J_2 in the last equation above reads

$$J_2 = -\frac{1}{D} (\psi_2(h_+) \psi_1'(h_+) - \psi_2'(h_+) \psi_1(h_+) + \quad (71)$$

$$\psi_2(h_-) \psi_1'(h_+) - \psi_1(h_-) \psi_2'(h_+)). \quad (72)$$

Interestingly enough, the first two terms in the parenthesis is the Wronskian which is positive, *c.f.* [7]. The last two terms also give a positive contribution since they can be written as $\partial_{h_+} D$ which is positive when $h_+ > h_-$. Consequently, $J_2 < 0$ and therefore $J_1 > 0$. Now, take the derivative of the second boundary condition wrt Γ

$$0 = \rho_\Gamma [\alpha_1'''(\Theta_0^+) \psi_1(x) + \alpha_2'''(\Theta_0^+) \psi_2(x)] + \frac{\partial \alpha_1''}{\partial \Gamma}(\Theta_0^+) \psi_1(x) + \frac{\partial \alpha_2''}{\partial \Gamma}(\Theta_0^+) \psi_2(x). \quad (73)$$

One immediately infers that

$$\rho_x \rho_\Gamma [\alpha_1'''(\Theta_0^+) \psi_1(x) + \alpha_2'''(\Theta_0^+) \psi_2(x)] = -J_1, \quad (74)$$

which in view of $\rho_x \rho_\Gamma < 0$, leads one to conclude

$$V_{\theta\theta\theta}(\Theta_0^+) \equiv \alpha_1'''(\Theta_0^+) \psi_1(x) + \alpha_2'''(\Theta_0^+) \psi_2(x) > 0. \quad (75)$$

References

- [1] B. Tóth, Y. Lempérière, C. Deremble, J. De Lataillade, J. Kockelkoren, and J.-Ph. Bouchaud, *Anomalous price impact and the critical nature of liquidity in financial markets*, Physical Review X, 1:021006 (2011)
- [2] J. Donier, J. Friedrich Bonart, I. Mastromatteo, and J.-Ph. Bouchaud. *A fully consistent, minimal model for non-linear market impact*, Quantitative Finance, 15(7):1109-1121 (2015), and refs. therein.

- [3] J. Donier, J. Bonart, *A million metaorder analysis of market impact on Bitcoin*, [arXiv:1412.4503](#)
- [4] R. F. Almgren and N. Chriss, *Optimal execution of portfolio transactions*, *Journal of Risk*, 3, 5-39 (2000)
- [5] A. Obizhaeva, J. Wang, *Optimal trading strategy and supply/demand dynamics*, *Journal of Financial Markets*, 16, 1-32 (2013).
- [6] N. B. Gârleanu, L. H. Pedersen, *Dynamic Trading with Predictable Returns and Transaction Costs*, *Journal of Finance*, vol. 68 (2013), issue 6, p. 2309-2340
- [7] R. Martin, T. Schöneborn, *Mean Reversion Pays, but Costs*, *RISK*, February 2011, p. 96-101
- [8] J. de Lataillade, C. Deremble, M. Potters, J.-Ph. Bouchaud, *Optimal Trading with Linear Costs*, *Journal of Investment Strategy* 01/2012, 1(3): p. 91-115
- [9] F. Passerini, S. Vázquez, *Optimal Trading with Alpha Predictors*, [arXiv:1501.03756](#)
- [10] R. Liu, J. Muhle-Karbe, M. Weber, *Rebalancing with Linear and Quadratic Costs*, [arXiv:1402.5306](#)
- [11] J. Hinch, *Perturbation Methods*, Cambridge University Press, 1991.
- [12] D. E. Panayotounakos, Th. I. Zampoutis, *Construction of Exact Parametric or Closed Form Solutions of Some Unsolvable Classes of Non-linear ODEs*, *International Journal of Mathematics and Mathematical Sciences* Volume 2011 (2011)