

Option Hedging with Smooth Market Impact

Tianhui Michael Li* and Robert Almgren[†]

April 8, 2016

Abstract

We consider intraday hedging of an option position, for a large trader who experiences temporary and permanent market impact. We formulate the general model including overnight risk, and solve explicitly in two cases which we believe are representative. The first case is an option with approximately constant gamma: the optimal hedge trades smoothly towards the classical Black-Scholes delta, with trading intensity proportional to instantaneous mishedge and inversely proportional to illiquidity. The second case is an arbitrary nonlinear option structure but with no permanent impact: the optimal hedge trades toward a value offset from the Black-Scholes delta. We estimate the effects produced on the public markets if a large collection of traders all hedge similar positions. We construct a stable hedge strategy with discrete time steps.

Contents

1	Introduction	2
2	Problem Setup	5
2.1	Market Impact Model	5
2.2	European Option	6
2.3	Wealth dynamics	7
2.4	Mean-variance optimization	8
2.5	Overnight risk	9
3	Solutions and Effects	10
3.1	HJB equation	11
3.2	Constant Gamma approximation	11
3.3	Effect on the price process	14
3.4	No permanent impact, general Gamma	16
3.5	Discrete time	18
4	Conclusion	20

*Princeton Bendheim Center for Finance and Operations Research and Financial Engineering. Research supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE-0646086 and a Fannie and John Hertz Foundation Graduate Fellowship.

[†]Quantitative Brokers and NYU Courant Institute. a1mgren@cims.nyu.edu

1 Introduction

Dynamic hedging of an option position is one of the most studied problems in quantitative finance. But when the position size is large, the optimal hedge strategy must take account of the transaction costs that will be incurred by following the Black-Scholes solution. This large position may be the position of a single large trader, or may be the aggregate position of a collection of traders, for example the entire sell-side community, who all hold similar positions, and all of whom hedge while their counterparties do not. In addition to private costs, hedging activity by one or more large position holders may have observable effects on the public markets.

On the morning of July 19, 2012, an unusual “sawtooth” pattern was observed in US equity markets [Hwang et al., 2012]. Four large stocks exhibited substantial price swings on a regular half-hour schedule. Each stock hit a local minimum price on each hour, and a local maximum on each half-hour (Figure 1). No significant news was released on this day, but CBOE options expiration was the next day. The most plausible explanation [Lehalle and Lasnier, 2012] is that these patterns were the result of a delta-hedging strategy executed by a large options position holder with no regard for market impact. Each half-hour, he or she evaluated the necessary trade to obtain a delta-neutral position, and executed this trade across the next half-hour. Market impact caused the price to move, and at the next evaluation at the new price, the position was partially reversed. The action was similar to a forward Euler discretization of an ordinary differential equation [Ascher and Petzold, 1998] which can introduce instability into a stable problem.

As an additional example, on Oct. 15, 2014, the US Treasury market underwent the largest intraday move since 2009 [Almgren, 2015]. Devasabai [2014] cites numerous market participants who attribute the market instability in part to hedging of short option positions held by the dealer community, that were large in aggregate if not individually: “ ‘These things don’t happen unless there is a big short gamma position,’ says a senior fixed-income portfolio manager at a firm in New York.” This paper explains how this hedging activity can increase market volatility. (Such collective effects are likely significant primarily in special market conditions; for example, Kambhu [1998] estimated hedging effects to be small for interest rate options, including over-the-counter positions.)

We use a simple market impact model like those used for optimal execution to study the hedging problem faced by a large investor. The optimal hedge strategy depends on a balance between the risk of mishedge and the cost of temporary market impact. Permanent market impact causes an increase or decrease in realised volatility of the public market price depending on whether the large investor, or the entire community of hedging traders, are net long or short options. A simplistic implementation of the hedge strategy in discrete time intervals can lead to the behavior shown in Figure 1, but in Section 3.5 we show how to stably execute such hedging using an implicit discretization.

There is a substantial literature on the effect of transaction costs on Black-Scholes hedging. Leland [1985] introduced a discrete-time model in which the trading in each time interval affects the market price at the next interval. With suitable dependence of the market impact on the time interval, he was able to obtain a preference-free option price calculated using a modified implied volatility. Subsequent work [Kabanov and Safarian, 1997, Zhao and Ziemia, 2007] has clarified some aspects of Leland’s model, but the

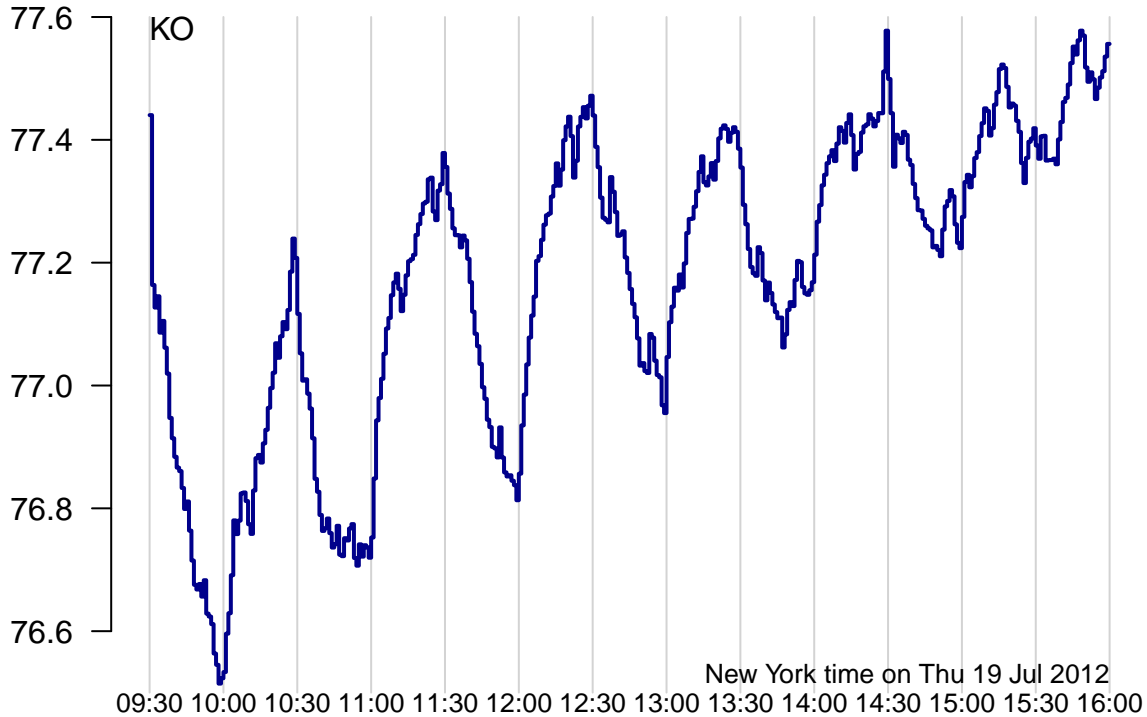


Figure 1: Price swings for Coca-Cola (KO) on July 19, 2012, one day before options expiration. Similar swings were observed in Apple (AAPL), IBM (IBM), and McDonald's (MCD). These were likely caused by a naïve hedge strategy for a large options position.

model is essentially derived as a limit of time-discrete hedging rather than continuous-time Brownian motion [Kabanov and Safarian, 2009].

More recent literature is interested in super-replication [Çetin et al., 2010, Soner et al., 1995]. We relax this requirement by having a finite mishedge penalty. Our paper is more closely linked to these using a utility-based framework [Cvitanic and Wang, 2001].

Transaction costs themselves have been modeled via various mechanisms. A large strand of the literature models trading frictions as a cost proportional to trade size, typically interpreted as arising from the bid-ask spread. This branch of the literature uses singular control and the optimal solution is typically in the form of a tracking band [Davis and Norman, 1990, Shreve and Soner, 1994]. As the portfolio exits this band, the trader makes singular corrections to his holdings to keep it strictly within the limits of the band. In the upper panel of Figure 2, we illustrate this hedging strategy.

