Portfolio optimization with transient linear and non-linear price impact under exponential utility

Martin Forde* Marko Weber† Hongzhong Zhang‡§

January 30, 2016

Abstract

We consider a market with one risky asset and one riskless asset with temporary linear price impact where the price paid per share is given by $\tilde{S}_t = S_t(1+\lambda S_t\dot{\theta}_t)$, where S_t is the unaffected price which evolves as $dS_t = S_t(\mu dt+\sigma dW_t)$ and θ_t is the number of shares held at time t. In this setting, we compute the optimal trading strategy and implied welfare for an investor with exponential (CARA) utility and long time horizon, and we describe the asymptotic behaviour of these quantities as $\lambda \to 0$. Letting $u_t = S_t\dot{\theta}_t$, the optimal strategy $\hat{u}_t$ is characterized via the solution to a non-linear parabolic PDE and we find that $\hat{u}_t = \hat{u}_t(Y_t) = -\frac{\sqrt{\alpha}\,\sigma(Y_t-\bar{Y})}{\sqrt{2}}\frac{1}{\sqrt{\lambda}} + o(\frac{1}{\sqrt{\lambda}})$ where Y_t is the risky wealth, i.e. for $\lambda \ll 1$ a myopic strategy is optimal where the agent trades towards the frictionless target $\bar{Y} = \frac{\mu}{\alpha\sigma^2}$, and we find that $\hat{u}(Y_t) \approx -\frac{1}{2\lambda}$ for $Y_t \gg 1$, i.e. we sell at constant speed. The $Q(y)$ function in the exponent of the ansatz for the value function is unbounded (unlike the power utility case in [GW15]) so the verification theorem is more difficult to establish. We later extend this setup to the case of non-linear price impact where $\tilde{S}_t = S_t(1+\lambda|S_t\dot{\theta}_t|^\gamma \mathrm{sgn}(\dot{\theta}))d\theta_t$ for $\gamma \in (0,1)$. In this case, the solution is governed by a non-linear ODE which exhibits a delicate bifurication-type behaviour as the β parameter approaches a critical value β^* which measures the welfare loss due to illiquidity. We also discuss the limit as $\gamma \to 0$ which corresponds to proportional transaction costs.

1 Introduction

The impact of trades on execution prices is a key component in determining optimal rebalancing strategies. Classical models assume prices are unaffected by trades, and conclude that a constant proportion of risky wealth is optimal (for power utility) and constant risky wealth is optimal for exponential utility (cf. Merton[Mer69]), which results in infinite trading volume. Small bid-ask spreads with proportional transaction costs preclude trading when the portfolio is near its target inside the so-called no trade region (see Constantinides[Con86], Davis&Norman[DN90] et al.). Price impact models are in a sense a continuous time idealization of a limit order book structure where traders are penalized with a higher buy price or lower sell price if they wish to fill orders faster in large quantities. In contrast to the aforementioned models, linear price-impact models recommend a trading rate which is proportional to its distance from the frictionless target (see [GW15]), although empirical evidence would suggest that price impact is nonlinear.

[GW14] consider a non-linear price impact model for a financial market with one risky asset and a risk free bond with zero interest rate where the price paid per share is given by $\tilde{S}_t = S_t(1+\lambda|\frac{S_t}{X_t}\dot{\theta}_t|^\alpha \mathrm{sgn}(\dot{\theta}))d\theta_t$, where S_t is the best quote that can be attained which satisfies $dS_t = S_t(\mu dt + \sigma dW_t)$, θ_t is the number of shares held at time t and X_t is the total investor wealth. This generalizes their linear price impact model in [GW15], and by linear we mean the case when $\alpha = 1$. We remark that these are *transient* price impact models, in the sense that the effect of illiquidity is a incurred cost, but this illiquidity does not affect the price of the underlying asset, but rather just the price that the agent pays for asset while his trading speed is non-zero. In this framework, [GW14] compute the optimal trading strategy, its implied welfare, and the resulting trading volume for long-term investors with power utility. The optimal quantities are determined in terms of a solution to a non-linear ordinary differential equation, and these quantities admit explicit asymptotic formulas, which depend on a universal constant that depends only on α, which can be computed numerically. Trading rates are finite as with linear impact (unlike transaction costs), and are lower near the frictionless target and higher away from this

*Dept. Mathematics, King's College London, Strand, London, WC2R 2LS (Martin.Forde@kcl.ac.uk)
†Department of Mathematical Sciences, Dublin City University, and Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa Italia
‡Dept. Statistics, Columbia University, New York, NY 10027 (hzhang@stat.columbia.edu)
§We thank Paolo Guasoni for many invaluable insights.

target, and the model includes the square-root impact law and, as extreme cases, linear impact ($\alpha = 1$) and proportional transaction costs (the limit as $\alpha \to 0$). The economic intuition/motivation for such a model is that the faster an agent wants to buy (or sell) the stock, the deeper into the limit order book he will have to go, and the higher (lower) will be the price for the units of the asset which are bought (sold) at a later time. However, once a transaction is completed, the limit order book quickly fills up again, and that the transaction has no permanent lasting effect on the price of the stock. Compare this to an Almgren-Chriss (AC) type model (cf. [AC01],[Alm03]) which captures both temporary and permanent price impact, where $\tilde{S}_t = S_t + \eta \dot{x}_t + \gamma(x_t - x_0)$, where x_t is the number of shares held at time t and S_t is typically modelled as geometric or arithmetic Brownian motion. Non-linear generalizations of this type of model (if mispecified) can lead to price manipulation or more specifically the free "round trip" phenomenon, whereby a trader can start and end with zero risky position over a fixed time interval and have a negative expected trading cost (e.g. a dominant market player can rapidly sell stock, forcing the price down and then buy it all back, with a positive expected Profit/Loss), see e.g. Gatheral&Schied[GS13] for precise results in this direction.

Motivated by modelling of limit order books, Rogers&Singh[RS10] consider the problem of optimal hedging of a European contingent claim under a Black-Scholes model with transient linear price impact. The objective functional in [RS10] is the square of the difference between the option price and the hedging strategy, plus the accrued costs from the effect of the price impact, which can be re-written in a more usable form using the Ito isometry. This problem leads to a HJB equation, for which they substitute an ansatz which is quadratic in the current number of shares and leads to three parabolic PDEs, one of which is non-linear. For the latter, they construct an approximating sequence using an implicit equation for the solution which arises from the Feynmac-Kac formula. The other two equations are then straightforward to solve in terms of this solution. The numerical solution requires solving three PDEs of Black-Scholes type, and they compute the leading order asymptotics when the price impact parameter is small.

As in this article, Schied et al.[SST10] consider an investor with CARA utility, but they look at the *optimal liquidation* problem where the investor has to close out his position in the risky asset by a fixed terminal time. They find the surprising result that the investor fares no worse if he restricts his trading strategies to those which are deterministic in time (so there is no intertemporal updating of strategies); in particular, this means the the (optimal) terminal revenue is normally distributed, so the optimal liquidation exponential utility maximization problem reduces to the problem of finding a deterministic strategy which minimizes the usual mean-variance functional (where the risk aversion parameter α is determined by the exponent of the utility function), and we can compute the expected exponential utility of the revenue just using the known mgf of a normal distribution.

[SS08] consider the problem of maximizing expected utility from liquidating a risky asset when the time horizon is infinite, for a von-Neumann-Morgenstern investor in a Bachelier model with temporary and permanent linear price impact. They consider a general utility function, and rather than seeking closed-form solutions for the optimal trading strategies, they employ a stochastic control approach and show that the value function and associated optimal control satisfy certain nonlinear parabolic PDEs. They show that investors with increasing absolute risk aversion (IARA) should sell faster when the risky asset price rises than when it falls, and investors with decreasing absolute risk aversion (DARA) should sell slower when asset prices rise. Their PDE formulation also facilitates a qualitative sensitivity analysis of the optimal strategy and the value function.

In this article we work in a similar setting to [GW14] but we now assume that $\tilde{S}_t = S_t(1 + \lambda S_t \dot{\theta}_t)$ and we consider an investor who maximizes his equivalent safe rate under exponential utility in the large-time limit, i.e. trade so as to maximize $\liminf_{T\to\infty} -\frac{1}{\alpha T} \log \mathbb{E}(e^{-\alpha X_T})$. We show that compute the optimal dynamic trading strategy $u_t = \dot{\theta}_t S_t$ and the implied welfare, and compute asymptotic expressions for these quantities as $\lambda \to 0$. In particular, we find that

$$u_t \quad = \quad \hat{u}(Y_t) \quad = \quad -\frac{\sqrt{\alpha}\,\sigma(Y_t - \bar{Y})}{\sqrt{2}} \frac{1}{\sqrt{\lambda}} + o(\frac{1}{\sqrt{\lambda}})$$

as $\lambda \to 0$, where Y is the risky wealth, $\bar{Y} = \frac{\mu}{\alpha\sigma^2}$ is the frictionless optimizer, i.e. the optimal risky wealth for the frictionless case, which is constant (see section 2 for a derivation); thus at leading order, we see that we sell when Y is above $\bar{Y}$, and buy when Y is below $\bar{Y}$. In section 6, we perform similar (formal) computations for the case of non-linear price impact where $S_t(1 + \lambda |S_t \dot{\theta}_t|^\gamma \mathrm{sgn}(\dot{\theta}))d\theta_t$ (note that our γ parameter is the α paramater in the [GW15] article discussed above), and for $\gamma \in (0, 1)$, we find that the asymptotic behaviour of the optimal trading strategy is essentially governed by a non-linear ODE and a universal constant as for the power utility case discussed in [GW14].

2 Review of the frictionless case - the Merton problem with finite time horizon and exponential utility

We work on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ throughout, supporting a single Brownian motion and satisfying the usual conditions. We consider a financial market with a standard safe asset earning zero interest rate and a risky asset whose best quoted price S_t follows a geometric Brownian motion:

$$dS_t \quad = \quad S_t(\mu dt + \sigma dW_t)$$

with $S_0 > 0$, where W is standard Brownian motion. Let X_t denote the investor's total wealth at time t, θ_t the amount of stocks owned, and $Y_t := \theta_t S_t$ denote his risky position, and assume that $X_0 > 0$. Then X_t evolves as

$$dX_t \quad = \quad Y_t \frac{dS_t}{S_t} \quad = \quad Y_t(\mu dt + \sigma dW_t)\,. \tag{1}$$

Let $\mathcal{A}$ denote the set of progressively measurable processes Y such that $\int_0^T Y_t^2 dt < \infty$ a.s. The Hamilton-Jacobi-Bellman (HJB) equation for the value function $V(t,x) = \sup_{Y\in\mathcal{A}} \mathbb{E}_{x,t}(U(X_T))$ with exponential utility $U(x) = -e^{-\alpha x}$ is

$$V_t \quad = \quad \inf_y [-y\mu V_x \, - \, \frac{1}{2} y^2 \sigma^2 V_{xx}]$$

with terminal condition $V(T,x) = -e^{-\alpha x}$. Solving for y we have

$$y^* \quad = \quad -\frac{\mu V_x}{\sigma^2 V_{xx}}\,.$$

Using the ansatz $V(x,t) = -e^{-\alpha x} e^{-\alpha\beta(T-t)}$, we find that

$$y^* \quad = \quad \frac{\mu}{\alpha\sigma^2} \quad , \quad \beta \quad = \quad \frac{\mu^2}{2\alpha\sigma^2}\,.$$

β is the so-called *equivalent safe rate*, because an annuity which is worth $A_t = \beta t$ at time t has the same exponential utility, i.e. $e^{-\alpha\beta t}$ (ignoring the constant in front of the exponential). Thus the optimal risky wealth is equal to the constant value

$$\bar{Y} \quad = \quad \frac{\mu}{\alpha\sigma^2}\,.$$

From (3), we see that the total wealth evolves as

$$dX_t \quad = \quad \bar{Y}(\mu dt + \sigma dW_t)$$

i.e. arithmetic Brownian motion, and the optimal number of shares θ_t evolves as

$$\theta_t \quad = \quad \frac{\bar{Y}}{S_t}$$

which clearly is not differentiable in t a.s. (this analysis can be made rigorous using verification theorems, see e.g. section 3.6.1 and Theorem 3.5.2 in [Pham09].

