1. Transaction Costs: Much Smaller than We Thought

Summary: We analyze the economics of transaction costs in dynamic hedging of contingent claims, making the case for a representative dynamic hedging agent. We argue that the most efficient dynamic hedger incurs no transaction costs when owning options. Furthermore, as options are rarely replicated until maturity, the expected transaction costs of the short options depend mostly on the dynamics of the order flow in the option markets — not on the direct costs of transacting. The conclusion is that transaction costs are a fraction of what has been assumed to be in the literature. For the efficient operators (and those operators only), markets are more dynamically complete than anticipated.

1.1. Introduction

A retired investor in Florida equipped with a telephone and a high speed modem should perhaps refrain from dynamically replicating an option, an activity that would prove so onerous that it may rob him of his nest egg. Furthermore, assume he owns an option in his inventory. Owing to the cost of replication of the option, selling it would reduce his transaction costs, thus providing the curious situation of positive transaction costs. This paper’s aim is the use of the economists’ available tools to study the properties of transaction costs in option replication for the professional and specialized dynamic hedger, operating in (alas) discrete time, not the unspecified and occasional one.

Since the original paper by Leland (1985), transaction costs in option replication have been extensively investigated in the literature. Yet we still do not have any particularly satisfying way to incorporate them into conventional option valuation. Incorporating transaction costs in an Arrow-Debreu model that does not include financial institutions, as has been done, causes distortions and paradoxes. We will see that adding the financial sector resolves such paradoxes.

This paper’s primary motivation is to address the question that a naïve factoring of transaction costs into the Black-Scholes(1973) and Merton (1973) model framework would clearly disturb the risk-neutral argument¹ and set option valuation.

¹ Empirical studies show the implied volatility at which options trade in the market to be sufficiently close to historical measures of actual volatility (with small biases, if any) to convince us of the absence of transaction costs. Should market makers incur the full
Transaction Costs: Much Smaller than We Thought

theory back a few decades. Transaction costs induce, for obvious reasons, dynamic incompleteness, thus preventing valuation as we know it, which would cause the double effect of creating a need for options and preventing them from trading, a condition known as the Hakansson paradox, after Hakansson (1979). It will be argued that a resolution of the Hakansson paradox could come from market structure. The market is segmented between two classes of operators; for the first category, it will be shown that options are redundant (or nearly so); for the other category, options are not redundant, thus justifying a demand for their use.

The first category here is deemed to be the most efficient dynamic hedgers, hence MEDH, specialized market makers who compete with each other to provide liquidity in option instruments, and maintain inventories in them. They rationally limit their dynamic replication to their residual exposure (what will be defined as net gamma), not their gross exposure. In addition, the fact that they do not hold options until maturity greatly reduces their costs of dynamic hedging. They have an incentive in the acceleration of financial intermediation (what Merton (1992) calls the “spiraling towards dynamic market completeness”).

The second category includes those who merely purchase or sell financial instruments that are subjected to dynamic hedging; they are those agents market microstructure theory calls “price takers”. They, accordingly, neither are equipped for dynamic hedging, nor have the need for it, thanks to the existence of specialized and more efficient market makers. The examination of their transaction costs in the event of their decision to dynamically replicate their options is of no true theoretical contribution.

Leland transaction costs, they would have passed on the expenditure. It would have appeared in the data and volatility would have been abnormally high (or abnormally low) as compared to historical measures. See Scott (1992), Xu and Taylor (1994), or Campa and Chang(1995) for an indirect evidence of the absence of transaction costs through their testing of the expectation hypothesis.

2 It can be argued, using the Allen and Gale(1988) argument of instrument migration, and the Ross (1989) notion of near-spanning that, when replication costs force the dynamic hedger to offer the option at a price far in excess of the Black-Scholes-Merton value, the demand for option would be easily satisfied by other means—other instruments linked to the same source of uncertainty. If high transaction costs in the replication of the option by the specialized dynamic hedger lead to a high option price, then financial innovation would
A second distinction in this analysis, between a long option and short option optimal dynamic hedging policy for the MEDH, will show asymmetries in the optimal policy. A third distinction, concerning rebalancing policies, split between time-based rules and price-based (or return-based) ones, will be shown to yield non trivial differences between them from a microstructure standpoint. The passage from the economics of continuous time to that of discrete time is not as smooth as it was thought out to be. Just as the relaxation of the assumption of the Walrasian auctioneer led to the thriving market microstructure theory, the relaxation of infinitesimal time increments $dt$ in option replication yields non trivial consequences on the nature of the optimal policy. The intuition of the results are shown in Table I.