Our simple market impact model is phenomenological and not directly based in the details of microstructure. Following Almgren and Chriss [2000], we decompose price impact into temporary and permanent price impact. We can think of the temporary impact as connected to the liquidity cost faced by the agent while the permanent impact as linked to information transmitted to the market by the agent's trades. The temporary impact depends on the *rate* of execution, while the permanent impact depends on the total number of shares executed. Under this model, the optimal solution is to trade aggressively

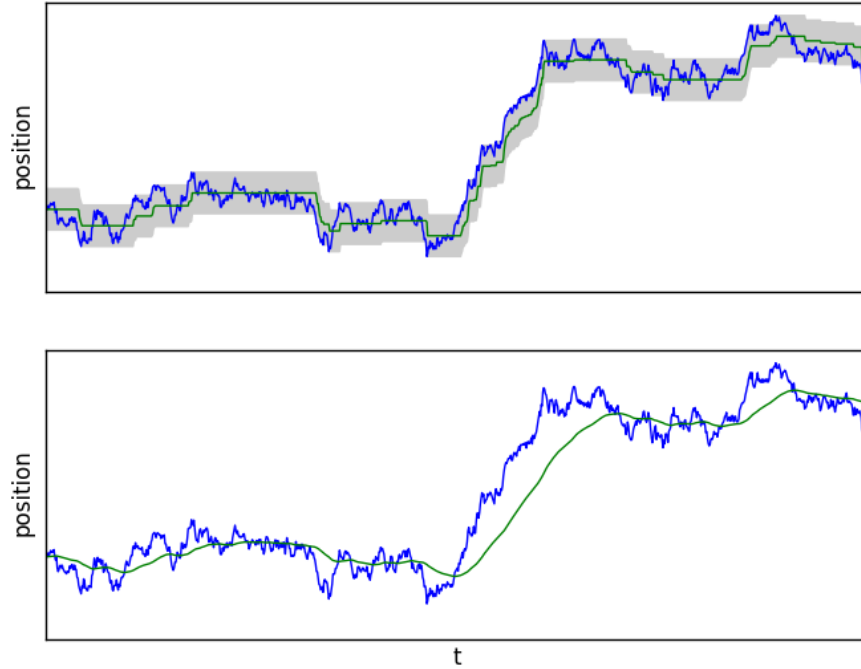


Figure 2: Comparison of the proportional-transaction-cost fixed-tracking-band strategy (top) and our dynamic strategy (bottom). In the former, our (green) trading position changes only when the (blue) target leaves the (gray) trading-band. In the latter, our (green) trading position smoothly adjusts to the same (blue) target. Compare the smooth trading flow and position of the latter strategy to the abrupt trading of the former.

towards being hedged, taking account both the available liquidity and the degree of the mishedge. Our trading strategy is smooth: we approximate the impact-free Black-Scholes Delta, an infinite variation process, with trading positions that are differentiable. In Figure 2 we compare our strategy to the strategy using a tracking band.

Our solution is similar to that of Gârleanu and Pedersen [2013], who solve the infinite-horizon “Merton Problem” under only temporary market-impact assumptions. As in our setup, they use a linear-quadratic objective rather than the traditional expected utility setup. They find that trading intensity at time t is given by

$$\theta_t = -\kappa h \cdot (X_t - \Delta_t),$$

where X_t is the number of shares, κ is an urgency parameter with units of inverse time, $h > 0$ is a dimensionless constant of proportionality, and the “target portfolio” Δ_t is related to the frictionless Merton-optimal portfolio. The intensity of trading θ_t is proportional to the distance between the current holdings and target, and is inversely proportional to the square root of the illiquidity coefficient.

Rogers and Singh [2010] obtain a similar solution for option hedging, with temporary impact but no permanent impact, in which the coefficient h depends on time to expiration. In their model, as in ours, the role of the target portfolio is played by the Black-Scholes delta Δ_t . Without permanent impact, they do not obtain any effect of the hedging on the public markets as in Figure 1.

Lions and Lasry [2006, 2007] have studied the effect on volatility of hedging by a large options trader. In our language, they include permanent impact but not temporary. Thus the trader’s position is always perfectly hedged, but there is an observable effect on the realised volatility. In our model, this modified volatility appears on time scales longer than the hedge scale, which is controlled by risk aversion and temporary impact. A key feature of our model is this interplay between temporary and permanent price impact.

Guéant and Pu [2015] have solved a model very similar to ours, including both temporary and permanent price impact. They use a utility function rather than our mean-variance optimization. They distinguish between cash settlement and physical delivery, whereas our intraday model assumes the position will be marked to market the morning, so delivery does not appear. Their focus is on the modifications to the option value introduced by market impact, whereas we are most interested in the qualitative behavior of the hedge strategy, and the likely effects to be observed in the public market. We concentrate on a few special cases in which exact solutions can give insight, whereas they undertake numerical solution in more general cases. The general behavior of their solutions is very similar to ours.

Avellaneda and Lipkin [2003] have studied “stock pinning:” the tendency of the underlying asset price to approach an option strike price at expiration. They use a model similar to ours, but their analysis is based on a local approximation near expiration at the money. Jeannin et al. [2008] have performed a more refined analysis, and determine a modified volatility coefficient near expiration.

In Section 2 below, we motivate our assumptions and formally set up the problem. In Section 3.1 we present the general solution approach. We solve this problem for a quadratic option value in 3.2, and in 3.3 we consider the impact of hedging on the price process. In 3.4 we consider the special case of no permanent impact, which we can solve for general option structure. In 3.5, we give a stable discrete-time solution, that avoids the problems shown in Figure 1. Finally, in Section 4 we summarize and suggest possible future empirical work.

2 Problem Setup

We first present our market model including both temporary and permanent impact. We then discuss hedging of a European option, and present our objective function possibly including overnight risk.

2.1 Market Impact Model

Let X_t be the number of shares held by the agent at time $t \geq 0$. The fundamental price at t is given by

$$P_t = P_0 + \nu(X_t - X_0) + \sigma W_t \tag{1}$$

where $\nu > 0$ is the coefficient of permanent impact, $\sigma > 0$ is the absolute volatility of the fair value, and W_t is a standard Brownian motion with filtration \mathcal{F}_t . We use arithmetic Brownian motion rather than geometric: the difference is negligible over the intraday time horizons considered in this paper and this leads to dramatic simplifications. We neglect interest rates and dividends. Lions and Lasry [2006, 2007] considered option hedging for a large trader who faces this form of permanent impact, but with no temporary impact.

Under this model, if a position of size X_0 shares with initial market price P_0 is fully liquidated, the expected value of the resulting cash will be $\int_0^X (P_0 - \nu x) dx = X_0 P_0 - \frac{1}{2} \nu X_0^2$, independently of the time taken or strategy used to execute the liquidation. That is, we cannot avoid a cumulative impact cost of $\frac{1}{2} \nu X_0^2$, quadratic in portfolio size. Nonetheless, we shall assume that such a position is marked to market at value $X_0 P_0$.

We denote by θ_t the instantaneous intensity of trading, so that

$$X_t = X_0 + \int_0^t \theta_u du.$$

Implicit in this formulation is the assumption that the trade rate θ_t is almost everywhere defined and bounded (see (8) below), and hence that the portfolio position X_t is differentiable. As in Almgren and Chriss [2000], trading at instantaneous rate θ_t requires payment of a price premium linear in θ_t . That is, the effective trade price is

$$\tilde{P}_t(\theta_t) = P_t + \eta \theta_t. \quad (2)$$

where $\eta > 0$ is the coefficient of temporary impact. Nonlinear temporary impact functions are more consistent with empirical data [Almgren et al., 2005] but would complicate our analysis; they have been used for stock pinning by Avellaneda et al. [2012]. Rogers and Singh [2010] considered option hedging with this form of temporary impact, but with no permanent impact.

2.2 European Option

We hedge a European contingent claim (option) over a time period $[0, T]$. We shall think of the time T as the end of the trading day, when the trader's position is marked to market. Thus we consider intraday hedging. Let $g(t, p)$ denote the value of the option for a small trader in a complete market whose execution has no price impact. This trader's total portfolio value at time t is $X_t P_t + g(t, P_t)$.

If the option value at $t = T$ is specified as $g_0(p)$, then a Black-Scholes hedging argument gives the option value for $t < T$, neglecting market impact, as the solution of the partial differential equation

$$\dot{g}(t, p) + \frac{1}{2} \sigma^2 g''(t, p) = 0 \quad \text{for } t, p \in [0, T) \times \mathbb{R} \quad \text{and} \quad g(T, p) = g_0(p). \quad (3)$$

Here \dot{g} is derivative with respect to t and g' and g'' are derivatives with respect to p .