3 Linear price impact

We now incorporate linear price impact into the model above; we assume that trades in the risky asset are no longer realized at the best quote S_t but rather at a less favorable price $\tilde{S}_t$, which effectively penalizes the trader for making large trades in a short period of time. Specifically, we assume that the price paid per share is given by

$$\tilde{S}_t \quad := \quad S_t(1 + \lambda S_t \dot{\theta}_t) \tag{2}$$

where θ_t is the number of shares held at time t, which we assume is differentiable in t (in contrast to the frictionless case discussed above). Let C_t denote the cash position, which (from the self-financing condition) must evolve as

$$dC_t \quad = \quad -S_t(1 + \lambda S_t \dot{\theta}_t) d\theta_t \quad = \quad -S_t \dot{\theta}_t dt - \lambda S_t^2 \dot{\theta}_t^2 dt\,.$$

We now let $u_t = \dot{\theta}_t S_t$ denote the wealth turnover, $X_t = \theta_t S_t + C_t$ denote the total wealth, and $Y_t = \theta_t S_t$ the risky wealth. Then from Ito's rule we have

$$\begin{aligned} dX_t &= S_t d\theta_t + \theta_t dS_t + dC_t &&= \theta_t S_t(\mu dt + \sigma dW_t) - \lambda S_t^2 \dot{\theta}_t^2 dt \\ &&&= Y_t(\mu dt + \sigma dW_t) - \lambda u_t^2 dt\,. \end{aligned}$$

and

$$dY_t = d(\theta_t S_t) = \theta_t dS_t + S_t \dot{\theta} dt$$

which we can re-write as

$$dY_t = Y_t(\mu dt + \sigma dW_t) + u_t dt\,. \tag{3}$$

Moreover, we see that

$$\theta_t = \int_0^t \dot{\theta}_s ds = \int_0^t \frac{u_s}{S_s} ds\,.$$

Following Definition 1 in [GW15], we define an admissible trading strategy in terms of its u_t process, and require that the resulting wealth process remain positive:

Definition 3.1 *For initial wealth $X_0 > 0$ and initial risky wealth Y_0, an admissible strategy is a process u_t, which is bounded and adapted to the natural filtration of S_t such that $\int_0^T |u_t| dt < \infty$ a.s. for all T), and such that the stochastic differential equation* (3) *for Y has a unique strong solution on for $t \in [0, \infty)$ such that $Y_t \geq 0$ for all t, i.e. no short selling allowed. For any such admissible strategy, from* (3) *we see that the corresponding wealth process X satisfies*

$$X_t = X_0 \exp[\int_0^T ((\mu - \frac{1}{2}\sigma^2 - \lambda u_t^2) dt + \sigma dW_t)]\,.$$

We assume the investor has exponential utility function $U(x) = -e^{-\alpha x}$ and trades so as to maximize his *equivalent safe rate*, which we now define.

Definition 3.2 *The equivalent safe rate is defined as*

$$ESR_\alpha := \liminf_{T \to \infty} -\frac{1}{\alpha T} \log \mathbb{E}(e^{-\alpha X_T})\,.$$

This corresponds to the linear rate that gives the same long-run utility as the trading strategy of the investor:

$$\mathbb{E}(U(X_T)) \approx \mathbb{E}(U(ESR_\alpha \cdot T))).$$

Note that this definition differs from the one presented for power utility in [GW15], where

$$\mathbb{E}(U(X_T)) \approx \mathbb{E}(U(\exp\{ESR \cdot T\}))).$$

4 Heuristics

This section contains a sketch derivation of the main result, based on heuristic control arguments, while the rigorous proof is in section 5.

Suppose that the investor maximizes $\mathbb{E}(-e^{-\alpha X_T})$ for some finite time horizon T. For a trading strategy $u_t = \dot{\theta}_t S_t$, (assuming sufficient smoothness) the value function $V(t, X_t, Y_t)$ evolves as

$$\begin{aligned} dV(t, X_t, Y_t) &= V_t dt + V_x dX_t + V_y dY_t + \frac{1}{2} V_{xx} d\langle X \rangle_t + \frac{1}{2} V_{yy} d\langle Y \rangle_t + V_{xy} d\langle X, Y \rangle_t \\ &= V_t dt + V_x(Y_t(\mu dt + \sigma dW_t) - \lambda u_t^2 dt) + V_y(Y_t(\mu dt + \sigma dW_t) + u_t dt) \\ &+ \frac{1}{2} V_{xx} Y_t^2 \sigma^2 + \frac{1}{2} V_{yy} Y_t^2 \sigma^2 + V_{xy} Y_t^2 \sigma^2 dt\,. \end{aligned}$$

By the martingale optimality principle of stochastic control, the value function $V(t, X_t, Y_t)$ must be a supermartingale for any admissible strategy, and a martingale for the optimal one, i.e. the drift of $V(t, X_t, Y_t)$ cannot be positive is zero for the optimizer. The corresponding HJB equation for $V(t, x, y)$ is thus obtained by maximizing the drift over the strategy u, and setting it equal to zero, which yields

$$\begin{aligned} 0 &= H := \sup_u [V_t + V_x(y\mu - \lambda u^2) + V_y(y\mu + u) + \frac{1}{2}V_{xx}y^2\sigma^2 + \frac{1}{2}V_{yy}y^2\sigma^2 + V_{xy}y^2\sigma^2] \\ &= V_t + \mu y V_x + \mu y V_y + \frac{1}{2}y^2\sigma^2 V_{xx} + \frac{1}{2}y^2\sigma^2 V_{yy} + y^2\sigma^2 V_{xy} + \sup_u [-\lambda u^2 V_x + V_y(y\mu + u)]\,. \end{aligned} \tag{4}$$

Substituting

$$V(t, x, y) = -e^{-\alpha x} e^{\alpha\beta t} e^{\alpha \int_0^y q(\zeta)d\zeta} \tag{5}$$

and solving for $\frac{\partial H}{\partial u} = 0$ yields that

$$\hat{u}(y) = -\frac{q(y)}{2\lambda}$$

and $\frac{\partial^2 H}{\partial u^2} = 2\alpha\lambda V(t, x, y) < 0$ (and independent of u), so $\hat{u}$ is indeed the global maximizer. Substituting $u = \hat{u}$, the HJB equation reduces to the ordinary differential equation:

$$\frac{dq}{dy} = f(y, q) \tag{6}$$

where

$$f(y, q) = \frac{1}{\sigma^2 y^2}\big[(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q(y)^2 - 2y(\mu - \alpha\sigma^2 y)q(y) + (2\mu y - \alpha\sigma^2 y^2 - 2\beta)\big]. \tag{7}$$

Remark 4.1 Note that Eq (6) has no boundary conditions, and the correct value for β is an unknown of the problem (these issues will be dealt with in section 5).

An admissible solution to (6) should be bounded, and (in analogy with the frictionless case) be such that $u(0+) > 0$ and $u(+\infty) < 0$.

On a compact interval $[0, \tilde{y}]$ that contains $\bar{Y} = \mu/(\alpha\sigma^2)$, we study the behavior of the solution for small λ. Following [GW15], we guess a solution of the form $q(y) = q_1(y)\sqrt{\lambda} + o(\sqrt{\lambda})$ and $\beta(\lambda) = \beta^* - c(\lambda)$ where $\beta^* = \frac{\mu^2}{2\alpha\sigma^2}$ (the equivalent safe rate for the frictionless case) and $c(\lambda)$ represents the welfare loss from illiquidity (recall that the utility function and $V(t, x, y)$ are negative so we expect $c(\lambda)$ to be positive), and we guess that $c(\lambda) = \tilde{c}_1\sqrt{\lambda} + o(\sqrt{\lambda})$. Then we find that

$$q_1(y) = \sqrt{2\alpha}\,\sigma(y - \bar{Y}) \tag{8}$$

where $\bar{Y}$ is the *frictionless target*. This implies that

$$\hat{u}(y) = -\frac{\sqrt{\alpha}\,\sigma(y - \bar{Y})}{\sqrt{2}}\frac{1}{\sqrt{\lambda}} + o(\frac{1}{\sqrt{\lambda}})\,. \tag{9}$$

Thus at leading order, we see that we sell when Y is above $\bar{Y}$, and buy when Y is below $\bar{Y}$ and (formally at least) we see that for small λ, Y evolves as

$$dY_t = Y_t(\mu dt + \sigma dW_t) + u_t dt \approx -\frac{\sqrt{\alpha}\,\sigma(Y_t - \bar{Y})}{\sqrt{2}}\frac{1}{\sqrt{\lambda}}dt + Y_t\sigma dW_t$$

for $\lambda \ll 1$, so we expect Y to have mean-reverting behaviour around $\bar{Y}$ (this kind of result is made rigorous via weak convergence arguments in [GW15]). We expect $\hat{u}(y)$ to blow up as $\lambda \to 0$, because in the frictionless case $\theta_t = \frac{\bar{Y}}{S_t}$ which clearly is not differentiable in t almost surely. Proceeding as in [GW15], we now evaluate the HJB equation at $y = \bar{Y}$, divide both sides by $\sqrt{\lambda}$ and take again the limit for $\lambda \downarrow 0$ to get

$$\sqrt{2\alpha}\,\sigma = \frac{2\tilde{c}_1\alpha^2\sigma^2}{\mu^2}$$

which we can re-arrange as

$$\tilde{c}_1 = \frac{\mu^2}{\sqrt{2}\,\alpha^{\frac{3}{2}}\sigma}\,. \tag{10}$$

Note that if we change the problem to maximize exponential utility of riskless wealth $X_t - Y_t$, then trivially we find that $u_t^* = -\frac{1}{2\lambda}$.

5 Verification arguments

We now give the verification lemmas which make the arguments in section 4 rigorous.

Proposition 5.1 *For $\lambda > 0$ sufficiently small, there exists a β and a solution to the ODE $q(y)$ in (6) defined on $(0, +\infty)$, which is bounded from below, increasing, and satisfies $q(0+) = -2\sqrt{\lambda\beta^*}$ and $q(\infty) = 1$.*

Proof. See Appendix A. ■

Remark 5.1 Using that $q(\infty) = 1$, we see that the optimal trading strategy $\hat{u}(Y_t) \approx -\frac{1}{2\lambda}$ for $Y_t \gg 1$ i.e. we sell at constant speed.

We will require the following intermediate lemma.

Lemma 5.2 *Let*

$$F(y) := \bar{Y} - y(1 - q(y)), \quad \forall y > 0, \tag{11}$$

$$c_\pm(\lambda, y) := \frac{1}{2}\big[-1 \pm \sqrt{1 + 4\alpha(\frac{q^2(y)}{2\lambda\sigma^2} + \frac{2(\beta^* - \beta)}{\sigma^2} - \bar{Y})}\big], \quad \forall y > 0, \tag{12}$$

where we assume that $\lambda > 0$ is sufficiently small such that $c_+(\lambda, \infty) > 0$ and $-c_-(\lambda, \infty)/\alpha > \bar{Y}$. Then

$$-\sqrt{\frac{2(\beta^* - \beta + 1/(4\lambda))}{\alpha\sigma^2}} \leq \liminf_{y\to\infty} F(y) \leq \limsup_{y\to\infty} F(y) \leq 0. \tag{13}$$

Proof. See Appendix B. ■

Lemma 5.3 *There exists a probability measure $\hat{P}$, equivalent to P, such that the terminal wealth of X_T of any admissible strategy satisfies*

$$-\mathbb{E}(e^{-\alpha(X_T - X_0)}) \leq -e^{-\alpha\beta T}\,\mathbb{E}^{\hat{P}}(e^{-\alpha(Q(Y_T) - Q(Y_0))})$$

where $Q(y) = \int_0^y q(\zeta)d\zeta$ and $q(.)$ is defined as in Proposition 5.1, and equality holds for $u_t = \hat{u}(Y_t)$.

Proof. Recall that

$$\begin{aligned} dX_t &= Y_t(\mu dt + \sigma dW_t) - \lambda u_t^2 dt\,, \\ dY_t &= Y_t(\mu dt + \sigma dW_t) + u_t dt \end{aligned}$$

and our ansatz $V(t, x, y) = -e^{-\alpha x}e^{\alpha\beta t}e^{\alpha Q(y)}$, and let $\mathcal{L}$ and $\mathcal{A}$ denote the infinitesimal generators of X and Y respectively. Then for any admissible control u_t we have

$$\begin{aligned} -Z_T := V(t, X_t, Y_t) &= -e^{-\alpha(X_T - X_0) + \alpha\beta t + \alpha(Q(Y_t) - Q(Y_0))} \\ &= -e^{\int_0^T \alpha[-\mathcal{L}x(X_t)dt - \alpha\sigma Y_t dW_t + \beta dt + \mathcal{A}Q(Y_t)dt + \alpha\sigma Y_t q(Y_t)dW_t]} \\ &= -e^{\int_0^T \alpha(-\mathcal{L}x(X_s) + \beta + \mathcal{A}Q(Y_s))ds + \int_0^t \gamma_s dW_t} \\ &= -e^{\int_0^T \alpha(-\mathcal{L}x(X_t) + \beta + \mathcal{A}Q(Y_t))ds + \frac{1}{2}\int_0^t \gamma_s^2 ds}\,\frac{d\hat{P}}{dP} \end{aligned} \tag{14}$$

where $\gamma_t = -\alpha Y_t\sigma(1 - q(Y_t))$, and $\frac{d\hat{P}}{dP}|_{\mathcal{F}_t} = e^{\int_0^t \gamma_s dW_s - \frac{1}{2}\int_0^s \gamma_s^2 ds}$. $\frac{d\hat{P}}{dP}|_{\mathcal{F}_t}$ is the stochastic exponential of γ_t, but we know that γ_t is uniformly bounded from (13), so we can use the Novikov condition to conclude that $\frac{d\hat{P}}{dP}|_{\mathcal{F}_t}$ is a true martingale.