We call the modified Leland implied standard deviation, $\sigma_A$, after Leland (1985), any modifications to the implied standard deviation in the Black-Scholes formula to take into account the transaction costs under a well set utility-free option rebalancing policy.

<table>
<thead>
<tr>
<th>Long Gamma</th>
<th>Time-based revision policy ($\Delta t$ rule)</th>
<th>Price based revision policy ($\Delta S$ or $\Delta S/S$ rule)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not optimal</td>
<td>Optimal, No transaction costs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Short Gamma</th>
<th>Time-based revision policy ($\Delta t$ rule)</th>
<th>Price based revision policy ($\Delta S$ or $\Delta S/S$ rule)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not feasible (capital constraint)</td>
<td>Optimal, Modified Leland $\sigma_A$ small and depending on the expected option holding period.</td>
</tr>
</tbody>
</table>

Note that the equilibrium restrictions on the transaction costs are introduced in the literature by Constantinides(1998).
The rest of this paper is organized as follows. Section 1.2 presents the concept of adjusted sample path in its nuances and critically reviews the different approaches to transaction costs in the literature. Section 1.3 discusses the inapplicability of the time-based hedging rule for a short option operator with finite capital and presents the rationale for and the dynamics of the most efficient dynamic hedger and explains, on economic grounds, the inapplicability of the $\Delta t$ rule for the long option hedger, which leads to the result that such operator incurs no transaction costs. Section 1.4 discusses the importance of offsetting trades. Section 1.5 makes the case for a representative agent and proves that the option owner incurs no transaction costs. Section 1.6 presents an order flow model for the options and computes the $\lambda$, which is the proportion of the Leland adjustment that the operator is expected to incur. The conclusions are presented in section 1.7.

1.2. The Notion Of Adjusted Sample Path and Modified Leland Adjustment.

Dynamic hedging is defined here as an attempt to replicate the payoff of a European option $V$ using a strategy pair sequence $(Q_t, B_t)_{t=0}^n$ where $Q_t$ is the quantity of units (or shares) of the underlying asset $S$ (assumed to follow a geometric Brownian motion) held at time $t$, $t_0 < t_i < t_n$ and $B_t$ are the cash balances held in a default-free interest bearing money market account that satisfies all of the following:

i) $Q_t$ is $F_t$ adapted, $F_t$ being the filtration $F_{t,j} \subset F_{t,i}$, all information related to $S_t$. We are assuming a left-continuous sample path for $S$.

ii) $E_{t_j, t_i} (q V(S_t, t_i) - B_{t_i} + Q_{t_i} S_{t_i})_{i=0}^n = 0$ for all $t_i$ and $t_j$, where $V$ is the market value of the option and $q$ the (positive or negative) quantities of the option held in inventory.

iii) $n < \infty$

iv) $q (V(S_t, t_i) - B_{t_i} + Q_{t_i} S_{t_i}) - V(S_t, t_0) - B_{t_0} + Q_{t_0} S_{t_0} < K_{t_i}$ for all $t_0 < t_i < t_n$, where $K_{t_i}$ is the capital of the operator. As the option replicating operator has finite capital, the value of the package should not drop below some threshold.

We further define the two distinguishable types of dynamic hedges. A positive gamma dynamic hedging strategy over an interval $[S-a, S+b]$ between periods $t_i$ and $t_j > t_i$ is a sequence where $\text{sign}(Q_{t_i} - Q_{t_j}) = \text{sign}(S_{t_i} - S_{t_j})$ and a negative gamma strategy corresponds to a strategy where $\text{sign}(Q_{t_j} - Q_{t_i}) = \text{sign}(S_{t_j} - S_{t_i})$. One remarks that the sign of the gamma corresponds to that of the $q$ in the case examined here of a European option –not for more complex options like digital options, and not for portfolios include more than one option.

While in continuous time the self financing condition is imposed in the literature, that $(q V(S_t, t_i) - B_{t_i} + Q_{t_i} S_{t_i})_{i=0}^n = 0$ for all $t_i$ and $t_j$, here we allow for a small relaxation by requiring that $E((q V(S_t, t_i) - B_{t_i} + Q_{t_i} S_{t_i})_{i=0}^n) = 0$. Such relaxation
Transaction Costs: Much Smaller than We Thought

is necessary owing to the fact that time increments are finite\(^3\). The package will be
unavoidably stochastic owing to the impracticality, in the real world, of performing
continuous time dynamic hedges (condition iv). Given that the revisions are not
done at infinitesimal time lapses, all the operator can do is attempt to come as close
as possible to the self financing strategy so as to allow for the pricing of the option
instrument as if it were as redundant as possible, namely by reducing the variance
of the package.