We identify the delta and gamma

$$\Delta(t, p) = -g'(t, p), \quad \Gamma(t, p) = g''(t, p).$$

The negative sign on $\Delta(t, p)$ is because our trader is long the payoff, rather than hedging a short position; a portfolio perfectly hedged against small price fluctuations will have $X_t = \Delta(t, P_t)$. A trader who is long a call option or short a put will have $\Delta < 0$; one who is short a call or long a put will have $\Delta > 0$. A trader who is long a put or call option will have $\Gamma > 0$; a trader who is short will have $\Gamma < 0$. Thus Γ reflects both the sign and size of the trader's net option position. More generally, for the purposes of modeling the effect of permanent price impact, Γ should be thought of as the total position (sign and size) of all traders who are hedging their option positions (see Section 3.3).

We assume that the option is such that

$$\Gamma(t, p) = g''(t, p) \text{ is uniformly bounded above and below, and} \quad (4)$$

$$\text{is Lipschitz in } p \text{ with a constant that is independent of } t.$$

This assumption holds for most options except at expiry, and so this formula will be valid for intraday hedging except on the expiration day.

We also assume that permanent impact ν is small enough that

$$\text{There is a constant } G > 0 \text{ with } 1 + \nu\Gamma(t, p) \geq G \text{ for all } (t, p). \quad (5)$$

For an option contract having $\Gamma < 0$ and a fixed value of ν , condition (5) can always be violated by scaling up the position size. Thus our model captures realistic behavior in an intermediate range, where the position is large enough to have some effect on the market, but not so large that option hedging overwhelms the intrinsic market dynamics. We expect $|\nu\Gamma| \ll 1$, so option hedging is not the primary factor driving the price process.

Condition (5) is necessary so that hedging does not lead to unbounded price motions; Figure 1 illustrates a situation in which this condition is close to being violated. To understand this, suppose that at a current asset price of P_t , we are imperfectly hedged so that $X_t \neq \Delta(t, P_t)$. Suppose that we ignore temporary impact, so we would be able to execute an instantaneous trade to an arbitrary new position \hat{X}_t . With permanent impact, the price following this trade will be $\hat{P}_t = P_t + \nu(\hat{X}_t - X_t)$. In order to be hedged against small fluctuations, we want \hat{X}_t to be such that $\hat{X}_t = \Delta(t, \hat{P}_t)$, or

$$\hat{X}_t + \int_{P_t}^{P_t + \nu(\hat{X}_t - X_t)} \Gamma(t, p) dp = \Delta(t, P_t).$$

The right side of this equation is independent of \hat{X}_t . Denoting the left side by $F(\hat{X}_t)$, condition (5) says that $F'(\hat{X}_t) = 1 + \nu\Gamma \geq G > 0$ everywhere, so the condition $F(\hat{X}_t) = \Delta(t, P_t)$ always has a solution and there always exists a unique optimal hedge portfolio. If ν or $\Gamma < 0$ is large enough so that (5) is violated, then according to this model the price will recede to $\pm\infty$ as we attempt to hedge; in reality, of course other effects will intervene.

This reproduces behavior familiar to practitioners, that for a long position $\Gamma > 0$, our trading moves the price toward our hedge target and hedging is easy, while for a short position $\Gamma < 0$, our own trading pushes the price away from us and hedging is hard. Condition (5) says that it is not infinitely hard.

2.3 Wealth dynamics

We assume that our trader's position is marked to market using the Black-Scholes option value, as well as the book value for the underlying shares, ignoring market impact that

would be incurred in converting these positions into cash. This could be the case, for example, because of institutional rules. Thus we define the initial and terminal wealth

$$\begin{aligned} R_0 &= g(0, P_0) + X_0 P_0 \\ R_T &= g(T, P_T) + X_T P_T - \int_0^T \tilde{P}_t(\theta_t) \theta_t dt \end{aligned} \quad (6)$$

where the last term in (6) denotes the capital spent or gained from trading. Integrating by parts, using the stock dynamics (1), the total temporary impact (2), and Feynman-Kac (3), these quantities are related by

$$\begin{aligned} R_T &= R_0 + \int_0^T Y_t dP_t - \eta \int_0^T \theta_t^2 dt \\ &= R_0 + \sigma \int_0^T Y_t dW_t + \nu \int_0^T Y_t \theta_t dt - \eta \int_0^T \theta_t^2 dt, \end{aligned} \quad (7)$$

in which the portfolio's instantaneous net delta exposure is

$$Y_t = X_t - \Delta(t, P_t) = X_t + g'(t, P_t).$$

The final wealth R_T is the sum of the fluctuation during the trading day and the liquidity cost from permanent and temporary impacts. For a perfectly hedged portfolio, $Y_t = 0$. But since $g'(t, P_t)$ is typically of infinite variation in t while X_t must be differentiable, perfect hedging is impossible.

We assume that the control $\theta_t \in \Theta$, with

$$\Theta = \left\{ \theta \text{ predictable, with } \mathbb{E} \int_0^T \theta_s^2 ds < \infty \text{ and } \theta_t \leq C(1 + |Y_t|) \text{ a.s. for all } t \right\}. \quad (8)$$

The state variables have dynamics (an Itô term in \dot{g}' cancels using (3))

$$\begin{aligned} dP_t &= \nu \theta_t dt + \sigma dW_t \\ dY_t &= (1 + \nu \Gamma(t, P_t)) \theta_t dt + \sigma \Gamma(t, P_t) dW_t \end{aligned} \quad (9, 10)$$

where the same Brownian process W_t appears in both. It is easy to see that P_t is a continuous semimartingale and hence predictable. With (4), the system (9,10) has a strong solution (P_t, Y_t) for $\theta_t \in \Theta$ [Touzi, 2013, sect. 2.1].

Each share purchased ($\theta_t dt$) pushes P_t up by ν because of the permanent impact, and also increases the net delta position Y_t by 1, for the increase in the stock position, plus $\nu \Gamma(t, P_t)$, for the effect of permanent price impact effect on the underlying price and hence the option's delta.

It only remains to specify the trader's objective function.

2.4 Mean-variance optimization

The trader chooses the strategy θ_t to maximise the final wealth R_T (6). With no market impact, we may maintain $Y_t = 0$, eliminating all randomness in R_T due to market fluctuations dW_t , and the zero-risk solution defines the option value. With market impact,

perfect hedging is impossible and we must maximise the expected value of R_T , while also minimizing its uncertainty.

We use a mean-variance criterion rather than a utility function. Although mean-variance optimization occasionally can have unexpected properties, it is extremely straightforward and familiar to practitioners. The expected value of R_T is

$$\mathbb{E}R_T = R_0 + \nu \mathbb{E} \int_0^T Y_t \theta_t dt - \eta \mathbb{E} \int_0^T \theta_t^2 dt.$$

The variance of R_T is complicated, since all terms in (6) are random and dependent. But a reasonable approximation is that the largest source of uncertainty is the price motions. The terms involving market impact are important because they have consistent sign, but their variances are small compared with market dynamics. This is the “small-portfolio” approximation of Lorenz and Almgren [2011] and further explored by Tse et al. [2013]. Effectively, it is an analog for temporary impact to condition (5) for permanent impact. Thus we make the approximation

$$\text{Var } R_t \approx \sigma^2 \text{Var} \int_0^T Y_t dW_t = \sigma^2 \mathbb{E} \int_0^T Y_t^2 dt.$$

We introduce a variance penalty $\lambda > 0$ (this λ is half the risk aversion parameter a of an exponential utility function $\exp(-aR_T)$ if R_T had a normal distribution), and define our mean-variance objective function to be the approximate version of $\lambda \text{Var } R_T - \mathbb{E}R_T$. If the position is marked to market at time T , so that overnight risk does not enter, then this preliminary version of our problem is

$$\inf_{\theta \in \Theta} \mathbb{E} \left[\lambda \sigma^2 \int_0^T Y_t^2 dt - \nu \int_0^T Y_t \theta_t dt + \eta \int_0^T \theta_t^2 dt \right].$$

More realistically, overnight risk should be included as we see in the next section.