From section 4 we know that $V(t, X_t, Y_t)$ has non-positive drift for any choice of u (note that we are not using the martingale optimality principle to justify this statement because we don't know the form of the true value function $\tilde{V}(t, x, y)$, rather it just follows because we maximize H in (4) over all u values in the lines that immediately follow (4), but we know that H is zero when q is given by the solution $q(y)$ defined in Proposition 5.1, and from Proposition 5.1 we also know that $q(Y_t)$ is bounded, and hence $\hat{u}_t = -\frac{q(Y_t)}{2\lambda}$ is also bounded, consistent with our assumption for an admissible u_t in Definition 3.1. Hence the drift of the right hand side of (14) is non-positive, and from Ito's formula we know this drift is given explicitly by

$$-Z_t[\alpha(-\mathcal{L}x(X_t) + \beta + \mathcal{A}Q(Y_t)) + \frac{1}{2}\gamma_t^2] \leq 0\,.$$

Thus we have

$$-\alpha(X_T - X_0) + \alpha\beta T + \alpha(Q(Y_T) - Q(Y_0)) \quad \geq \quad \log \frac{d\hat{P}}{dP}$$

and the result follows. Finally (from Proposition 5.1) we know that the drift is zero when $u_t = \hat{u}(Y_t)$. ■

Remark 5.2 Under $\hat{P}$ the dynamics of Y are

$$dY_t \quad = \quad Y_t[(\mu - \alpha Y_t \sigma^2(1 - q(Y_t))dt + \sigma d\hat{W}_t)] \; + \; u_t dt$$

where $\hat{W}$ is a $\hat{P}$-Brownian motion.

We now have the following corollary of Lemma 5.3:

Corollary 5.4 *For any admissible control we have*

$$\liminf_{T\to\infty} -\frac{1}{\alpha T} \log \mathbb{E}(e^{-\alpha(X_T - X_0)}) \quad \leq \quad \beta \; - \; \liminf_{T\to\infty} \frac{1}{\alpha T} \log \mathbb{E}^{\hat{P}}(e^{-\alpha(Q(Y_T) - Q(Y_0))}) \tag{15}$$

with equality for $u_t = \hat{u}(Y_t)$.

Remark 5.3 In the proof of Proposition 5.1 we showed that $c(\lambda) := \frac{\mu^2}{2\alpha\sigma^2} - \beta$ converges to 0 for small λ. In addition, $f(y, 0) > 0$ is equivalent to $y \in (\bar{Y} - \sqrt{\frac{2c(\lambda)}{\alpha\sigma^2}}, \bar{Y} + \sqrt{\frac{2c(\lambda)}{\alpha\sigma^2}})$. Since $q(y)$ is increasing, the value y_* such that $q(y_*) = 0$ belongs to the interval $(\bar{Y} - \sqrt{\frac{2c(\lambda)}{\alpha\sigma^2}}, \bar{Y} + \sqrt{\frac{2c(\lambda)}{\alpha\sigma^2}})$. Therefore, for any $\varepsilon > 0$, if λ is chosen small enough, $y_* < \bar{Y} + \sqrt{\frac{2c(\lambda)}{\alpha\sigma^2}}) < \bar{Y} + \varepsilon$. Thus, the integral $Q(y) = \int_0^y q(\zeta)d\zeta$ is bounded from below by $(\bar{Y} + \varepsilon) \times q(0)$.

We now give the main result of the article.

Theorem 5.5 *For any bounded control u_t such that $Y_t \geq 0$ for all $t > 0$, we have*

$$-\liminf_{T\to\infty} \frac{1}{\alpha T} \log \mathbb{E}^{\hat{P}}(e^{-\alpha(Q(Y_T) - Q(Y_0))}) \quad = \quad 0$$

and hence (from (15)*) we see that*

$$\liminf_{T\to\infty} -\frac{1}{\alpha T} \log \mathbb{E}(e^{-\alpha(X_T - X_0)}) \quad \leq \quad \beta \tag{16}$$

and we have equality in (16) *for the optimal control $u_t = \hat{u}(Y_t)$.*

Proof. Set $Z_t = \log Y_t$. From Ito's formula we see that

$$dZ_t = [\alpha\sigma^2(\bar{Y} - e^{Z_t}(1 - q(e^{Z_t}))) + u_t e^{-Z_t} - \frac{\sigma^2}{2}]dt + \sigma d\hat{W}_t$$

and recall that $q(y) = Q'(y) \leq 1$ so

$$Q(y) = \int_0^y q(\zeta)d\zeta \quad \leq \quad y\,.$$

From Lemma 5.2 we now see that for large enough z, the drift of Z_t satisfies (assuming u is uniformly bounded from above by some $\bar{u}$)

$$\lim_{z\to\infty} [\alpha\sigma^2(\bar{Y} - e^z(1 - q(e^z))) + ue^{-z} - \frac{\sigma^2}{2}] \leq -\frac{\sigma^2}{2}.$$

Let us define

$$L := \inf\{z > 0 : \alpha\sigma^2(\bar{Y} - e^v(1 - q(e^v))) + \bar{u}e^{-v} - \frac{\sigma^2}{2} \leq -\frac{\sigma^2}{4}, \quad \forall v > z\}$$

From the above we know that L is finite. For any $t<0$, let us define

$$\varrho_t := \sup\{s<t : Z_s < L\}.$$

Then for all $s\in(\varrho_t,t]$, $Z_t\geq L$. By comparison, we also know that,

$$Z_s \leq U_s, \quad \forall s\in(\varrho_t,t],$$

where

$$U_{\rho_t}\equiv L, \quad \forall s\in[0,\varrho_t] \quad ; \quad dU_s = -\frac{\sigma^2}{4}ds+\sigma d\hat{W}_s, \quad \forall s>\varrho_t.$$

It follows that, for any $t>0$, we have

$$\begin{aligned}\mathbb{E}[e^{-\alpha Y_t}] =&\mathbb{E}[e^{-\alpha e^{Z_t}}1_{Z_t\leq L}]+\mathbb{E}[e^{-\alpha e^{Z_t}}1_{Z_t>L}]\\ \geq& e^{-\alpha e^L}P(Z_t\leq L)+\mathbb{E}[e^{-\alpha e^{U_t}}1_{Z_t>L}]\\ \geq& e^{-\alpha e^L}P(Z_t\leq L)+\mathbb{E}[e^{-\alpha e^{\sup_{s\geq\varrho_t}U_s}};Z_t>L]\\ \geq&\mathbb{E}[e^{-\alpha e^{\sup_{s\geq 0}U'_s}}]\end{aligned} \tag{17}$$

where U'_s is defined as

$$dU'_s = -\frac{\sigma^2}{4}ds+\sigma d\hat{W}_s, \quad \forall s>0\,,\, U'_0=L.$$

It is well known that $\sup_{s\geq 0}U'_s - L$ follows an exponential distribution with parameter $\Phi(0)$ (see e.g. page 213 in [Kyp06] or just use the reflection principle and then apply Girsanov's theorem and let $t\to\infty$), where

$$\Phi(0) := \sup\{m>0 : \frac{\sigma^2}{2}m^2-\frac{\sigma^2}{4}m=0\}$$

i.e. $\Phi(0)=\frac{1}{2}$. Hence we have

$$\mathbb{E}[e^{-\alpha e^{\sup_{s\geq 0}U'_s}}] = \int_0^\infty e^{-\alpha e^{L+x}}\frac{1}{2}e^{-\frac{1}{2}x}dx > \int_0^L e^{-\alpha e^{L+x}}\frac{1}{2}e^{-\frac{1}{2}x}dx \geq e^{-\alpha e^{2L}}[1-e^{-\frac{1}{2}L}]>0.$$

Putting this together and using (17), we find that

$$\liminf_{T\to\infty}\frac{1}{T}\log\mathbb{E}[e^{-\alpha(Q(Y_T)-Q(Y_0))}] \quad\geq\quad \liminf_{T\to\infty}\frac{1}{T}\log\mathbb{E}[e^{-\alpha Y_T}] \quad\geq\quad \liminf_{T\to\infty}\frac{1}{T}\log\mathbb{E}[e^{-\alpha e^{\sup_{s\geq 0}U'_s}}] \quad=\quad 0\,. \tag{18}$$

Moreover, from Remark 5.3 we know that $Q(y)\geq(\bar{Y}+\varepsilon)\times q(0)$, hence

$$\liminf_{T\to\infty}\frac{1}{T}\log\mathbb{E}[e^{-\alpha(Q(Y_T)-Q(Y_0))}] \quad\leq\quad 0\,.$$

Finally, for the optimal control $u_t=\hat{u}(Y_t)$ we can easily show $Y_t\geq 0$, because the solution to $dY_t = Y_t(\mu dt+\sigma dW_t)+\hat{u}(y)dt$ sits above the solution to $dY_t = Y_t(\mu dt+\sigma dW_t)+\hat{u}(y)dt$ for Y_t small because $u(0+)>0$ (to rigourize this we just compute the scale function for Y and apply Proposition 5.22 in [KS91]. ■

Remark 5.4 Intuitively, $-q(y)$ represents the change in the certainty equivalent if the investor has one extra dollar in stock (i.e. the Y variable), while keeping his total wealth (i.e. cash plus stock, the X variable) constant. It also makes sense that $q(\infty)=1$, because as the stock position becomes huge an extra dollar in stock with the riskless position held constant (i.e. a extra 1 dollar in total wealth) becomes useless; looking at the exponent of the ansatz to the ODE (divided by $-\alpha$), i.e. $x-\beta t+\int_0^y q(u)du$ we see that there will be an increase of one in the total capital, offset by a decrease of one in the $\int_0^y q(\zeta)d\xi$ term, i.e.no net gain in welfare.

6 Non-linear price impact

In this section, we posit the same dynamics for S as in section 3, but we now assume that the price paid per share is

$$\tilde{S}_t \quad:=\quad S_t(1+\lambda|S_t\dot{\theta}_t|^\gamma\mathrm{sgn}(\dot{\theta})) \tag{19}$$

for some $\gamma \in (0,1]$, so the cash position evolves as

$$dC_t \;=\; -S_t(1+\lambda|S_t\dot{\theta}_t|^{\gamma}\mathrm{sgn}(\dot{\theta}))d\theta_t\,.$$

Then from Ito's rule we have

$$\begin{aligned} dX_t \;=\; S_t d\theta_t + \theta_t dS_t + dC_t \;&=\; \theta_t S_t(\mu dt + \sigma dW_t) \,-\lambda|S_t\dot{\theta}_t|^{\gamma+1}dt \\ &=\; Y_t(\mu dt + \sigma dW_t) \,-\lambda|u_t|^{\gamma+1}dt\,. \end{aligned}$$

and

$$dY_t \;=\; d(\theta_t S_t) \;=\; \theta_t dS_t \,+\, S_t\dot{\theta}dt \;=\; Y_t(\mu dt + \sigma dW_t) + u_t dt\,.$$

6.1 Heuristics for non-linear price impact

Proceeding as in section 4 in [GW14], we expect the value function $V(t,X_t,Y_t)$ to evolve as

$$\begin{aligned} dV(t,X_t,Y_t) \;&=\; V_t dt + V_x dX_t + V_y dY_t + \frac{1}{2}V_{xx}d\langle X\rangle_t + \frac{1}{2}V_{yy}d\langle Y\rangle_t + V_{xy}d\langle X,Y\rangle_t \\ &=\; V_t dt + V_x(Y_t(\mu dt + \sigma dW_t) \,-\lambda|u_t|^{\gamma+1}dt) + V_y(Y_t(\mu dt + \sigma dW_t) + u_t dt) \\ &+\; \frac{1}{2}V_{xx}Y_t^2\sigma^2 + \frac{1}{2}V_{yy}Y_t^2\sigma^2 + V_{xy}Y_t^2\sigma^2 dt\,. \end{aligned}$$

Applying the usual HJB argument we obtain

$$\begin{aligned} 0 \;&=\; \sup_u[V_t + V_x(y\mu \,-\lambda|u|^{\gamma+1}) + V_y(y\mu+u) + \frac{1}{2}V_{xx}y^2\sigma^2 + \frac{1}{2}V_{yy}y^2\sigma^2 + V_{xy}y^2\sigma^2] \\ &=\; V_t + \mu y V_x + \mu y V_y + \frac{1}{2}y^2\sigma^2 V_{xx} + \frac{1}{2}\sigma^2 y^2 V_{yy} + \sigma^2 y^2 V_{xy} \,+\, \sup_u[-\lambda|u|^{\gamma+1}V_x + V_y(y\mu+u)]\,. \end{aligned}$$

Substituting

$$V(t,x,y) \;=\; -e^{-\alpha x}e^{\alpha\beta t}e^{\alpha\int_0^y q(\zeta)d\zeta}$$

and maximizing over u (for $u>0$) yields the optimal monetary trading rate:

$$\hat{u}(y) \;=\; -\frac{\mathrm{sgn}(q(y))}{(1+\gamma)^{\frac{1}{\gamma}}\lambda^{\frac{1}{\gamma}}}\,|q(y)|^{\frac{1}{\gamma}}$$

which reduces the HJB equation to the ordinary differential equation to

$$\begin{aligned} &-2c(\lambda) - 2y\mu + \frac{\mu^2}{\alpha\sigma^2} + y^2\alpha\sigma^2 + y^2\alpha\sigma^2 q(y)^2 + 2\lambda|\frac{q(y)}{\lambda(1+\gamma)}|^{1+\frac{1}{\gamma}} \\ +\;\; &-2q(y)(-y\mu + y^2\alpha\sigma^2 + \mathrm{sgn}(q)|\frac{q(y)}{\lambda(1+\gamma)}|^{\frac{1}{\gamma}}) + y^2\sigma^2 q'(y) \;=\; 0 \end{aligned} \tag{20}$$

where $\beta(\lambda) = \frac{\mu^2}{2\alpha\sigma^2} - c(\lambda)$.