Leland (1985) builds on the Boyle and Emanuel (1980) pioneering work on
the properties of the Black-Scholes & Merton portfolio revised discretely. It
introduces the trade-off between the variance of return in a portfolio and
the costs of reduction through frequent dynamic hedging\(^4\).

Two general policies are available to the dynamic hedger. The first, that we
call a \(\Delta t\) rule, presented by Boyle and Emanuel (1980) and adopted by Leland
(1985), assumes that the operator revises the portfolio at exogenously set time
increments. The second, that we call a \(\Delta S / S\) rule, assumes that the revisions take
place according to set variations in the returns, which leads to a stochastic time
lapses between interventions. The utility free approach, examined by Boyle and
Vorst (1992), assumes a constant \(\Delta S / S\). The utility based approach, pioneered by
Constantinides(1979), examined by Hodges and Neuberger (1989), Davis, Panas
and Zariphoupoulou (1993) and other authors, assumes that the dynamic revisions
take place upon variations in \(\Delta S / S\) that are large enough to create a second order
exposure in the portfolio. It can be casually termed “hedging whenever it matters”,
as the optimal rebalancing policy becomes the solution of a stochastic control
problem, taking the policy that realizes the supremum of expectation of the
expected utility of a stochastic cash flow. Note that Constantinides (1979) proved
that the joint assumptions of concave utility and proportional transaction costs
imply that the optimal policy is unique. While these models present the advantage
of showing the existence of a solution to the problem, they are not retained here
because they make it difficult to accommodate decision making based on option
order flow –and mostly because it has been difficult to derive tangible valuation
under utility. We argue that a utility framework is not applicable if, under the
optimal hedging policy, transaction costs are deemed either non existent or too

\(^3\) Should the operator revise his portfolio in continuous time (assuming that would be
possible), it can be shown the smallest amount of transaction costs would lead to an infinite
price for the option for a short option replicator (and conversely to a negative value for a
long option hedger) –the optimal policy becomes that of a buy-and hold (for a call option),
indeed not a dynamic hedging policy.

\(^4\) See Gilster(1990) for an analysis of the variance of the replicating portfolio.
small. Accordingly, we will assume, in the remaining part of the paper, risk
neutrality.

1.2.1. The $\Delta t$ adjustment rule

Leland (1985) presents the concept of adjusted sample discrete Brownian path with
a higher volatility than that of the equivalent Black-Scholes-Merton frictionless
one, due to the need for the dynamic hedger to buy the underlying security at a
proportionally marked-up price and sell it at a proportionally marked-down price.
Such sample path is the one that would be obtained should the dynamic hedger
revise his portfolio at a set increment $\Delta t$, while transacting the underlying security
at the following prices: The operator buys at $S_{it}^{\text{ask}} = S_{it} (1 + k/2)$ and sells at $S_{it}^{\text{bid}} = S_{it} (1-k/2)$. The resulting variance for the equivalent sample path for the underlying
security becomes:

$$\sigma^{+2} = \sigma^2 (1 + \frac{2}{\sqrt{\pi}} \frac{k}{\sigma \sqrt{\Delta t}})$$  \hspace{1cm} (1)

and the contingent claim can thus be priced using a variance of the portfolio that
can be adjusted up or down accordingly.\footnote{Whaley and Wilmot (1993), Avellaneda and Paras (1994) generalized the approach to
price the contingent claim including a positive or negative adjustment and rewrite a
stochastic differential equation in which the variance would be adjusted up and down
according to whether the operator has locally a positive second derivative with respect to
the underlying price. The latter paper introduces a solution for the pathological case of an
operator owning options better left not hedging.}

It is easy to see how, as in the Boyle and Emanuel (1980) framework, the
number of revisions, called positive and negative gamma adjustment, reduces the
variance.

Define $M[n]$ as the quadratic norm of the variance of the replicating
portfolio during the life of the option and $n$ the number of adjustments:

$$n \equiv \frac{t_n - t_0}{\Delta t}$$

$$M_n = \sum_{i=1}^{n} \left[ C(S_{i-1}) - C(S_i) + (S_i - S_{i-1}) \frac{\partial C(S_{i-1})}{\partial S_{i-1}} \right]^2$$

it can be shown that:

$$M_{\frac{n}{j}} \sim M(n) \frac{1}{\sqrt{j}}$$

Thus the frequency of hedging reduces the variance by a factor of $1 / \sqrt{n}$.\footnote{Whaley and Wilmot (1993), Avellaneda and Paras (1994) generalized the approach to
price the contingent claim including a positive or negative adjustment and rewrite a
stochastic differential equation in which the variance would be adjusted up and down
according to whether the operator has locally a positive second derivative with respect to
the underlying price. The latter paper introduces a solution for the pathological case of an
operator owning options better left not hedging.}
1.2.2. The $\Delta S$ Rules and the Boyle and Vorst (1992) Model in Continuous Time