2.5 Overnight risk

In a more realistic version of the model, the position is marked to market at tomorrow’s open T_* rather than today’s close T . Hence market risk is incurred for $T < t < T_*$, although trading is not possible during the night so thus $X_{T_*} = X_T$. The trader must choose the close position X_T to hedge this overnight risk as much as possible. We do not necessarily assume that the overnight price process is arithmetic Brownian motion (1).

As in (6),

$$\begin{aligned} R_{T_*} &= g(T_*, P_{T_*}) + X_T P_{T_*} - \int_0^T \tilde{P}_t(\theta_t) \theta_t dt \\ &= R_T + g(T_*, P_{T_*}) - g(T, P_T) + X_T (P_{T_*} - P_T) \\ &= R_T + \int_T^{T_*} [g'(t, P_t) - g'(T, P_T)] dP_t + Y_T (P_{T_*} - P_T) \\ &= R_T - \xi + Y_T \Pi \end{aligned}$$

with

$$\xi = \int_T^{T_*} (\Delta(t, P_t) - \Delta(T, P_T)) dP_t \quad \text{and} \quad \Pi = P_{T_*} - P_T.$$

Since each of ξ and Π is an Itô integral involving price motions dP_t only for $t > T$, each has mean zero and has zero correlation with any variable depending on P_t for $t \leq T$, such as Y_T and R_T (though ξ is not necessarily independent of them). Thus overnight risk simply adds an additional variance of $\mathbb{E}[(Y_T \Pi - \xi)^2]$, controlled by the mishedge at market close Y_T , and the optimization problem with overnight risk is

$$\inf_{\theta \in \Theta} \mathbb{E} \left[\lambda (Y_T \Pi - \xi)^2 + \lambda \sigma^2 \int_0^T Y_t^2 dt - \nu \int_0^T Y_t \theta_t dt + \eta \int_0^T \theta_t^2 dt \right]. \quad (11)$$

We denote $\Sigma^2 = \mathbb{E}[\Pi^2 | P_T = p]$ and $\Xi(p)^2 = \mathbb{E}[\xi^2 | P_T = p]$. If the overnight price process is arithmetic Brownian motion, then $\Sigma^2 = \sigma^2(T_* - T)$ independent of P_T . In general, we shall assume that Σ^2 is independent of P_T . Even more generally, we may abstract from the details of the overnight price process to consider any L^2 random variables ξ and Π of mean zero, measurable on \mathcal{F}_T , such that the distribution of ξ depends only on P_T , and the distribution of Π is independent of \mathcal{F}_T .

We shall use (11) as our objective function from now on, since the additional motivation to “close flat” is important in practice.

Example We could take T to be the maturity of the option (ignoring the unboundedness of Γ if P_t is near the strike). In this case, there is no risk beyond expiration and formally ξ and Π would be zero. But the individual trader may choose to incorporate a penalty in order to drive the portfolio toward more precise hedging at expiration.

Example On intraday trading time scales, far from expiration, it is plausible to model the option as having a constant gamma, $\Gamma(t, p) \equiv \Gamma \in \mathbb{R}$ constant, so

$$\Delta(t, P_t) = \Delta(T, P_T) - \Gamma(P_t - P_T). \quad (12)$$

If ξ and Π arise from the Brownian price process continued on $T < t < T_*$, then

$$\xi = -\Gamma \int_T^{T_*} (P_t - P_T) dP_t = -\frac{1}{2} \Gamma (\Pi^2 - \Sigma^2).$$

The overnight terms are uncorrelated, and the corresponding objective term is

$$\mathbb{E}[\lambda (Y_T \Pi - \xi)^2] = \lambda (\Sigma^2 Y_T^2 + \Xi^2). \quad (13)$$

where Ξ as well as Σ is independent of \mathcal{F}_T . For general ξ and Π , the “constant Γ ” assumption on the option value will include the assumption that ξ as well as Π is independent of \mathcal{F}_T and that ξ and Π are uncorrelated with each other.

3 Solutions and Effects

We now use standard techniques of optimal control to identify the partial differential equation satisfied by the value function, and we exhibit solutions in two special cases.

3.1 HJB equation

Let $J(t, p, y)$ denote the optimal value function beginning at time t :

$$J(t, p, y) = \inf_{\theta \in \Theta_t} \mathbb{E} \left[\lambda (Y_T \Pi - \xi)^2 + \lambda \sigma^2 \int_t^T Y_s^2 ds - \nu \int_t^T Y_s \theta_s ds + \eta \int_t^T \theta_s^2 ds \right]$$

where Θ_t denotes the allowable control set θ_s for $t \leq s \leq T$, and the expectation is conditional on initial values $P_t = p$ and $Y_t = y$. The actual share holding X_t does not enter into the trading cost on the remaining time, only the mishedge Y_t . We temporarily assume that $J \in C^{1,2,2}([0, T] \times \mathbb{R} \times \mathbb{R})$. Then from the Martingale Principle of Optimal Control, J must satisfy the HJB equation

$$0 = J_t + \lambda \sigma^2 y^2 + \frac{1}{2} \sigma^2 J_{pp} + \Gamma \sigma^2 J_{py} + \frac{1}{2} \Gamma^2 \sigma^2 J_{yy} + \inf_{\theta \in \mathbb{R}} \left\{ \left[(1 + \nu \Gamma) J_y + \nu J_p - \nu y \right] \theta + \eta \theta^2 \right\},$$

in which subscripts on J denote partial derivatives (subscripts t, T on θ, X, P , etc. continue to denote evaluation at the given time), and $\Gamma = \Gamma(t, p)$. The optimal strategy is

$$\theta = -\frac{1}{2\eta} \left((1 + \nu \Gamma) J_y + \nu J_p - \nu y \right) \quad (14)$$

and hence the value function satisfies

$$0 = J_t + \lambda \sigma^2 y^2 + \frac{1}{2} \sigma^2 J_{pp} + \Gamma \sigma^2 J_{py} + \frac{1}{2} \Gamma^2 \sigma^2 J_{yy} - \frac{1}{4\eta} \left[(1 + \nu \Gamma) J_y + \nu J_p - \nu y \right]^2 \quad (15)$$

with terminal data

$$J(T, p, y) = \lambda \mathbb{E} \left(y \Pi - \xi \right)^2. \quad (16)$$

In the expectation, the distribution of ξ is conditional on $P_T = p$; recall that Π is assumed independent of \mathcal{F}_T .

In the optimal strategy (14), if $\nu = 0$ then we trade so as to move our position y in the direction of decreasing J , with rate controlled by the coefficient η of temporary impact. If $\nu > 0$, then we also take account of the effect of our permanent impact on the stock price, both directly via the term νJ_p , and indirectly via the change in Δ by the term $\nu \Gamma J_y$.

The last term in (14), $\theta = \dots + \nu y / 2\eta$, expresses an arbitrage. Since our position is marked to market via the public price of the underlying, we increase the value of our holdings by trading so as to increase the price if we are long relative to the optimal hedge, and to decrease the price if we are short. This effect is intrinsic in our wealth specification (7), and will be controlled by risk aversion.

We will not solve the equation in full generality but will stick to two major sub-cases.

3.2 Constant Gamma approximation

The most illuminating case is the approximation that Γ is constant, as at the end of Section 2.5. This considerably simplifies the problem by eliminating the dependence on the

state variable P_t , and allows us to exhibit the essential features of local hedging without losing ourselves in complexities due to the global shape of the option price. The problem becomes essentially the well-known stochastic linear regulator with time dependence.

With this assumption, the option's delta varies linearly with the stock price, as in (12), the terminal penalty is as in (13), and Γ is constant in the state dynamics (9,10). Assumption (5) says that the constant value $G = 1 + \nu\Gamma > 0$.