6.2 Asymptotic behaviour of $q(y)$ for λ small and $y \neq \bar{Y}$ fixed

If we expand the HJB equation in (20) for small λ and assume that $y \neq \bar{Y}$ and set $\beta = \frac{\mu^2}{2\alpha\sigma^2} - c(\lambda)$ for $c(\lambda) = o(1)$ and we further assume that $\lim_{\lambda\to 0} q(\lambda) = 0$ we find that

$$-\frac{q(y)^{1+\frac{1}{\gamma}}\alpha\gamma(\frac{1}{\lambda(1+\gamma)})^{\frac{1}{\gamma}}}{1+\gamma} \,+\, \frac{(\mu - y\alpha\sigma^2)^2}{2\sigma^2} \,+\, h.o.t. \;=\; 0$$

and re-arranging we see that a first order approximation for q is given by

$$\begin{aligned} q_1(y) \;&:=\; 2^{-\frac{\gamma}{1+\gamma}}\big[\frac{(1+\gamma)[\lambda(1+\gamma)]^{\frac{1}{\gamma}}(\mu - y\alpha\sigma^2)^2}{\alpha\gamma\sigma^2}\big]^{\frac{\gamma}{1+\gamma}} \\ &=\; 2^{-\frac{\gamma}{1+\gamma}}(1+\gamma)\gamma^{-\frac{\gamma}{1+\gamma}}(\frac{\alpha\sigma^2}{\gamma})^{\frac{\gamma}{1+\gamma}}\,\mathrm{sgn}(\bar{Y}-y)\,\lambda^{\frac{1}{1+\gamma}}\,|\bar{Y}-y|^{\frac{2\gamma}{1+\gamma}} \\ &=\; const.\times\mathrm{sgn}(\bar{Y}-y)\,\lambda^{\frac{1}{1+\gamma}}\,|\bar{Y}-y|^{\frac{2\gamma}{1+\gamma}} \end{aligned} \tag{21}$$

as $\lambda \to 0$. Moreover, setting $\gamma = 1$, we recover the same behaviour as in section 3, namely that $q(y) \sim q_1(y)\sqrt{\lambda}$ with $q_1(y)$ defined as in (8). (21) is similar to Eq 23 in [GW14]. For $\gamma \in (0,1)$, we see that $q_1(y)$ is sublinear and its derivative explodes at $y = \bar{Y}$. However, the true solution to the HJB should be bounded with bounded derivative. Hence to achieve sensible asymptotics in ths region, following [GW14] we now "zoom in" close to the Merton proportion $\bar{Y}$ by letting λ and $|y - \bar{Y}|$ go small simultaneously.

6.3 Asymptotic behaviour of $q(y)$ when λ and $|y - \bar{Y}|$ go small together

Inspired by [GW14], we now set $c(\lambda) = \tilde{c}_1 \lambda^{\frac{2}{\gamma+3}}(1 + o(1))$ as $\lambda \to 0$ and $y = \bar{Y} + \lambda^{\frac{1}{\gamma+3}} z$ and $q(y) = r(z)\lambda^{\frac{3}{\gamma+3}}$ as $\lambda \to 0$, so we see that $|y - \bar{Y}|$ now also tends to zero as $\lambda \to 0$. In this new regime, taking the limit $\lambda \downarrow 0$, the HJB equation now reduces to

$$-2\tilde{c}_1 + z^2\alpha\sigma^2 - 2\gamma(1+\gamma)^{-1-\frac{1}{\gamma}}|r(z)|^{1+\frac{1}{\gamma}} + \frac{\mu^2 r'(z)}{\alpha^2\sigma^2} = 0 \tag{22}$$

Trivially, we can re-write as (22)

$$-\frac{1}{2}z^2\alpha\sigma^2 + \tilde{c}_1 - \frac{\mu^2}{2\alpha^2\sigma^2}r'(z) + \gamma(1+\gamma)^{-1-\frac{1}{\gamma}}|r(z)|^{1+\frac{1}{\gamma}} = 0 \tag{23}$$

(note the similarity of this ODE to the ODE below Eq 25 in [GW14] for their $r_0(z)$ function). If we let $r(z) = As(w)$ where $w = Bz$ for some constants A, B, then setting

$$\begin{aligned} A &= 2^{-\frac{\gamma}{1+\gamma}}(\alpha\gamma\sigma^2)^{\frac{\gamma}{1+\gamma}}, \\ B &= -2^{\frac{\gamma}{1+\gamma}}\alpha^3(\alpha\gamma\sigma^2)^{-\frac{\gamma}{1+\gamma}}\sigma^4/\mu^2 \end{aligned}$$

we find that (22) transforms to

$$-c + w^2 - \gamma(\gamma+1)^{-\frac{\gamma+1}{\gamma}}|s(w)|^{\frac{\gamma+1}{\gamma}} - s'(w) = 0 \tag{24}$$

where $c = 2\tilde{c}_1/(\alpha\sigma^2)$, which is now the same ODE as Eq 12 in [GW14]. From Lemma 22 in [GW14], we know that there exists a unique constant $c_\gamma > 0$ and a unique solution $s_\gamma(w)$ to the ODE (24) such that

$$\lim_{w\to\pm\infty} \frac{s_\gamma(w)}{|w|^{\frac{2\gamma}{1+\gamma}}} = \mp(\gamma+1)\,\gamma^{-\frac{\gamma}{\gamma+1}}.$$

Translating back to q, we has thus found that

$$q(\bar{Y} + \lambda^{\frac{1}{\gamma+3}}z) \sim r(z)\lambda^{\frac{3}{\gamma+3}} = As_\gamma(Bz)r(z)\lambda^{\frac{3}{\gamma+3}} \tag{25}$$

as $\lambda \to 0$ (note the similarity of this formula to Eq 14 in [GW14]).

6.4 The limiting cases $\gamma = 1$ and $\gamma = 0$

For $\gamma = 1$, we can easily verify that $s(w) = -2w$ is the solution to (24), so $c_\gamma = 2$ (see the top of the proof of Lemma 6 in [GW14]) and $\tilde{c}_1 = \frac{1}{2}\alpha\sigma^2$. Note that in section 4 we found that $\tilde{c}_1 = \frac{\mu^2}{\sqrt{2}\alpha^{\frac{3}{2}}}$, which shows that this constant is different in the two regimes.

From Lemma 7 in [GW14], we know that $c_0 := \lim_{\gamma\to 0} c_\gamma = (\frac{3}{2})^{\frac{2}{3}}$.

6.5 Small-λ asymptotics for the optimal trading policy

From the expression for $\hat{u}(y)$ above, we see that for $z \neq 0$ fixed we have

$$\begin{aligned} \hat{u}(\bar{Y} + \lambda^{\frac{1}{\gamma+3}}z) \sim -\frac{\mathrm{sgn}(q(y))}{(1+\gamma)^{\frac{1}{\gamma}}\lambda^{\frac{1}{\gamma}}}|q(y)|^{\frac{1}{\gamma}} &= -\mathrm{sgn}(y-\bar{Y})\frac{|r(z)\lambda^{\frac{3}{\gamma+3}}|^{\frac{1}{\gamma}}}{(1+\gamma)^{\frac{1}{\gamma}}\lambda^{\frac{1}{\gamma}}} \\ &= -\mathrm{sgn}(z)\frac{|r(z)|^{\frac{1}{\gamma}}}{(1+\gamma)^{\frac{1}{\gamma}}}\frac{1}{\lambda^{\frac{1}{3+\gamma}}} \\ &= -\mathrm{sgn}(z)\frac{|As_\gamma(Bz)|^{\frac{1}{\gamma}}}{(1+\gamma)^{\frac{1}{\gamma}}}\frac{1}{\lambda^{\frac{1}{3+\gamma}}} \end{aligned}$$

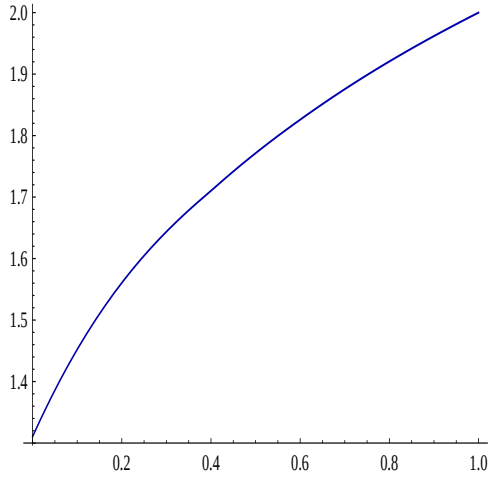

Figure 1: Here we have plotted and tabulated the universal constant $\gamma \mapsto c_\gamma$ for $\gamma \in (0, 1]$, using a bisection trial and error approach in Mathematica with the NDSolveValue command. Note this is the same graph as Figure 2 in [GW14].

as $\lambda \to 0$ (note the similarity of this formula to Eq 17 in [GW14]). The formula shows that (at leading order), the optimal policy is to buy when the weight is below the frictionless target $\bar{Y}$, and to sell when it is above, as with linear impact. In contrast to the linear case) $r(z)$ is not linear in general for $\gamma \neq 1$, so $\hat{u}$ is no longer linear in $y - \bar{Y}$. Setting $\gamma = 1$, the final expression here is $O(\lambda^{-\frac{1}{4}})$, which agrees with the $\gamma = 1$ calculations in (9) when we let $y = \bar{Y} + \lambda^{\frac{1}{\gamma+3}} z = \bar{Y} + \lambda^{\frac{1}{4}} z$.

The equivalent safe rate is given by

$$\begin{aligned} \beta(\lambda) &= \frac{\mu^2}{2\alpha\sigma^2} - \tilde{c}_1 \lambda^{\frac{2}{\gamma+3}} [1 + o(1)] \\ &= \frac{\mu^2}{2\alpha\sigma^2} - \frac{1}{2} c_\gamma \alpha\sigma^2 \lambda^{\frac{2}{\gamma+3}} [1 + o(1)] \end{aligned}$$

as $\lambda \to 0$.