Boyle and Vorst (1992) provide the first adumbration of a fixed, non-utility based $\Delta S / S$ rule by using the binomial tree, where the operator revises his position at every node. Their result is that their adjusted volatility becomes higher than (1) by a factor of $\sqrt{\frac{\pi}{2}}$.

$$\frac{\sigma}{\Delta t} = \sigma^2 (1 + \frac{k}{\sigma \sqrt{\Delta t}})$$  \hspace{1cm} (2)

The binomial trees do not assume that between two legs the maxima (minima) can be higher (lower) than the price at the node. Assuming $S$ is either followed over period $\Delta t$ by the up-price $S_u$ or the down price $S_d$, the maximum (minimum) cannot be higher (lower) than $S_u$ ($S_d$).

We remark that their results hold in continuous time (i.e. holding that the underlying security diffuses but that the dynamic hedger executes a discrete hedging policy) using the stopping time for a risk-neutral process subjected to two barriers $S(1+\nu)$ and $S(1-\nu)$, the two levels of interventions. Take

$$\tau = \inf \left\{ t_i : S_a \not\in \{S_u (1-\nu), S_d (1+\nu) \} \right\}$$  \hspace{1cm} (3)

The $\Delta S / S$ rule, by eliminating the absorbing barrier problem, becomes more costly. Thus, assuming no drift in the Brownian motion\(^6\) to simplify:

$$E\left(\frac{|\Delta S|}{S} | \Delta t \right) = \frac{\pi}{\sqrt{2 \sigma \sqrt{\Delta t}}}$$ \hspace{1cm} (4)

can be misleading. We need the stopping time at the revision price.

The expected length of time $\int_{t_0}^{\tau} \left( s - t_0 \right) \phi(s-t_0) ds$ between intervention $t_0$ and $\tau$, with levels equally spaced above and below the market, can be calculated as

$$E[\tau - t_0] = \frac{\nu^2}{\sigma} - \int_{t_0}^{\tau} \phi(s-t_0) ds$$ \hspace{1cm} (5)

---

\(^6\) The drift has at best a residual effect on the analysis here since the time lapses are very small. It would be equivalent to shifting the upper and lower bands by the same proportional amount, which has only a second order effect on the distribution of the stopping time.
where $\phi(s)$ is the risk-neutral density of the stopping time. When the option has time until expiration ($T$ large) or $v$ is sufficiently small, i.e. when $(1+v)$ and $(1-v)$ can be approximated by $e^v$ and $e^{-v}$ respectively, using standard results\(^7\), we get

$$E[\tau-t_0] \approx \left( \frac{v}{\sigma^2} \right)^2$$

Thus $E[\tau-t_0]$ is higher than the equivalent $\Delta t$ in Leland (1985) by a factor the upper bound of which is $\sqrt{2/\pi}$, which explains the markup over the Leland (1985) result. Thus we can use the markup in the Boyle and Vorst for a revision policy based on moves in the underlying asset values (not just time steps):

$$\sigma'^2 = \sigma^2 (1+\frac{k}{v})$$  \(6\)

We end this section by calling modified Leland adjustment to implied volatility $\sigma$, the adjusted standard deviation of the sample path given an unspecified utility free but fixed hedging policy and a $\Delta S$ rule any policy that is based on asset moves, proportional or simple. The latter policy should embrace the dynamic programming families of utility based policies.

### 1.3. On the Inadequacy of the $\Delta t$ Rule for a Short Gamma Dynamic Hedger.

How much leeway the operator has in the choice of his policy is limited by condition iv) since there is the absorbing barrier hanging over the operator during the life of the option. Clearly no operator has infinite capital.

\textit{Proposition 1.1} The $\Delta t$ rebalancing rule does not satisfy condition iv) for a short call not fully covered by the asset (in any quantity) $(Q_t/V)$ or a short put with any hedge (in amounts where $Q_t < 0$), while $\Delta S$ always satisfies condition iv).

\textit{Proof:}

1) Assume the instrument is a call. With $q<0$, $P[q V(S,t_i) + Q_t S_{t_i} - K_{t_i}] > 0$ when we assume left continuous sample path for $S$. Even with the generally assumed compact support of the

\textit{Proposition 1.1} The $\Delta t$ rebalancing rule does not satisfy condition iv) for a short call not fully covered by the asset (in any quantity) $(Q_t/V)$ or a short put with any hedge (in amounts where $Q_t < 0$), while $\Delta S$ always satisfies condition iv).