Further, $J(t, p, y) = J(t, y)$ independent of p , since the terminal data does not depend on p and the PDE (15) introduces no p -dependence. We look for a solution quadratic in the mishedge y

$$J(t, y) = \frac{1}{2} A_2(T-t) y^2 + A_0(T-t). \quad (17)$$

To solve (15), $A_0(\tau)$ and $A_2(\tau)$ must satisfy the ordinary differential equations

$$\dot{A}_2 = 2\lambda\sigma^2 - \frac{1}{2\eta} (GA_2 - \nu)^2 \quad (18)$$

$$\dot{A}_0 = \frac{1}{2} \Gamma^2 \sigma^2 A_2. \quad (19)$$

for $\tau \geq 0$, with

$$A_2(0) = 2\lambda\Sigma^2 \quad \text{and} \quad A_0(0) = \lambda\Xi^2.$$

To solve (18), note that the graph of the function of A_2 on the right is a parabola opening downwards, crossing $\dot{A}_2 = 0$ at the critical points

$$A_2^\pm = \frac{1}{G} \left(\nu \pm 2\sqrt{\eta\lambda\sigma^2} \right). \quad (20)$$

For an initial value $A_2(0) > A_2^-$, $A_2(\tau)$ moves monotonically towards the stable point A_2^+ as τ increases: it increases to A_2^+ if $A_2^- < A_2(0) < A_2^+$ and it decreases to A_2^+ if $A_2(0) > A_2^+$. If $A_2(0) < A_2^-$, then $A_2(\tau)$ explodes to $-\infty$ at a finite time $\tau > 0$.

We assume that the initial data is in the stable region: $2\lambda\Sigma^2 > A_2^-$ or

$$\nu < 2\left(\sqrt{\eta\lambda\sigma^2} + G\lambda\Sigma^2\right). \quad (21)$$

This requires that either $\lambda\sigma^2$ or $\lambda\Sigma^2$ be sufficiently large compared to ν . It is in this sense that risk aversion controls the potential arbitrage opportunity introduced by permanent impact and our mark to market formulation, as noted at the end of Section 3.1.

Under assumption (21), $A_2(\tau)$ and hence $A_0(\tau)$ exist for all $\tau \geq 0$, $A_2(\tau)$ is uniformly bounded, $A_0(\tau)$ grows linearly, and $J(t, y)$ exists for all $t \leq T$. Indeed,

$$A_2(\tau) = \frac{1}{G} \left(\nu + 2\eta\kappa h(\kappa G\tau) \right) \quad (22)$$

with the function $h(s)$ given by

$$h(s) = \frac{1-u}{1+u}, \quad u = \frac{1-h_0}{1+h_0} e^{-2s}, \quad (23)$$

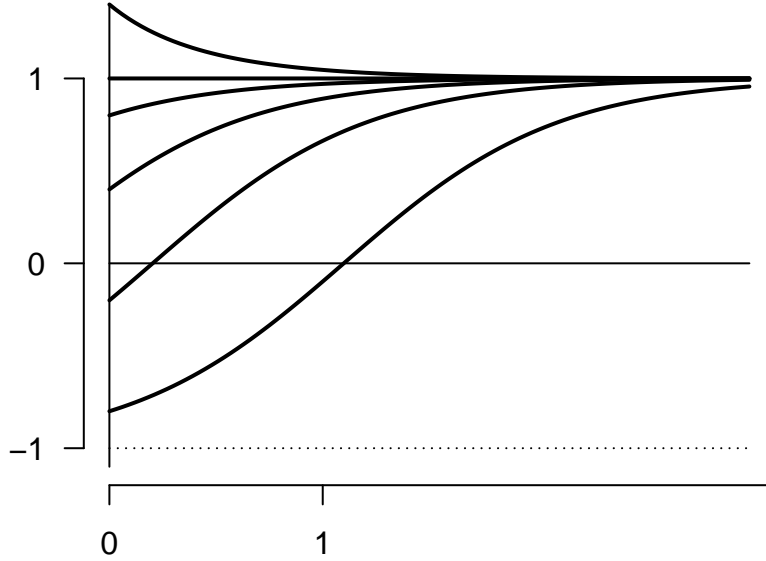


Figure 3: The control coefficient $h(\tau)$. The horizontal axis is scaled time to close of trading $s = \kappa G(T - t) \approx \kappa(T - t)$. The vertical axis is the coefficient h in the linear response $\theta_t = -\kappa h Y_t$. When time to close is greater than $1/\kappa$, the coefficient is 1. As close approaches, the coefficient may increase, decrease, or even become negative, depending on the relative magnitudes of overnight risk and permanent impact.

(see Figure 3), and the constants

$$h_0 = \frac{G\lambda\Sigma^2 - 2\nu}{\eta\kappa} \quad \text{and} \quad \kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}}. \quad (24)$$

Condition (21) assures us that $h_0 > -1$. Also, $h(s) \geq h_0$ for $h_0 \leq 1$, and hence $A_2(\tau) \geq 0$: the value function is convex in the mishedge. The optimal trading intensity (14) is

$$\theta_t = -\kappa h(\kappa G(T - t)) Y_t. \quad (25)$$

This depends only relatively weakly on the value of Γ , via the term $\nu\Gamma$ in $G = 1 + \nu\Gamma$. This argument can v

The agent's trading share target for shares X_t is the Black-Scholes delta hedge $\Delta(t, P_t)$, which varies in time due to price motions caused both by volatility and by his own trading. He constantly trades towards this target but is prevented from holding the exact Black-Scholes delta hedge by the temporary impact stemming from limited liquidity. The trading intensity θ_t is proportional to the degree of mishedge Y_t and the urgency parameter κ . There is a greater penalty to being mishedged with higher underlying volatility σ and risk aversion λ so these parameters increase urgency. Similarly, a more illiquid market (higher η) makes trading more costly, which decreases trading intensity.

Our κ is the same as in Almgren and Chriss [2000] where it was an "urgency parameter" dictating the speed of liquidation as a fraction of the position size. The a/λ in Proposition

5 of Gârleanu and Pedersen [2013] is equivalent to κh in our setup, that is, the higher κ , the faster the agent trades towards the Merton-optimal portfolio.

Far from expiration, where $\kappa(T - t) \gg 1$ (we suppose $G = 1 + \nu\Gamma \approx 1$), $h \equiv 1$ and the trade rule is $\theta_t = -\kappa Y_t$. Near expiration, the solution falls into two cases depending on the value of h_0 , that is, depending on the relative values of terminal risk and market impact.

- If $h_0 > 1$, then overnight risk dominates permanent impact. The trade rate increases as expiration is approached. The trader is willing to pay more in temporary impact costs to reduce the overnight risk of an imperfectly hedged position.
- If $h_0 < 1$, then permanent impact dominates overnight risk. If $h_0 < 0$, then in fact near expiration the trader trades so as to *increase* the mishedge. This is the arbitrage possibility noted in Section 3.1. If $0 < h_0 < 1$, then permanent impact only partially controls the dynamics, and trade rate is reduced but not reversed.

If $[0, T]$ is the trading day, then the case $h_0 > 1$ is the most realistic. Options market-makers typically increase their hedging towards the close of trading to minimize overnight exposure. The trader may choose a lower h_0 as options expiry approaches, selecting $h_0 = 0$ on the day of options expiry. However, this strategy may be deemed too risky and the trader may choose $h_0 > 0$ even at expiry to avoid the risk of being mishedged.

3.3 Effect on the price process

What would be the effect on the publicly observable price process, as the result of hedging by large traders? We first observe that since options are bilateral contracts, each long position has a corresponding short position and conversely. If all position owners hedge their positions, and if all have roughly similar market impact, then there will be no net effect on the price. The only effect will be a net total positive cost from temporary impact, which may be interpreted as premia paid by the hedgers to liquidity suppliers in order to complete their trades.

Presumably at least some market participants are trading the options because they want the options exposure to hedge other risks in their portfolio. Let us identify *buy side* traders as options position holders who do not hedge. *Sell side* traders will be the Wall Street firms who have sold these contracts. The sell side traders have no interest in owning the options exposure and hence will hedge their positions.

We therefore take Γ to be the total exposure of all the sell side traders, that is, of all options position holders who hedge their options exposure. This Γ may be positive or negative depending on whether the “street” is a net buyer or seller. It is not necessarily related to the option open interest. But conversations with market participants indicate that most professional traders are generally aware of the net positions of their counterparts across the industry.

Note that the example shown in Figure 1 shows something slightly different. There, the price dynamics is caused by a *single* large position holder who uses strongly suboptimal hedging techniques. Here we consider a population of hedgers who use optimal hedge strategies as outlined in this paper, and we determine the unavoidable effects that they would have on the market dynamics.