γ	c_γ
0.0	1.31037
0.2	1.5605551
0.4	1.71006801
0.6	1.82587907
0.8	1.92060895
1.0	2

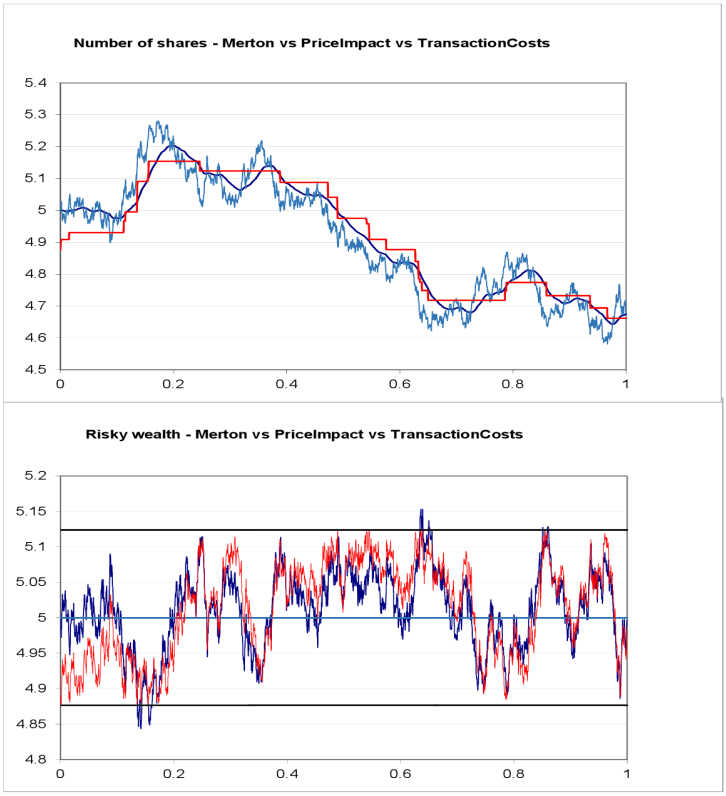

Figure 2: In the upper graph here, we have plotted a single Monte Carlo simulation of how the number of shares $\theta_t = Y_t/S_t$ evolves using the (asymptotically) optimal trading strategy $\hat{u}(y) = -\frac{\sqrt{\alpha}\,\sigma(y-\bar{Y})}{\sqrt{2}}\frac{1}{\sqrt{\lambda}}$ under linear price impact from (9) (dark blue) against the evolution of θ_t in the frictionless Merton setting (light blue) and the optimal θ_t process under the Guasoni&Muhle-Karbe[GM15] model (red) with proportional transaction costs but no price impact. We see that the price impact curve smoothly tracks the Merton optimal portfolio and the transaction costs curve is piecewise constant because of the no-trade region. Here the parameters are $\mu = .05$, $\sigma = .1$, $\lambda = .0.00001$, $\alpha = 1$, the size of the proportional transaction costs is $\varepsilon = .0001$ and the time horizon here is $t = 1$ year. The lower graph shows the corresponding evolution of the risky wealth under all three models. For the model in [GM15], the process $d\Upsilon_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t + dL_t - dU_t$ is arithmetic Brownian motion with two reflecting barriers (hence the local processes L_t and U_t) at $\Upsilon = 0$ and $\Upsilon = b = \log\frac{u}{l}$ and their risky wealth process $Y_t = le^{\Upsilon_t}$ where $l = (\bar{\mu} - \bar{\lambda})/\alpha$, $\bar{\mu} = \mu/\sigma^2$ and $\bar{\lambda} = (\frac{3}{4}\bar{\mu}^2)^{\frac{1}{3}}\varepsilon^{\frac{1}{3}}$, $u = (\bar{\mu} + \bar{\lambda})/\alpha$, and the optimal number of shares evolves as $d\theta_t/\theta_t = d\log\theta_t = dL_t - dU_t$ (θ_t here is φ_t in their notation); hence we see that Y_t evolves within the so-called no-trade region $[l, u]$, and trading only occurs when Υ hits the reflecting barriers; specifically we sell when hit the upper barrier u and buy when we hit the lower barrier b.

References

[AC01] Almgren, R. and N.Chriss, "Optimal execution of portfolio transactions", *J. Risk*, 3:5,39, 2001.

[Alm03] Almgren, R., Optimal execution with nonlinear impact functions and trading-enhanced risk, *Applied Mathematical Finance*, 10, 1-18, 2003.

[BSV15] Bank, P., H.M.Soner, M.Voss, "Hedging with Transient Price Impact", preprint, 2015.

[Con86] Constantinides, G., 'Capital market equilibrium with transaction costs", *Journal of Political Economy*, pp. 842-862, 1986.

[DN90] Davis, M. and Norman, A., "Portfolio selection with transaction costs", *Mathematics of Operations Research*, pp. 676-713, 1990.

[GS13] Gatheral, J. and A.Schied, "Dynamical models for market impact and algorithms for optimal order execution", In: Handbook on Systemic Risk (eds.: J.-P. Fouque and J. Langsam), Cambridge University Press, 579-602, 2013.

[GW14] Guasoni, P. and M.Weber, "Nonlinear Price Impact and Portfolio Choice", preprint.

[GW15] Guasoni, P. and M.Weber, "Dynamic trading volume", *Mathematical Finance*, forthcoming.

[GM15] Guasoni, P. and J.Muhle-Karbe, "Horizons, High Risk Aversion, and Endogeneous Spreads", *Mathematical Finance*, 25, no.4, 724-753, 2015 .

[KS91] Karatzas, I. and S.Shreve, "Brownian motion and Stochastic Calculus", Springer-Verlag, 1991.

[Kyp06] Kyprianou, A., "Introductory Lectures on Fluctuations of Lévy Processes with Applications", Springer, 2006

[Mer69] Merton, R., "Lifetime portfolio selection under uncertainty: The continuous-time case", *The review of Economics and Statistics*, 51(3), 247-257, 1969.

[Pham09] Pham, H., "Continuous-time Stochastic Control and Optimization with Financial Applications", Springer, 2009.

[RS10] L.C.G.Rogers and S.Singh, "The cost of illiquidity and its effects on hedging", *Mathematical Finance*, 20:597-615, 2010.

[SS08] Schied, A. and T. Schoneborn, "Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets", 2008.

[SST10] Schied, A., T. Schoneborn and Michael Tehranchi, "Optimal basket liquidation for CARA investors is deterministic", *Applied Mathematical Finance*, 17(6): 471–489, 2010.

A Proof of Proposition 5.1

Remark A.1 Given two slope fields $f_1(y,q)$, $f_2(y,q)$ and suppose that $f_1(y,q) > f_2(y,q)$ for all $(y,q) \in \mathbb{R}_+ \times \mathbb{R}$. If $q_i(\cdot)$ solves $q_i'(y) = f_i(y, q_i(y))$, $i = 1, 2$, with initial condition $q_1(y_0) = q_2(y_0)$ for some $y_0 \in \mathbb{R}$. Denoting the common domain of $q_1(\cdot)$ and $q_2(\cdot)$ by $\mathcal{D}$, then it can be easily seen that $q_1(y) < q_2(y)$ for all $y \in (0, y_0) \cap \mathcal{D}$; and $q_1(y) > q_2(y)$ for all $y \in (y_0, \infty) \cap \mathcal{D}$.

The following Proposition summarizes the qualitative behaviour of the zero contours of $f(y,q)$ (see also Figure 3 below), and will be needed for the rest of the proof which follows.

Proposition A.1 *Assume $0 < \lambda \leq \frac{\alpha\sigma^2}{8\mu^2} = \frac{1}{16\beta^*}$ (where β^* is defined in section 4). Then for any $\beta \in (0, \beta^*]$, there exist two functions $h_-(y)$ and $h_+(y)$ such that $f(y, h_\pm(y)) = 0$ such that*

1. *$h_-(y)$ and $h_+(y)$ are strictly increasing and decreasing respectively, for all $y \in (0, y_-)$, and $h_\pm(0+) = \pm 2\sqrt{\beta\lambda}$;*
2. *$h_-(y)$ is strictly decreasing for all $y \in (y_+, \hat{y})$, with $\lim_{y\uparrow\hat{y}} h_-(y) = -\infty$;*
3. *$h_-(y)$ is strictly decreasing for all $y \in (\hat{y}, \infty)$, with $\lim_{y\downarrow\hat{y}} h_-(y) = \infty$ and $\lim_{y\to\infty} h_-(y) = 1$;*
4. *$h_+(y)$ is strictly increasing for all $y \in (y_+, \infty)$, with $\lim_{y\to\infty} h_+(y) = 1$*

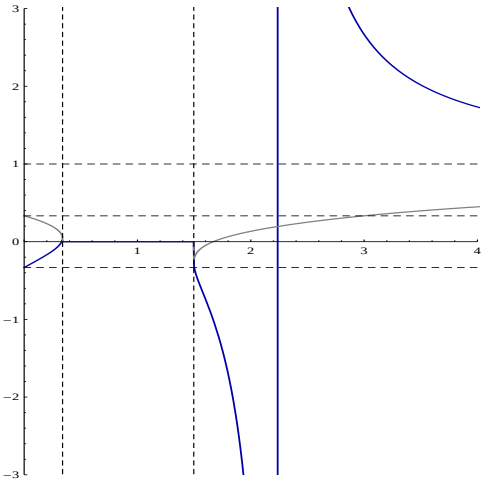

Figure 3: Here we have plotted the zero contours of $f(y,q)$ for $\lambda = .1$, $\sigma = 1$,$\mu = 1$, $\alpha = 1$ and β estimated using $\beta \approx \beta^* - \tilde{c}\lambda^{\frac{1}{2}}$ as derived in section 4 (note that the contour plot looks qualitatively the same for all parameter choices as proved in Proposition A.1, see Appendix C for full mathematical details). The blue line is h_- and the grey line is h_+. The three vertical asymptotics are (from left to right) y_-, y_+ and $\hat{y}$ respectively, and the three horizontal asymptotes (from top to bottom) 1, $2\sqrt{\beta\lambda}$ and $-2\sqrt{\beta\lambda}$. The solution $q(y)$ for the correct β value is extremely difficult to solve for numerically due to the peculiar nature of the ODE in (6), but from the analysis below we know that $\lim_{y\to 0} q(y) = -2\sqrt{\beta\lambda}$ and $\lim_{y\to\infty} q(y) = 1$ and $q(y)$ is increasing.

5. $h_\pm$ have the asymptotic behaviour

$$h_\pm(y) \;=\; \pm 2\sigma\sqrt{\beta\lambda} \,+\, \lambda\mu(2 \mp \frac{1}{\sigma\sqrt{\beta\lambda}})y \,+\, o(y) \qquad (y \to 0) \tag{A-1}$$

$$h_\pm(y) \;=\; 1 - \frac{2\lambda\mu \pm \sqrt{2}\sqrt{\lambda(\alpha\sigma^2 + 2\lambda(\mu^2 - 2\alpha\beta\sigma^4))}}{2\alpha\lambda\sigma^2}\,\frac{1}{y} \,+\, o(\frac{1}{y}) \qquad (y \to \infty) \tag{A-2}$$

where

$$y_\pm \;:=\; \frac{\bar{Y} \pm \sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)}}{1 + 4\lambda(\beta^* - \beta)} \qquad , \qquad \hat{y} \;:=\; \frac{1}{\sqrt{2\lambda\alpha\sigma^2}}\,. \tag{A-3}$$

Proof. See Appendix C. ■

Corollary A.2 *If $\tilde{q}$ is a solution to (6) that crosses h_- over $(0, y_-)$, then $\tilde{q}(0+) = h_+(0+) = 2\sqrt{\beta\lambda}$. If $\bar{q}$ is a solution to (6) that crosses h_+ over (y_+, ∞), then $\bar{q}(\infty) = -\infty$.*

- *There exists a unique solution $q_0(y)$ to (6) such that $-\infty < q_0(0+) < 0$ and q_0 is increasing on $(0, y_+)$.*
  Taking the limit for $y \downarrow 0$ of the slope field $f(y,q)$ yields $f(0^+, q) < 0$ for $q \in (-2\sqrt{\lambda\beta}, 2\sqrt{\lambda\beta})$ and $f(0^+, q) > 0$ for $q \in (-\infty, -2\sqrt{\lambda\beta}) \cup (2\sqrt{\lambda\beta}, +\infty)$ and from (A-1) we see that for λ sufficiently small, we have that $\lim_{y\downarrow 0} h'_+(y) < 0$ and $\lim_{y\downarrow 0} h'_-(y) > 0$.

  We now show that if a solution $q(y)$ has finite limit as $y \to 0$, then $q^2(0+) = 4\lambda\beta$. Since the functions $h_\pm(y)$ are monotone on a small interval $(0, \varepsilon)$, also any solution $q(y)$ is monotone close to 0. By taking the limit $y \downarrow 0$ in the equation $q'(y) = f(y, q(y))$, since we assume $q^2(0+) \neq 4\lambda\beta$, we get $q'(y) \sim \frac{K}{y^2}$ close to 0 for some constant K. This implies that $q(y) \sim -\frac{K}{y}$ close to 0, which contradicts the assumption that $q(0+)$ is finite.

  Proceeding along similar lines for the top of page 25 in [GW15], for each $z > 0$ we define

$$\underline{q}(z) \;:=\; \inf\{k : q(0+; z, k) = 2\sqrt{\lambda\beta}\}$$

  where $q(y; z, k)$ denotes the solution to the ODE in (6) which passes through the point (z, k). Fix y_0 sufficiently small and consider the solutions to the ODE given by $q_0(y) := q(y; y_0, \underline{q}(y_0))$ and $\tilde{q}_0(y) := q(y; y_1, \underline{q}(y_1))$ for some

$y_1 \in (0, y_0)$. Now assume that $\tilde{q}_0(y_1) < q_0(y_1)$. Then $\tilde{q}_0(y_0) < q_0(y_0)$ because solutions cannot cross. But this contradicts the definition of $q_0(.)$, hence we must have $\tilde{q}_0(y_0) \geq q_0(y_0)$. Conversely, if $\tilde{q}_0(y_0) > q_0(y_0)$ then $\tilde{q}_0(y_1) > q_0(y_1)$ which contradicts the definition of $\tilde{q}_0(.)$. Thus we see that $\tilde{q}_0(.) = q_0(y)$ for all $y \leq y_0$ so we can say that $q_0(.) = \inf\{q(.) : q(0+) = 2\sqrt{\lambda\beta}\,,\ q \text{ satisfies (6)}\}$.