As above, we assume that Γ can be taken constant during the period of interest. We also assume for simplicity that we are far enough from expiration so $\kappa(T - t) \gg 1$ and hence $h = 1$ so that the hedge strategy is $\theta_t = -\kappa Y_t$. Then (10) gives

$$Y_t = \sigma \Gamma \int_0^t e^{-\tilde{\kappa}(t-s)} dW_s. \quad (26)$$

We assume that the position is initially correctly hedged so $Y_0 = 0$. We denote

$$\tilde{\kappa} = G \kappa$$

with $\tilde{\kappa} \approx \kappa$ if permanent impact ν is not too large (recall $G = 1 + \nu\Gamma$).

The approximate instantaneous size of the mishedge Y_t is

$$\mathbb{E}Y_t^2 = \frac{\sigma^2 \Gamma^2}{2\tilde{\kappa}} = \frac{\sigma \Gamma^2}{2G} \sqrt{\frac{\eta}{\lambda}}.$$

The mishedge size, measured in shares (recall that $Y = X - \Delta$) increases in proportion to the total position size Γ as expected, except for feedback effects contained in the factor $G = 1 + \nu\Gamma$. For a given position size, the mishedge increases as risk aversion λ decreases, and it decreases as temporary impact η decreases. Permanent impact ν does not appear in this expression at leading order, except as a small adjustment of the value to which hedging is made.

The total amount lost by the hedgers to temporary market impact is approximately

$$\int_0^T 2\eta \theta_t^2 dt = \int_0^T 2\eta \kappa^2 Y_t^2 dt \sim 2T \sigma^3 \Gamma^2 \lambda^{1/2} \eta^{1/2}.$$

As temporary impact $\eta \rightarrow 0$, not only does the optimal hedge position track the Black-Scholes value more and more closely, but also the total cost of this hedge decreases to zero, even though trading becomes more and more active. We eliminate the plausible but false scenario in which the hedge error decreases to zero but the trading cost increases. In the limit of zero temporary cost we fully recover the Black-Scholes solution.

To determine the price process, note that (12) since $\Delta(t, P_t) = \Delta_0 - \Gamma(P_t - P_0)$, we have

$$Y_t = X_t - \Delta(t, P_t) = X_t - X_0 + \Gamma(P_t - P_0).$$

Solving for $P_t - P_0$ between this and (1), we obtain

$$G(P_t - P_0) = \nu Y_t + \sigma W_t$$

or

$$P_t = P_0 + \frac{\sigma}{G} \left(W_t + \nu \Gamma \int_0^t e^{-\tilde{\kappa}(t-s)} dW_s \right). \quad (27)$$

The price process given by (26) has momentum or mean reversion across time scales of length $\sim \kappa^{-1}$, depending on the sign of Γ . One way to describe such a process is to compute the effective variance $\sigma_{\text{eff}}^2(\delta t)$ that would be measured on a time interval of fixed length δt . In the market microstructure literature, this is often called the “signature” of

the process, though it is usually taken to reflect effects such as bid-ask bounce rather than the impact effects considered here. In this case we obtain

$$\sigma_{\text{eff}}^2(\delta t) = \frac{1}{\delta t} \mathbb{E}(P_{t+\delta t} - P_t)^2 = \left(\frac{\sigma}{G}\right)^2 \left(1 + (2 + \nu\Gamma)\nu\Gamma \frac{1 - e^{-\tilde{\kappa}\delta t}}{\tilde{\kappa}\delta t}\right).$$

We readily see that

$$\sigma_{\text{eff}}(\delta t) \sim \begin{cases} \sigma & \text{for } \tilde{\kappa}\delta t \ll 1, \text{ and} \\ \frac{\sigma}{G} & \text{for } \tilde{\kappa}\delta t \gg 1. \end{cases} \quad (28)$$

While the instantaneous price process has the original volatility σ , a modified volatility will be observed on time scales longer than the hedge time scale. If $\Gamma > 0$, then this modified volatility will be smaller than the original volatility; as observed at the end of Section 2.2, the long Γ position is easy to hedge since trading towards the hedge portfolio moves the price in, reducing effective volatility. This is related to the “pinning” near expiration modeled by Avellaneda and Lipkin [2003], when market makers are net long. If $\Gamma < 0$ then volatility is enhanced, since hedge trading pushes the price away.

Temporary impact η sets the shortest time scale on which this modified volatility can be observed, but its magnitude is determined entirely by the permanent impact ν and the net position Γ of the hedgers. Thus Lions and Lasry [2006, 2007], with no temporary impact, obtained a modified Brownian motion on infinitesimal time scales.

In principle this effect could be observed from market data, if a reliable estimate for the total hedge position Γ were available.

3.4 No permanent impact, general Gamma

We can relax the constant gamma assumption and allow for general options. Thus $\Gamma = \Gamma(t, p)$ is a function of arbitrary form, satisfying (4) and (5). However, to make the solution tractable, we need to dispense with permanent impact ($\nu = 0$). We are still able to obtain a fairly explicit solution for the control θ_t , which illustrates an important asymmetry.

Now we look for a solution to (15) in the form

$$J(t, p, y) = \frac{1}{2} A_2(T-t) y^2 + A_1(T-t, p) y + A_0(T-t, p). \quad (29)$$

from which we obtain the system of one ordinary and two partial differential equations

$$\begin{aligned} \dot{A}_2 &= 2\lambda\sigma^2 - \frac{1}{2\eta} A_2^2 & A_2(0) &= 2\lambda\Sigma^2 \\ \dot{A}_1 &= \frac{1}{2}\sigma^2 A_1'' - \frac{1}{2\eta} A_2 A_1 & A_1(0, p) &= -2\lambda F(0, p) \\ \dot{A}_0 &= \frac{1}{2}\sigma^2 A_0'' + \sigma^2 \Gamma A_1' + \frac{1}{2}\sigma^2 \Gamma^2 A_2 - \frac{1}{4\eta} A_1^2 & A_0(0, p) &= \lambda \Xi(p)^2. \end{aligned}$$

with

$$F(0, p) = \mathbb{E}[\Pi \xi \mid P_T = p]. \quad (30)$$

(If $\nu \neq 0$ and Γ is not constant, then terms with $\nu\Gamma$ in \dot{A}_2 force A_2 to depend on p and make the problem intractable.)

The solution for A_2 is the same as in (22) with $\nu = 0$:

$$A_2(\tau) = 2\eta\kappa h(\kappa\tau)$$

with $h_0 = \lambda\Sigma^2/\eta\kappa$. This h_0 is nonnegative and positive if overnight risk is nonzero. Hence $A_2(\tau)$ is nonnegative, strictly positive except possibly at $\tau = 0$, and bounded.

To solve for $A_1(\tau, p)$, write

$$A_1(\tau, p) = \exp\left(-\frac{1}{2\eta} \int_0^\tau A_2(s) ds\right) B(\tau, p)$$

so that $\dot{B} = \frac{1}{2}\sigma^2 B''$ and $B(0, p) = A_1(0, p) = -2\lambda F(0, p)$. Hence $B(\tau, p) = -2\lambda F(\tau, p)$, where $F : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$ is the solution of the heat equation $\dot{F} = \frac{1}{2}\sigma^2 F''$ with initial data (30). From (14), the optimal control is

$$\theta_t = -\frac{1}{2\eta} J_y = -\frac{1}{2\eta} (A_2 Y_t + A_1) = -\frac{1}{2\eta} A_2 (Y_t - \bar{Y}), \quad \bar{Y} = -\frac{A_1}{A_2}$$

Substituting the expressions for A_1 , A_2 , and h , and carrying out the integration, we determine the optimal control (compare (25))

$$\theta_t = -\kappa h(\kappa(T-t))(Y_t - \bar{Y}_t), \quad (31)$$

with

$$\bar{Y}(t, P_t) = \frac{\lambda}{\eta\kappa \sinh \kappa(T-t) + \lambda\Sigma^2 \cosh \kappa(T-t)} F(T-t, P_t). \quad (32)$$

We trade not towards the perfect Black-Scholes hedge $Y_t = 0$ as for the constant- Γ case, but towards an offset value \bar{Y} . It is evident from (32) that $\bar{Y}_t \rightarrow 0$ as $\kappa(T-t) \rightarrow \infty$, since $F(\tau, P_t)$ obeys the maximum principle, and the sinh and cosh in the denominator tend to ∞ . The offset is negligible when we are more than one typical hedge time away from the close of trading. The offset is also zero if the overnight risk is such that $F(0, p) = 0$. The asymmetry is thus due entirely to overnight hedging.

It may be surprising that the asymmetry in the option value does not appear before we are near to the close of trading. An explanation for this is that the hedge strategy is given in terms of the change in Δ rather than in terms of the underlying price change. Thus a positive price change δP and its opposite $-\delta P$ may cause changes of different size in the mishedge Y_t , which will cause trading at different rates.

To understand the nature of this terminal asymmetry, we note that for ξ and Π evaluated from an overnight process (rather than the more general formulation mentioned in Section 2.5), by Itô's Isometry

$$F(0, p) = \sigma^2 \mathbb{E} \left[\int_T^{T^*} (\Delta(t, P_t) - \Delta(T, P_T)) dt \mid P_T = p \right].$$

Since the price process has zero drift, this quantity is zero if Γ is constant and $\Delta' = 0$.

If Γ is increasing in p near P_T , then $\Delta'' < 0$, $\Delta(t, p)$ is concave down in p , and $F(0, p) < 0$; also, $F(T - t, P_t) < 0$ for t near T and P_t near P_T . Thus $\dot{Y}_t < 0$ and we trade towards a state with $X_t < \Delta(t, P_t)$. We desire to end the day "underhedged," because during the unhedgeable overnight moves, the expected decrease in optimal hedge if P_t decreases is smaller than the expected increase if P_t increases. The situation is the reverse if Γ is increasing in p .

Although this asymmetry appears explicitly only near the close of trading, in the middle of the trading day, the asymmetry appears implicitly via the definition of the mihedge $Y_t = X_t - \Delta(t, P_t)$. If the price changes by a small amount δP_t , then the change in the value of Δ may be larger for changes in one direction than in the other. Thus the change in Y_t will be larger on one side, and the rate of trading will be larger on that side. The target portfolio is always the perfect hedge, unless near close when we are facing an unhedgeable asymmetric risk.

3.5 Discrete time

The example in the Introduction has illustrated the risks of using a naive hedging strategy on a discrete time grid. We now show how to do a more correct computation of the hedge strategy with discrete time steps.

Suppose that we are allowed to reevaluate our trade strategy only at a discrete set of times t_0, \dots, t_{N-1} , with $t_0 = 0$ and $t_N = T$. We do not assume that these times are uniformly spaced, or that the time intervals $\delta_k = t_{k+1} - t_k$ are small. At each time t_k for $k = 0, \dots, N - 1$, we set our trade rate θ_k , which is to be held constant through the entire interval $t_k \leq t < t_{k+1}$. We denote by θ_k, P_k , etc. the values at $t = t_k$.

For simplicity, we use the constant- Γ approximation of Section 3.2. Then the share holdings, the stock price, and the mihedge evolve for t between t_k and t_{k+1} according to (compare (9,10))

$$\begin{aligned}\theta_t &= \theta_k \\ P_t &= P_k + \nu\theta_k(t - t_k) + \sigma(W_t - W_k) \\ Y_t &= Y_k + G\theta_k(t - t_k) + \sigma\Gamma(W_t - W_k)\end{aligned}$$

in which we again abbreviate $G = 1 + \nu\Gamma$.

The obvious strategy would evaluate the continuous-time rule (25) at (t_k, Y_k) :

$$\theta_k = -\kappa h Y_k$$

where h is evaluated at $\kappa G(T - t_k)$. But under this rule,

$$Y_{k+1} = (1 - Gh\kappa\delta)Y_k$$

(we denote $\delta = \delta_k$ for brevity). This gives the well-known Euler instability, with exponential growth in $|Y_k|$, unless $\kappa\delta < 2$ so that $|1 - Gh\kappa\delta| < 1$ (recall that G and h have values near one). For small temporary impact η , the relaxation rate κ may be large, and this is a very severe restriction on the maximum time step δ . We need a time discretization that does not depend on the value of $\kappa\delta$.

To compute the fully optimal discrete-time solution, we compute

$$\begin{aligned}
J(t_k, \mathcal{Y}) &= \inf_{\theta_k, \dots, \theta_{n-1}} \mathbb{E} \left[\lambda (Y_T \Pi - \xi)^2 + \sum_{j=k}^{N-1} \int_{t_j}^{t_{j+1}} (\lambda \sigma^2 Y_t^2 - \nu Y_t \theta_t + \eta \theta_t^2) dt \mid Y_k = \mathcal{Y} \right] \\
&= \inf_{\theta} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} (\lambda \sigma^2 Y_t^2 - \nu Y_t \theta_t + \eta \theta_t^2) dt + J(t_{k+1}, Y_{k+1}) \mid Y_k = \mathcal{Y}, \theta_k = \theta \right] \\
&= \inf_{\theta} \left[\left(\lambda \sigma^2 \left[\mathcal{Y}^2 + \left(\frac{1}{2} \sigma^2 \Gamma^2 + G \theta \mathcal{Y} \right) \delta + \frac{1}{3} G^2 \theta^2 \delta^2 \right] \right. \right. \\
&\quad \left. \left. - \nu \theta \left[\mathcal{Y} + \frac{1}{2} G \theta \delta \right] + \eta \theta^2 \right) \delta + \mathbb{E} [J(t_{k+1}, Y_{k+1}) \mid Y_k = \mathcal{Y}, \theta_k = \theta] \right].
\end{aligned}$$

We look for a solution in the form

$$J(t_k, \mathcal{Y}) = \frac{1}{2} A_k \mathcal{Y}^2 + C_k \quad \text{for } k = 0, \dots, N,$$

which gives

$$\begin{aligned}
\frac{1}{2} A_k \mathcal{Y}^2 + C_k &= \\
&\inf_{\theta} \left[\left(\lambda \sigma^2 \left[\mathcal{Y}^2 + \left(\frac{1}{2} \sigma^2 \Gamma^2 + G \theta \mathcal{Y} \right) \delta + \frac{1}{3} G^2 \theta^2 \delta^2 \right] - \nu \theta \left[\mathcal{Y} + \frac{1}{2} G \theta \delta \right] + \eta \theta^2 \right) \delta \right. \\
&\quad \left. + \frac{1}{2} A_{k+1} \left(\mathcal{Y}^2 + (\sigma^2 \Gamma^2 + 2G \theta \mathcal{Y}) \delta + G^2 \theta^2 \delta^2 \right) + C_{k+1} \right].
\end{aligned}$$

The optimal control is

$$\theta_k = -\frac{GA_{k+1} - \nu + \lambda \sigma^2 G \delta}{2\eta + (GA_{k+1} - \nu)G\delta + \frac{2}{3}\lambda \sigma^2 (G\delta)^2} Y_k, \quad (33)$$

and we obtain the iterative relation $A_N = 2\lambda \Sigma^2$ and $A_k = F(A_{k+1})$, with

$$F(A) = A + \left(2\lambda \sigma^2 - \frac{(GA - \nu + \lambda \sigma^2 G \delta)^2}{2\eta + (GA - \nu)G\delta + \frac{2}{3}\lambda \sigma^2 (G\delta)^2} \right) \delta. \quad (34)$$

In the limit $\delta \rightarrow 0$, (33) reproduces (14), and (34) reproduces (18).

Assertion Under the same stability condition (21) on the parameters as for the continuous-time case, and for arbitrary time steps δ_k , the dynamics given by (34) gives a well-behaved evolution for A_k , with $A_k \geq 0$ for each k . “Well-behaved” means that if δ_k is constant, then A_k tends monotonically to a fixed value as k decreases, and if δ_k varies, then A_k moves always in the direction of a variable target.

Proof $F(A) - A$ is rational in A . Its numerator is quadratic in A with zeros at

$$A_{\pm}(\delta) = \frac{1}{G} \left(\nu \pm \sqrt{4\eta\lambda\sigma^2 + \frac{1}{3}(\lambda\sigma^2)^2(G\delta)^2} \right)$$

which approach the stable points (20) for the differential equation as $\delta \rightarrow 0$. The numerator is positive for $A_-(\delta) < A < A_+(\delta)$ and negative for $A > A_+(\delta)$. Since $A_-(\delta) < A_-(0)$ for $\delta > 0$, we may say that the numerator is positive for $A_-(0) < A < A_+(\delta)$.