We now prove that $q_0(0+) = -2\sqrt{\lambda\beta}$. First, assume by contradiction that $q_0(0+) = 2\sqrt{\lambda\beta}$, then for y_0 sufficiently small we must have $q_0(y_0) > h_-(y_0)$. The solution that starts from $(y_0, h_-(y_0))$ also has limit $2\sqrt{\lambda\beta}$ as $y \to 0$ because $\frac{dq}{dy} < 0$ in the region between h_- and h_+, and this solution is below $q_0(y)$, contradicting its minimality. Now, assume by contradiction that $q_0(0+) = -\infty$. Then there is a point $(y_1, \tilde{q}_0) \in (0, \varepsilon) \times (-\infty, -2\sqrt{\lambda\beta})$ such that $\tilde{q}_0 > q_0(y_1)$. Since $f(y, q) > 0$ on $(0, \varepsilon) \times (-\infty, -2\sqrt{\lambda\beta})$, the solution that starts in $(y_1, \tilde{q}_0)$ has limit $-\infty$ as $y \to 0$, but is also above $q_0(y)$, which leads to a contradiction by the definition of $q_0(.)$. From this analysis we also see that q_0 must be increasing for $y \in (0, y_-)$ because $\frac{dq}{dy} > 0$ below h_- for $y \in (0, y_-)$ (see the previous Proposition and Figure 3). Moreover, $\frac{dq}{dy} > 0$ for all $y \in (y_-, y_+)$.

To emphasize the dependence of q_0 on β, we can denote the solution $q_0(\cdot)$ by $q_0^{(\beta)}(\cdot)$. Then since $q_0^{(\beta)}(0+) = -2\sqrt{\beta\lambda}$, we know that for any $0 < \beta_1 < \beta_2$, $q_0^{(\beta_1)}(y) > q_0^{(\beta_2)}(y)$ for sufficiently small $y > 0$. We claim that this inequality actually holds over the common domain of these two functions. This is because, if there is a smallest $y_2 > 0$ such that $q_0^{(\beta_1)}(y_2) = q_0^{(\beta_2)}(y_2)$, then Remark A.1 implies that $q_0^{(\beta_2)}(0+) > q_0^{(\beta_1)}(0+)$, which is a contradiction.

- *There exists a solution $q_\infty(y)$ defined up to $+\infty$ such that the limit as $y \to \infty$ exists and is positive, and q_∞ is increasing on (y_-, ∞).*
From (A-2) we can see that $\lim_{y\uparrow\infty} h_\pm(y) = 1$ and $h_+(y) < h_-(y)$. Moreover, from (A-2), we see that (for λ sufficiently small) $h_-(y)$ is strictly decreasing and $h_+(y)$ is strictly increasing for $y > \hat{y}$ (see (A-3) and Figure 3 for the definition and significance of $\hat{y}$).

Since for λ sufficiently small (say, $\lambda \in (0, \frac{\alpha\sigma^2}{8\mu^2}]$) $f(y, 1) > 0$, we have that $f(y, q) > 0$ for $q \in (h_+(y), h_-(y))$ and $f(y, q) < 0$ for $q \in (-\infty, h_+(y)) \cup (h_-(y), +\infty)$. We prove now that a solution with initial condition $(\bar{y}, h_+(\bar{y}))$ has limit $-\infty$ as $y \to \infty$. Since $h_+(y)$ is increasing, this solution $q(y)$ is decreasing after it crosses the curve $(y, h_+(y))$, in particular $q(y) \leq h_+(\bar{y})$ for any $y \geq \bar{y}$. Since $q(y)$ is monotone, it has a limit as $y \to \infty$; assume by contradiction that this limit is finite. As the derivative of any monotone function with finite limit at infinity converges to 0 at infinity, from the equality $0 = \lim_{y\uparrow\infty} q'(y) = \lim_{y\uparrow\infty} f(y, q(y))$, we get $\lim_{y\uparrow\infty} q(y) = 1$. But this cannot hold because $q(y)$ is decreasing for $y \geq \bar{y}$. A similar argument proves that a solution with initial condition $(\bar{y}, h_-(\bar{y}))$ has limit 1 for $y \uparrow +\infty$. Proceeding as in the previous bullet point, we now define

$$\bar{q}(z) \quad := \quad \sup\{k : q(+\infty; z, k) = -\infty\}\,. \tag{A-4}$$

and consider the solution to the ODE given by $q_\infty(y) := q(y; \bar{y}, \bar{q}(\bar{y}))$. Then from the preceding arguments, we know that $\bar{q}(\bar{y}) > h_-(\bar{y})$, and using the same comparison arguments as in the previous bullet point we can easily show that $q_\infty(y) = \sup\{q(y) : q(+\infty) = -\infty, q(y) \text{ satisfies (6)}\}$. Hence $q_\infty(y)$ is a solution to the ODE, $\lim_{y\uparrow\infty} q_\infty(y) = 1$ and it is increasing for $y > y_+$ because $\frac{dq}{dy} > 0$ in the region between h_- and h_-, and q_∞ cannot cross h_+. Moreover (as before) $\frac{dq}{dy} > 0$ for all $y \in (y_-, y_+)$.

- *If $\beta \geq \beta^*$, then $q_0(y) \leq 0 < q_\infty(y)$ for all $y > 0$.*
Notice that for all $y > 0$

$$\begin{aligned} f(y, 0) =& \frac{1}{\sigma^2 y^2}(2\mu y - \alpha\sigma^2 y^2 - 2\beta) = -\frac{\alpha}{y^2}[(y^2 - \frac{2\mu}{\alpha\sigma^2}y + \frac{2\beta^*}{\alpha\sigma^2}) + \frac{2(\beta - \beta^*)}{\alpha\sigma^2}] \\ =& -\frac{\alpha}{y^2}[(y^2 - 2\bar{Y}y + \bar{Y}^2) + \frac{2(\beta - \beta^*)}{\alpha\sigma^2}] = -\frac{\alpha}{y^2}[(y - \bar{Y})^2 + \frac{2(\beta - \beta^*)}{\alpha\sigma^2}] \leq 0. \end{aligned}$$

Since $q_0(0) = -2\sqrt{\lambda\beta} < 0$ and $q_\infty(+\infty) = 1 > 0$, neither $q_0(y)$ nor $q_\infty(y)$ can cross the line $q = 0$ (as they would have to do so with positive derivative).

- Fix a $\varepsilon \in (0, \beta^*)$ we let $\beta = \beta^* - \varepsilon$. Then we claim that for sufficiently small $\lambda > 0$, either $q_0(\cdot)$ explodes to $+\infty$ at some finite y-value (so $q_0(\cdot)$ has a finite domain) or there exists a $\bar{y}$ in the common domain of $q_0(\cdot)$ and $q_\infty(\cdot)$ such that $q_0(\bar{y}) \geq 1 > q_\infty(\bar{y})$. To that end, recall that the slope field $f(y, q)$ is

$$\frac{1}{\sigma^2 y^2}[(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q^2 - 2y\alpha\sigma^2(\bar{Y} - y)q - \alpha\sigma^2(\bar{Y} - y)^2 + 2\varepsilon].$$

Let us choose a fixed $\delta \in (0, \frac{1}{2}\sqrt{\varepsilon/(2\alpha\sigma^2)}))$ so that, for sufficiently small $\lambda > 0$ we have

$$0 < (\bar{Y} - \sqrt{\frac{2\varepsilon}{\alpha\sigma^2}(1-4\lambda\beta)})/(1+4\varepsilon\lambda) \equiv y_- < \bar{Y} - \delta < \bar{Y} + 2\delta < y_+ \equiv (\bar{Y} + \sqrt{\frac{2\varepsilon}{\alpha\sigma^2}(1-4\lambda\beta)})/(1+4\varepsilon\lambda).$$

Let us also impose that δ is also lies in $(0, \frac{1}{6}\bar{Y}(\sqrt{1+3\varepsilon/(\alpha\sigma^2\bar{Y}^2)}-1)$ so that

$$-12\alpha\sigma^2\delta^2 - 4\alpha\sigma^2\bar{Y}\delta + \varepsilon > 0 \tag{A-5}$$

(this will be needed in Eq (A-6) below). On the other hand, notice that for any $(y,q) \in [\bar{Y}-\delta, \bar{Y}+2\delta] \times [-1,1]$, we always have

$$\begin{aligned} \frac{1}{2\lambda} - \alpha\sigma^2 y^2 &> 0, \\ -2y\alpha\sigma^2(\bar{Y}-y)q &\geq -4(\bar{Y}+2\delta)\alpha\sigma^2\delta, \\ -\alpha\sigma^2(\bar{Y}-y)^2 &\geq -4\alpha\sigma^2\delta^2, \end{aligned}$$

where the first inequality just follows from $\bar{Y} + 2\delta < y_+ < \hat{y} = 1/\sqrt{2\alpha\sigma^2\lambda}$ (again see Proposition A.1 and Figure 3 for physical significance of $\hat{y}$). Adding the three inequalities above, we see that on $[\bar{Y}-\delta, \bar{Y}+2\delta] \times [-1,1]$ we have

$$\begin{aligned} f(y,q) \quad &\geq \quad \frac{1}{\sigma^2 y^2}[(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q^2 + \frac{1}{\sigma^2 y^2}[-12\alpha\sigma^2\delta^2 - 4\alpha\sigma^2\bar{Y}\delta + 2\varepsilon] \\ &\geq \quad \frac{1}{\sigma^2 y^2}(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q^2 + \frac{\varepsilon}{\sigma^2 y^2} \qquad \text{(A-6)} \\ &\geq \quad \frac{1}{\sigma^2 y^2}(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q^2 + \varepsilon^\star, \qquad \text{(A-7)} \end{aligned}$$

where $\varepsilon^\star := \frac{\varepsilon}{\sigma^2(\bar{Y}+2\delta)^2}$, and the second inequality follows from (A-5).

Now recall that $q_0(\cdot)$ is strictly increasing on the intersection of its domain and $(0, y_+)$, hence we can bound $q_0(\cdot)$ by $q_0(0+) = -2\sqrt{\beta\lambda}$ over $(0, \bar{Y}-\delta)$. Moreover, we just showed that the slope field is bounded by $\varepsilon^\star$ from below over the domain $[\bar{Y}-\delta, \bar{Y}+\delta] \times [-1,1]$, we know that we can bound $q_0(\cdot)$ by a function $g(\cdot)$ over $[0, \bar{Y}+\delta]$, where $g(y) = -2\sqrt{\lambda\beta} + \varepsilon^\star(y - \bar{Y} + \delta)^+$. As a consequence, $q_0(\cdot)$ either explodes to $+\infty$ by $\bar{Y}+\delta$, or $\infty > q_0(\bar{Y}+\delta) > g(\bar{Y}+\delta) = 2\delta\epsilon^\star - 2\sqrt{\beta\lambda}$. Let us assume further that $\lambda > 0$ is sufficiently small so that $q_0(\bar{Y}+\delta) > 2\delta\epsilon^\star - 2\sqrt{\beta\lambda} > 0$. Thus, over $[\bar{Y}+\delta, \bar{Y}+2\delta]$, we either have $q_0(\cdot)$ explodes to $+\infty$ or it is finite and increasing over $q_0(\bar{Y}+\delta) > 2\varepsilon^\star\delta - 2\sqrt{2\beta\lambda} > 0$, based on the comments about $q_0(\cdot)$ we made earlier. It follows that

$$\begin{aligned} q_0(\bar{Y}+2\delta) \quad &> \quad -2\sqrt{\lambda\beta} + \int_{\bar{Y}+\delta}^{\bar{Y}+2\delta} q_0'(y)dy \\ &\geq \quad -2\sqrt{\lambda\beta} + \int_{\bar{Y}+\delta}^{\bar{Y}+2\delta} [\frac{1}{\sigma^2 y^2}(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)q_0(y)^2 + \varepsilon^\star]dy \\ &\qquad \text{(from Eq (A-6))} \\ &\geq \quad -2\sqrt{\lambda\beta} + \int_{\bar{Y}+\delta}^{\bar{Y}+2\delta} [\frac{1}{\sigma^2 y^2}(\frac{1}{2\lambda} - \alpha\sigma^2 y^2)(-2\sqrt{\lambda\beta} + 2\varepsilon^\star\delta)^2 + \varepsilon^\star]dy \\ &> \quad 1 \qquad \text{(A-8)} \end{aligned}$$

for λ sufficiently small. This proves the our claim.

- *There exists β such that $q_0(y) = q_\infty(y)$.*
From the previous points we know that for some $\beta_{up} > \beta^*$ the function $q_0(y)$ is strictly negative on its domain, while $q_\infty(y)$ is strictly positive and defined on $(0, +\infty)$. For some $\beta_{low} < \beta^*$ there exists a point $\bar{y}$ such that $q_0(\bar{y}) \geq 1$. Consider a point $\tilde{y}$ close to 0 that belongs to the domain of q_0 for every $\beta \in [\beta_{low}, \beta_{up}]$. Assume for the moment that $\tilde{y}$ belongs to the domain of q_∞ for β_{low}, then we have that $q_0(\tilde{y}) < 0 < q_\infty(\tilde{y})$ for β_{up} and $q_0(\tilde{y}) > q_\infty(\tilde{y})$ for β_{low}. Since the solutions q_0 and q_∞ depend continuously on β, there exists a $\beta(\lambda) \in (\beta_{low}, \beta_{up})$ such that $q_0(\tilde{y}) = q_\infty(\tilde{y})$ (and therefore $q_0 = q_\infty$).
If $\tilde{y}$ does not belong to the domain of q_∞ for β_{low}, then since also this domain depends continuously on β there exists $\tilde{\beta} \in (\beta_{low}, \beta^*)$ such that for $\tilde{\beta}$ the point $\tilde{y}$ belongs to the domain of q_∞ and q_∞ has limit $-\infty$ at the left end of its domain. As before, for some $\beta \in (\beta_{low}, \tilde{\beta})$, $q_0(\tilde{y}) = q_\infty(\tilde{y})$.

- Now that there is at least a $\beta \in (0, \beta^*)$ such that there is a solution $q(\cdot)$ to the ODE, which is strictly increasing and $q(\cdot) < h_-(\cdot)$ over (y_+, ∞) and $q(\cdot) > h_+(\cdot)$ over $(0, y_+)$. Suppose there are $0 < \beta_1 < \beta_2 < \infty$ satisfies the above properties, let us denote by $q^{(\beta_i)}(\cdot)$, $h_\pm^{(\beta)}(\cdot)$, $y_\pm^{(\beta_i)}$, $i = 1, 2$, the corresponding solution, the 0-contours, and the critical points where $h_+^{(\beta_i)}(\cdot)$ collapses to $h_-^{(\beta_i)}(\cdot)$. Then from the discussion about $q_0(\cdot)$ and Remark A.1 we know that we must have
$$q^{(\beta_1)}(y) > q^{(\beta_2)}(y) \quad , \quad \forall y \in (0, \infty).$$
On the other hand, fix any $y_0 > \max\{y_+^{(\beta_i)}; i = 1, 2\}$, we denote by $\bar{q}^{(\beta_1)}(y) = q(\cdot; y_0, q^{(\beta_2)}(y_0))$. Then by Remark A.1 we have that
$$q^{(\beta_1)}(y) > \bar{q}^{(\beta_1)}(y; y_0, q^{(\beta_2)}(y_0)) > q^{(\beta_2)}(y), \quad \forall y \in (y_0, \infty).$$
In particular, by construction of $q_\infty^{(\beta_1)}(\cdot)$ we know that $\bar{q}^{(\beta_1)}(\cdot; y_0, q^{(\beta_2)}(y_0))$ will cross $h_+^{(\beta_1)}(\cdot)$ at some y-value and we will have $\bar{q}^{(\beta_1)}(+\infty; y_0, q^{(\beta_2)}(y_0)) = -\infty$. However, that implies that $q^{(\beta_2)}(+\infty) = -\infty$, which is impossible. Hence, there is a unique $\beta \in (0, \beta^*)$ such that the ODE has a bounded, strictly increasing solution $q(\cdot)$, that satisfies $q(0+) = -2\sqrt{\beta\lambda}$ and $q(+\infty) = 1$.

Remark A.2 From Corollary A.2 and the above construction we know that the global solution q satisfies $q(y) < h_-(y)$ over $(0, y_-]$ and $1 > q(y) > h_+(y)$ over $[y_+, \infty)$.

B Proof of Lemma 5.2

We begin by introducing
$$G(z) = zF(1/z), \quad p(z) = q(1/z), \quad \forall z > 0.$$
Then we have
$$G(z) = \bar{Y}z + p(z) - 1, \quad G(0+) = 0.$$
We will first establish that
$$-\sqrt{\frac{2(\beta^* - \beta + 1/(4\lambda))}{\alpha\sigma^2}} \le \liminf_{y\to\infty} F(y) \le \limsup_{y\to\infty} F(y) \le \bar{Y}. \tag{B-1}$$
Then we will improve the upper bound by comparison and contradiction. First, using that $p(z) < 1$ for all $z \in (0, \infty)$, we know that
$$F(1/z) = \frac{G(z)}{z} < \bar{Y}, \quad \forall z \in (0, \infty).$$
Recall that $G(z) := zF(1/z) = \bar{Y}z - (1 - p(z)) = \bar{Y}z - (1 - q(\frac{1}{z}))$. And we know that $h_+(y) < q(y)$ (see Remark A.2) for all $y \ge y_+$, so we have
$$\bar{Y}z + h_+(\frac{1}{z}) - 1 < \bar{Y}z + q(\frac{1}{z}) - 1 = G(z) < \bar{Y}z, \quad \forall z \in (0, 1/y_+).$$
Using the explicit expression for $h_+(\cdot)$ we know that
$$\lim_{z\downarrow 0} \frac{h_+(1/z) - 1}{z} = \lim_{y\to\infty} \frac{y[\bar{Y}y - \hat{y}^2 + \sqrt{\Delta(y)}/(\alpha\sigma^2)]}{\hat{y}^2 - y^2} = -\sqrt{\frac{2(\beta^* - \beta + 1/(4\lambda))}{\alpha\sigma^2}} - \bar{Y}.$$
and $\hat{y} = \frac{1}{\sqrt{2\alpha\sigma^2\lambda}}$ as in Proposition A.1.

$$\bar{Y}z + h_+(\frac{1}{z}) - 1 < G(z) < \bar{Y}z, \quad \forall z \in (0, 1/y_+).$$
Hence we have the upper bound in (B-1). On the other hand, by the fact that $p(1/z) > h_+(1/z)$ for all $z \in (0, 1/y_+)$, we know that
$$\begin{aligned}
\liminf_{z\downarrow 0} \frac{G(z)}{z} \ge& \bar{Y} + \liminf_{z\downarrow 0} \frac{h_+(1/z) - 1}{z} \\
=& \bar{Y} + \lim_{y\to\infty} \frac{y[\bar{Y}y - \hat{y}^2 + \sqrt{\Delta(y)}/(\alpha\sigma^2)]}{\hat{y}^2 - y^2} \\
=& -\sqrt{\frac{2(\beta^* - \beta + 1/(4\lambda))}{\alpha\sigma^2}}.
\end{aligned}$$

This proves the lower bound in (B-1).

We now prove that $G'(z) \leq 0$ for z sufficiently small, and hence $G(z) \leq 0$ for z sufficiently small (note that $G(0) = 0$, so if $G'(z) \leq 0$ for small z then G is negative for z small). This will prove the upper bound in (13). Suppose this is not the case, then there is a $z_0 > 0$ sufficiently small such that $p^2(z)$ is strictly decreasing over $(0, 2z_0)$, and that

$$G'(z_0) > 0, \tag{B-2}$$

$$-\frac{1}{\alpha} \cdot c_-(\lambda, \frac{1}{2z_0}) = \frac{1}{2\alpha}\left[1 + \sqrt{1 + 4\alpha\left(\frac{p^2(2z_0)}{2\lambda\sigma^2} + \frac{2(\beta^* - \beta)}{\sigma^2} - \bar{Y}\right)}\right] > \bar{Y}. \tag{B-3}$$

(B-3) holds due to the assumption $-c_-(\lambda, \infty)/\alpha > \bar{Y}$ and the fact that $p(z)$ continuously increases to 1 as z decreases to 0. On the other hand, from $p'(z) = q'(1/z) \cdot (-1/z^2) = -f(1/z, q(1/z))/z^2$, we see that $G(z)$ solves a Riccati equation:

$$G'(z) = \bar{Y} - f(\frac{1}{z}, p(z))/z^2 = \bar{Y} - \frac{1}{\sigma^2}[\frac{q^2}{2\lambda} + 2(\beta^* - \beta) - \alpha\sigma^2[y(1-q) - \bar{Y}]^2] \tag{B-4}$$

and noting that $y(1-q) - \bar{Y} = -G(z)/z$, we see that

$$G'(z) = A(z) + \frac{\alpha}{z^2}G^2(z), \quad \forall z > 0, \tag{B-5}$$

where

$$A(z) = \bar{Y} - \frac{2(\beta^* - \beta)}{\sigma^2} - \frac{p^2(z)}{2\lambda\sigma^2}.$$

Now consider the following proxy Riccati equation:

$$\mathcal{G}'(z) = A(2z_0) + \frac{\alpha}{z^2}\mathcal{G}^2(z), \quad \mathcal{G}(z_0) = G(z_0), \quad \forall z \in (0, z_0). \tag{B-6}$$

Making the substitution $\mathcal{G}(z) = -\frac{z^2}{\alpha}\frac{u'(z)}{u(z)}$, we see that $u(z)$ satisfies the linear Eq

$$\frac{A\alpha u(z)}{z^2} + \frac{2u'(z)}{z} + u''(z) = 0$$

where A is shorthand for $A(2z_0)$, which has solution $u(z) = B_+ z^{c^+} + B_- z^{c^-}$ for arbitrary constants B_+, B_-, where $c_\pm$ satisfy $c^2 + c + A\alpha = 0$, which has solutions given by $c_\pm = \frac{1}{2}(-1 \pm \sqrt{1 - 4\alpha A}) \equiv c_\pm(\lambda, \frac{1}{2z_0})$, and we see that

$$\begin{aligned} c_\pm &= \frac{1}{2}(-1 \pm \sqrt{1 - 4\alpha A(2z_0)}) \\ &= \frac{1}{2}(-1 \pm \sqrt{1 - 4\alpha(\bar{Y} - \frac{2(\beta^* - \beta)}{\sigma^2} - \frac{p^2(2z_0)}{2\lambda\sigma^2})}) \\ &= \frac{1}{2}\big[-1 \pm \sqrt{1 + 4\alpha(\frac{q^2(1/2z_0)}{2\lambda\sigma^2} + \frac{2(\beta^* - \beta)}{\sigma^2} - \bar{Y})}\big] \end{aligned} \tag{B-7}$$

which agrees with the definition of $c_\pm(\lambda, \frac{1}{2z_0})$. From this we find that

$$\mathcal{G}(z) = -\frac{z}{\alpha} \cdot \frac{c_+ z^{c_+} + \theta c_- z^{c_-}}{z^{c_+} + \theta z^{c_-}}$$

where $\theta = B_+/B_-$. We can now re-write this as

$$\mathcal{G}(z) = -\frac{z}{\alpha} \cdot \frac{wc_+ z^{c_+} + (1-w)c_- z^{c_-}}{wz^{c_+} + (1-w)z^{c_-}}, \quad \forall z \in (0, z_0), \tag{B-8}$$

where $w = \frac{1}{1+\theta} \in (-\infty, \infty]$, with $w = \infty$ if $\theta = -1$. Here w is a free parameter used to match the "terminal condition" $\mathcal{G}(z_0) = G(z_0)$. We claim that $w \neq 1$. If not, then we see that $\mathcal{G}(z) = -\frac{c_+}{\alpha}z$. Thus, by the monotonicity of $p^2(z)$ (and hence $A(z)$) we have

$$0 > -\frac{c_+}{\alpha} = \mathcal{G}'(z_0) = A(2z_0) + \frac{\alpha}{z_0^2}\mathcal{G}^2(z_0) > A(z_0) + \frac{\alpha}{z_0^2}G^2(z_0) = G'(z_0) > 0. \tag{B-9}$$

This is a contradiction; hence $w \neq 1$.

Recall that $\mathcal{G}(z_0) = G(z_0)$, so

$$\mathcal{G}'(z_0) = A(2z_0) + \frac{\alpha}{z_0^2}\mathcal{G}^2(z_0) > A(z_0) + \frac{\alpha}{z_0^2}G^2(z_0) = G'(z_0)$$

for $z \in (0, z_0)$ and hence $G(z) > \mathcal{G}(z)$ for $z \in (0, z_0)$. Thus (for $w \neq 1$) $z^{c-} \gg z^{c+}$ as $z \to 0$ so we see that

$$\limsup_{z \downarrow 0} \frac{G(z)}{z} \geq \limsup_{z \downarrow 0} \frac{\mathcal{G}(z)}{z} = -\frac{c_-}{\alpha} \equiv -\frac{1}{\alpha}c_-(\lambda, \frac{1}{2z_0}). \tag{B-10}$$

However, using (B-3) we know that this is contradictory to the upper bound in (B-1). Thus, in conclusion, we must have that $G'(z) \leq 0$ for sufficiently small $z > 0$. This proves the upper bound in (13).