The denominator of $F(A) - A$ is linear in A , with a single zero at

$$\bar{A}_0(\delta) = \frac{1}{G} \left(\nu - \frac{2\eta}{G\delta} - \frac{2}{3}\lambda\sigma^2 G\delta \right).$$

The denominator is positive for $A > \bar{A}_0(\delta)$. Maximising $\bar{A}_0(\delta)$ over δ , we can say that the denominator is positive for $A > \bar{A}_0^{\max}$ for all δ , with

$$\bar{A}_0^{\max} = \max_{\delta} \bar{A}_0(\delta) = \frac{1}{G} \left(\nu - \frac{4}{\sqrt{3}} \sqrt{\eta\lambda\sigma^2} \right) < A_-(0).$$

In particular, the denominator is positive for $A > A_-(0)$.

Thus $F(A) - A > 0$ for $A_-(0) < A < A_+(\delta)$ and $F(A) - A < 0$ for $A > A_+(\delta)$. Condition (21) assures us that $A_N > A_-(0)$, and hence that $A_k > A_-(0)$ for all $k \leq N$.

Furthermore,

$$F'(A) = \frac{\left(2\eta - \frac{1}{3}\lambda\sigma^2(G\delta)^2\right)^2}{\left(2\eta + (GA - \nu)G\delta + \frac{2}{3}\lambda\sigma^2(G\delta)^2\right)^2}$$

which is always nonnegative, and strictly positive except at a special value of δ such that $\eta = \frac{1}{6}\lambda\sigma^2(G\delta)^2$, that is, $\kappa^2\delta^2 = 6/G^2$. At that special value, $F'(A)$ is zero and indeed $F(A)$ has the constant value $A_+(\delta)$. Positivity of the derivative assures us that $F(A) \leq A_+(\delta)$ for $A_-(0) \leq A < A_+(\delta)$, and $A_+(\delta) \leq F(A)$ for $A \geq A_+(\delta)$.

Combining the two results above, we have $A < F(A) \leq A_+(\delta)$ for $A_-(0) \leq A < A_+(\delta)$, and $A_+(\delta) \leq F(A) < A$ for $A \geq A_+(\delta)$, and this gives us convergence to the stable point $A_+(\delta)$. If the time step δ varies from step to step, then the dynamics will track the moving stationary point.

For the positivity, the above give $F(A) \geq \min\{A, A_+(\delta)\}$. We have $A_N = 2\lambda\Sigma^2 \geq 0$; also $A_+(\delta) \geq 0$. Hence $A_k \geq 0$ for all $k \leq N$. That is, the value function is convex and the stationary point is indeed a minimum.

4 Conclusion

We have considered the problem of hedging an options position in the presence of both temporary and permanent impact. The solution consists of smooth trading in the direction of a target portfolio determined by the Black-Scholes delta-hedge, at a rate determined by the balance between temporary market impact and risk aversion. Permanent market impact causes a modification of the realised volatility in the public market, on time scales longer than the intrinsic hedge time.

This provides a way to think about market impact for options trading. Suppose that a buy-side trader purchases a large quantity of options from a sell-side trader. We assume that the buy-side trader purchases the option to hedge an external risk, and so does not hedge the position. The sell-side trader has no outside risk, and so hedges the option position. The market impact of this trade will be felt in two ways. First, the implied volatility of the option contract will rise, as the sell-side dealer raises his prices to counteract the buy-side demand. Second, the realised volatility of the underlying asset will also rise, as the sell-side dealer hedges his position in the open market. That is, in (28), we have $\Gamma < 0$ since the hedger is short the option; thus $G = 1 + \nu\Gamma < 1$, and $\sigma_{\text{eff}} > \sigma$. Liquidity-demanding trading in an options contract thus affects both types of volatility.

We make predictions that can be tested, if suitable market data can be found. The key variable is Γ which, as discussed in Section 3.3, represents the entire net position of traders who are hedging their options position rather than holding the option to hedge exogenous sources of risk. We typically assume these market participants to be sell-side dealers as well as market makers. This quantity is not related to total open interest in the option, but requires more detailed information about what types of market participants hold what net positions. This would be similar to the Liquidity Data Bank data for futures (see, for example, Almgren and Burghardt [2011]) previously available from the Chicago Mercantile Exchange. If similar data could be obtained for options (on any asset class) then the prediction of a relationship between net position of hedgers, realised volatility, and time to expiration could be tested.

References

- R. Almgren. Treasury price swings on October 15, 2014. Technical report, Quantitative Brokers, June 2015.
- R. Almgren and G. Burghardt. A window into the world of futures market liquidity. Research snapshot, Newedge Prime Brokerage, March 2011.
- R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3(2):5-39, 2000.
- R. Almgren, C. Thum, E. Hauptmann, and H. Li. Equity market impact. *Risk*, 18(7, July):57-62, 2005.
- U. M. Ascher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, 1998.
- M. Avellaneda and M. D. Lipkin. A market-induced mechanism for stock pinning. *Quant. Finance*, 3:417-425, 2003.
- M. Avellaneda, G. Kasyan, , and M. D. Lipkin. Mathematical models for stock pinning near option expiration dates. *Comm. Pure Appl. Math.*, 65(7):949-974, 2012.
- U. Çetin, H. M. Soner, and N. Touzi. Option hedging for small investors under liquidity costs. *Finance Stoch.*, 14(3):317-341, 2010.
- J. Cvitanić and H. Wang. On optimal terminal wealth under transaction costs. *J. Math. Econom.*, 35:223-231, 2001.

- M. H. A. Davis and A. R. Norman. Portfolio selection with transaction costs. *Math. Oper. Res.*, 15 (4):676–713, Nov. 1990.
- K. Devasabai. No flash crash: Paulson, Pimco and the US Treasury meltdown. *Risk*, Dec 2014.
- N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *J. Finance*, 68(6), Dec. 2013.
- O. Guéant and J. Pu. Option pricing and hedging with execution costs and market impact. *Math. Finance*, 2015. doi: 10.1111/mafi.12102.
- I. Hwang, W. Kisling, and N. Mehta. Investors whipsawed by price swings in IBM, Coca-Cola. *Bloomberg*, July 2012.
- M. Jeannin, G. Iori, , and D. Samuel. Modeling stock pinning. *Quant. Finance*, 8(8):823–831, 2008.
- Y. Kabanov and M. Safarian. *Markets with Transaction Costs: Mathematical Theory*. Springer, 2009.
- Y. M. Kabanov and M. M. Safarian. On Leland's strategy of option pricing with transactions costs. *Finance Stoch.*, 1(3):239–250, 1997.
- J. E. Kambhu. Dealers' hedging of interest rate options in the U.S. dollar fixed-income market. *FRTNY Economic Policy Review*, pages 35–57, June 1998.
- C.-A. Lehalle and M. Lasnier. What does the saw-tooth pattern on US markets on 19 July 2012 tell us about the price formation process. Technical report, CA Chevreux Quantitative Research, Aug. 2012.
- H. E. Leland. Option pricing and replication with transactions costs. *J. Finance*, 40(5):1283–1301, Dec 1985.
- P.-L. Lions and J.-M. Lasry. Towards a self-consistent theory of volatility. *J. Math. Pures Appl.*, 86: 541–551, 2006.
- P.-L. Lions and J.-M. Lasry. Large investor trading impacts on volatility. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 24(2):311–323, 2007.
- J. Lorenz and R. Almgren. Mean-variance optimal adaptive execution. *Appl. Math. Finance*, 18: 395–422, 2011.
- L. C. G. Rogers and S. Singh. The cost of illiquidity and its effects on hedging. *Math. Finance*, 20 (4):597–615, Oct. 2010.
- S. E. Shreve and H. M. Soner. Optimal investment and consumption with transaction costs. *Ann. Appl. Probab.*, 4(3):609–692, Aug. 1994.
- H. M. Soner, S. E. Shreve, and J. Cvitanić. There is no nontrivial hedging portfolio for option pricing with transaction costs. *Ann. Appl. Probab.*, 5(2):327–355, 1995.
- N. Touzi. *Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE*. Springer, 2013.
- S. T. Tse, P. A. Forsyth, J. S. Kennedy, and H. Windcliff. Comparison between the mean-variance optimal and the mean-quadratic-variation optimal trading strategies. *Appl. Math. Finance*, 20 (5):415–449, 2013.
- Y. Zhao and W. T. Ziemba. Hedging errors with Leland's option model in the presence of transaction costs. *Finance Research Letters*, 4:49–58, 2007.