C Proof of proposition A.1

Throughout we assume that $\beta \in (0, \beta^*]$. Recall that we have defined

$$\beta^* = \frac{\mu^2}{2\alpha\sigma^2}, \quad \hat{y} = \frac{1}{\sqrt{2\lambda\alpha\sigma^2}}, \quad \bar{Y} = \frac{\mu}{\alpha\sigma^2},$$

and that the ODE for $q(y)$ in the linear price impact case is given by:

$$q'(y) = \frac{1}{\sigma^2 y^2}\big[\big(\frac{1}{2\lambda} - \alpha\sigma^2 y^2\big)q(y)^2 - 2y(\mu - \alpha\sigma^2 y)q(y) + (2\mu y - \alpha\sigma^2 y^2 - 2\beta)\big] := f(y, q(y)). \tag{C-1}$$

We calculate the zero contour for f (ignoring the positive pre-factor $1/(\sigma^2 y^2)$):

$$\begin{aligned} 0 =& \big(\frac{1}{2\lambda} - \alpha\sigma^2 y^2\big)q^2 - 2y(\mu - \alpha\sigma^2 y)q + (2\mu y - \alpha\sigma^2 y^2 - 2\beta) \\ =& \alpha\sigma^2\left((\hat{y}^2 - y^2)q^2 - 2y(\bar{Y} - y)q + 2\bar{Y}y - y^2 - \frac{2\beta}{\alpha\sigma^2}\right) \end{aligned}$$

Then for any y such that $y \neq \hat{y}$, we have that the following functions for the 0-contours:

$$h_\pm(y) = \frac{2y(\mu - \alpha\sigma^2 y) \pm 2\sqrt{\Delta(y)}}{\frac{1}{\lambda} - 2\alpha\sigma^2 y^2} \equiv \frac{y(\bar{Y} - y) \pm \sqrt{\Delta(y)}/(\alpha\sigma^2)}{\hat{y}^2 - y^2}, \quad \forall y \in (0, \hat{y}) \cup (\hat{y}, \infty) \tag{C-2}$$

where

$$\Delta(y) = 2\alpha\sigma^2(\beta^* - \beta + \frac{1}{4\lambda})y^2 - \frac{\mu}{\lambda}y + \frac{\beta}{\lambda}. \tag{C-3}$$

We notice that for all $\beta \in (0, \beta^*]$ and $\lambda > 0$, $h_\pm(0+) = \pm 2\sqrt{\beta\lambda}$. Moreover, the coefficient of y^2 in $\Delta(y)$ is positive: Furthermore, straightforward calculation yields that $\Delta(y) < 0$ if and only if $y \in (y_-, y_+)$, where

$$y_\pm = \frac{\bar{Y} \pm \sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)}}{1 + 4(\beta^* - \beta)\lambda}, \tag{C-4}$$

where we have made the following assumption:

Assumption C.1 *We assume the following regarding λ henceforth,*

$$0 < \lambda \leq \frac{\alpha\sigma^2}{8\mu^2} := \frac{1}{16\beta^*} \Leftrightarrow 2\bar{Y} \leq \hat{y}. \tag{C-5}$$

Remark C.1 For any $\beta \in (0, \beta^*]$ and λ satisfying Assumption C.1, we have

1. $1 - 4\lambda\beta \geq 1 - 4\lambda\beta^* \geq 1 - \frac{1}{4} > 0$.

2. From
$$y_+ := \frac{\bar{Y} + \sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)}}{1 + 4\lambda(\beta^* - \beta)} < \bar{Y} + \sqrt{\frac{2}{\alpha\sigma^2}\beta^*} = 2\bar{Y} \leq \hat{y}$$
we know that $y_- \leq y_+ < 2\bar{Y} \leq \hat{y}$. Moreover,
$$y_+ - \bar{Y} = \frac{\sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)} - \bar{Y}4\lambda(\beta^* - \beta)}{1 + 4\lambda(\beta^* - \beta)} = \frac{\sqrt{(1 - \frac{\beta}{\beta^*})(1 - 4\lambda\beta)} - 4\lambda(\beta^* - \beta)}{1 + 4\lambda(\beta^* - \beta)}\bar{Y}$$
Notice that the denominator and $\bar{Y}$ are positive, and the pre-factor in the numerator is greater than
$$\sqrt{(1 - \frac{\beta}{\beta^*})(1 - \frac{1}{4})} - 4\lambda(\beta^* - \beta) =: F(\beta),$$
which is a concave function of β. Since $F(\beta^*) = 0$ and $F(0) = \frac{\sqrt{3}}{2} - 4\lambda\beta^* \geq \frac{\sqrt{3}}{2} - \frac{1}{4} > 0$, we may conclude that $F(\beta) \geq 0$ for all $\beta \in (0, \beta^*]$. Hence,
$$y_+ \geq \bar{Y},$$
where the equality holds if and only if $\beta = \beta^*$. Similarly, from
$$y_- := \frac{\bar{Y} - \sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)}}{1 + 4\lambda(\beta^* - \beta)} \leq \frac{\bar{Y}}{1 + 4\lambda(\beta^* - \beta)} \leq \bar{Y}$$
we know that $y_- \leq \hat{Y}$, with equality if and only if $\beta = \beta^*$.

3. From
$$\bar{Y} - \sqrt{\frac{2}{\alpha\sigma^2}(\beta^* - \beta)(1 - 4\lambda\beta)} > \bar{Y} - \sqrt{\frac{2}{\alpha\sigma^2}\beta^*} = \bar{Y} - \bar{Y} = 0$$
we know that $y_- > 0$.

Remark C.2 From the sign change of the denominator in (C-2), we see that for all $y \in (0, y_-)$, we have $h_-(y) < h_+(y)$; and for all $y \in (y_+, \hat{y})$, we have $h_-(y) < h_+(y)$; for all $y \in (\hat{y}, \infty)$, we have $h_+(y) < h_-(y)$.

We now give the proof of Proposition A.1:

Proof. (of Proposition A.1). We only prove the statement for $h_-(y)$. That for $h_+(y)$ can be proved in the same manner. We begin by applying implicit function theorem to $y^2\sigma^2 f(y, q)$ to obtain that
$$h'_-(y) = \frac{[y(q-1) + \bar{Y}](q-1)}{q(\hat{y}^2 - y^2) - y(\bar{Y} - y)}\Big|_{q=h_-(y)}, \quad \forall y \in (0, y_-) \cup (y_+, \hat{y}) \cup (\hat{y}, \infty). \tag{C-6}$$
Recall that
$$h_-(y) = \frac{y(\bar{Y} - y) - \sqrt{\Delta(y)}/(\alpha\sigma^2)}{\hat{y}^2 - y^2}, \tag{C-7}$$
By conditioning on whether $y > \hat{y}$, we know from (C-7) that
$$\begin{cases} h_-(y) < y(\bar{Y} - y)/(\hat{y}^2 - y^2), & \forall y \in (0, y_-) \cup (y_+, \hat{y}), \\ h_-(y) > y(\bar{Y} - y)/(\hat{y}^2 - y^2), & \forall y \in (\hat{y}, \infty). \end{cases}$$
This means that the denominator of $h_-(y)$ in (C-6) satisfies
$$h_-(y)(\hat{y}^2 - y^2) - y(\bar{Y} - y) < 0, \quad \forall y \in (0, y_-) \cup (y_+, \hat{y}) \cup (\hat{y}, \infty). \tag{C-8}$$
We now prove that the numerator in (C-6) is negative over $(0, y_-)$ and is positive over $(y_+, \hat{y}) \cup (\hat{y}, \infty)$. To that end, we notice that
$$h_-(y) - 1 = \frac{\bar{Y}y - \hat{y}^2 - \sqrt{\Delta(y)}/(\alpha\sigma^2)}{\hat{y}^2 - y^2} < 0 \quad \forall y \in (0, y_-) \cup (y_+, \hat{y}), \tag{C-9}$$

because $0 < \bar{Y} < \hat{y}$ and $0 < y < \hat{y}$. Moreover,

$$y(h_-(y)-1)+\bar{Y} = \frac{\hat{y}^2(\bar{Y}-y) - y\sqrt{\Delta(y)}/(\alpha\sigma^2)}{\hat{y}^2-y^2} =: \frac{k(y)}{\hat{y}^2-y^2}. \tag{C-10}$$

For $y \in (0, y_-) \cup (y_+, \hat{y})$, the sign of (C-10) depends on that of its numerator $k(y)$ only. We first treat $y \in (0, y_-)$, in which case we bound $k(y)$ from below, by bounding $\Delta(y)$ from above. To that end, recall that

$$\begin{aligned}\Delta(y) =& 2\alpha\sigma^2(\beta^* - \beta + \frac{1}{4\lambda})y^2 - \frac{\mu}{\lambda}y + \frac{\beta}{\lambda} = 2\alpha\sigma^2(\beta^* + \frac{1}{4\lambda})y^2 - \frac{\mu}{\lambda}y + 2\alpha\sigma^2\beta(\frac{1}{2\alpha\sigma^2\lambda} - y^2)\\ =& 2\alpha\sigma^2(\beta^* + \frac{1}{4\lambda})y^2 - \frac{\mu}{\lambda}y + 2\alpha\sigma^2\beta(\hat{y}^2 - y^2).\end{aligned} \tag{C-11}$$

Since $0 < y < y_- < \hat{y}$, (C-11) implies that, for any fixed $y \in (0, y_-)$, $\Delta(y)$ is increasing in β. Hence, for all $\beta \in (0, \beta^*]$,

$$\Delta(y) \leq 2\alpha\sigma^2(\frac{1}{4\lambda})y^2 - \frac{\mu}{\lambda}y + \frac{\beta^*}{\lambda} = \frac{\alpha\sigma^2}{2\lambda}(y^2 - 2\frac{\mu}{\alpha\sigma^2}y + \frac{2\beta^*}{\alpha\sigma^2}) = \frac{\alpha\sigma^2}{2\lambda}(y^2 - 2\bar{Y}y + \bar{Y}^2) = \frac{\alpha\sigma^2}{2\lambda}(\bar{Y}-y)^2. \tag{C-12}$$

where the first ineqality follows because its only an equality for $\beta = \beta^*$. By (C-12) we know that, for all $y < y_- \leq \bar{Y}$,

$$\begin{aligned}k(y) \geq& \hat{y}^2(\bar{Y}-y) - \frac{y}{\alpha\sigma^2}\sqrt{\frac{\alpha\sigma^2}{2\lambda}(\bar{Y}-y)^2} = \hat{y}^2(\bar{Y}-y) - \frac{y}{\sqrt{2\alpha\sigma^2\lambda}}(\bar{Y}-y)\\ =& \hat{y}^2(\bar{Y}-y) - \hat{y}y(\bar{Y}-y) = \hat{y}(\bar{Y}-y)(\hat{y}-y) > 0.\end{aligned} \tag{C-13}$$

(C-13) tells us that (C-10) is positive and hence $k(y) > 0$ and $y(h_-(y)-1)+\bar{Y} > 0$ (i.e. the left hand side of (C-10)) on $(0, y_)$. Thus the numerator of (C-6), i.e. $[y(h_-(y)-1)+\bar{Y}](h_-(y)-1)$ is negative, which implies that $h'_-(y) > 0$ for all $y \in (0, y_-)$ (see (C-6) and (C-8)).

For the case that $y \in (y_+, \hat{y})$, we notice that

$$k(y) = \hat{y}^2(\bar{Y}-y) - y\sqrt{\Delta(y)}/(\alpha\sigma^2) < \hat{y}^2(\bar{Y}-y) < 0, \tag{C-14}$$

because $\bar{Y} \leq y_+ < y$. From (C-9), (C-10), and (C-14), we know that the numerator in (C-6) is positive in this domain, which implies that $h'_-(y) < 0$ for all $y \in (0, y_-)$ (see (C-6) and (C-8)).

We now treat the monotonicity of $h_-(y)$ for $y \in (\hat{y}, \infty)$. We begin by using (C-2) to obtain that

$$\lim_{y\to\infty} h_-(y) = 1.$$

Moreover, for any fixed $\beta \in (0, \beta^*], \lambda \in (0, 1/(16\beta^*)]$ and all sufficiently large $y > 0$, it can be seen through asymptotics that $h'_-(y) < 0$ for large y. But it is easily seen that

$$(y(q-1)+\bar{Y})(q-1) > 0, \quad \forall y \in (0,\infty), q \in (1,\infty).$$

Hence, we may conclude from (C-6) and (C-8) that $h'_-(y) < 0$ for all $y \in (\hat{y}, \infty)$. ■