\textit{Proof:}

1) Assume the instrument is a call. With $q<0$, $P[q V(S,t_i) + Q_t S_{t_i} - K_{t_i}] > 0$ when we assume left continuous sample path for $S$. Even with the generally assumed compact support of the

---

\(^7\) The problems in obtaining a closed form solution to the expected conditional stopping time for a Brownian motion are circumvented here with the following approximations: 1) equally spaced barriers, 2) no drift, here too small to matter 3) too short a stopping time for the integral of the expected stopping time to be lengthier than expiration, i.e. $\int_0^T \phi(s) ds \#_{\tau^c}$ $s \phi(s)$ is residually small. See Borodin and Salminen (1996) for the derivations.
distribution the probability of exceeding a given amount is always positive. $V$ thus admits all possible values on the real line.

2) Assume the instrument to be a put, thus bounded by the (discounted value) of $S$ from above and 0 from below. When $q$ is negative, the delta hedge $Q_t$ is negative and there is a non-trivial probability of the losses on the hedge exceeding the variation in the put value by the amount of capital.

A long gamma operator, however, can elect the choice of a policy based on the revision rules $\Delta t$ or $\Delta S$ (a microstructure approach will be shown to prove the latter to be more efficient), while a short gamma operator would be required to use $\Delta S$.

Even when the short gamma operator claims to be following a time rebalancing rule, it would be inconceivable for him to ignore the absorbing barrier between two adjustment periods: the mere fact of being interested in what takes place in the market implies a hedging rule that depends on the price of the underlying security rather than the time interval. There is a non-trivial probability of bankruptcy when executing what is known in the business as a time stop - this policy is explicitly disallowed by most trading firms as short gamma traders are required to leave a price stop in markets, like foreign exchange, that trade 24 hours per day. Indeed we will show that a price stop, from a $\Delta S / S$ rule, can allow for convergence to risk-neutral pricing. While a nonleveraged investor can be short call (covered with the asset) with impunity, a dealer with leverage faces under the $\Delta t$ rule an unlimited loss, except in the cases where the lock-delta (i.e. the asymptotic delta) becomes zero (that can be achieved by being covered with 100% of the underlying security).

---

8 Sampling by the author involving 29 head traders during 1996 and 1997 at O’Connell and Piper-Taleb Research head trader discussion group in Chicago. The results (available from the author) confirm the general enforcement by trading institutions of the stop loss rule where mitigating action needs to be taken upon the changes in the market value of the portfolio dropping below a predetermined threshold. For an option portfolio this translates into mandatory revision.

9 The asymptotic delta is defined as the hedge ratio of the package at the limit of $S \rightarrow \infty$ or $S \rightarrow 0$. 
1.4. Importance of Offsetting Trades

The Leland and modified Leland markup $\sigma_A$ assume the extreme case that the dynamic hedger would keep the option in the inventory until expiration, that there are no offsetting trades that would cause the option to be hedged for a shorter period. There is a positive probability of the option being covered before should the order flow in the market dictate so. Merton(1990 chapter 14) discusses the importance of offsetting trades (although he applies it to the final delivery, not the dynamic hedging process). Intuitively if the MEDH is short a quantity of options, repurchasing an equal or smaller quantity would provide a positive transaction cost.

**Proposition 1-2**

Assuming a two-way order flow in the option markets, the expected risk-neutral value of the option under transaction costs is bounded from above by marked-up standard deviation of the modified Leland $V(\sigma^+) (\text{using policy } \Delta t)$ or $V'(\sigma^+) (\text{using policy } \Delta S/S)$.

**Proof:** Assume an operator with no inventory, who according to the modified volatility, sells at $V(\sigma^+)$ and purchases at $V(\sigma^-)$. Assume the operator bought the option at $V(\sigma^-)$. He would therefore gain at offering the option below $V(\sigma^+)$, as he would have an incentive to rid himself of it. Thus a dynamic hedger who has a positive probability of repurchasing the option prior to expiration $t_n$ runs a positive probability of deriving a profit from the sale of the option, as his holding will be shorter than $(t_n-t_d)$. Thus $E_{t_n<0} (\alpha V(S_{t_i}) - B_{t_i} + Q_{t_i} S_{t_i})_{i=0}^n >0$. The issue is discussed further in section 1.6 and the value $V$ adjusted accordingly.

1.5. The Case For the Representative Agent MEDH and Execution Asymmetries

What about the proportional costs of transacting $k$? In the Whaley and Wilmot (1993) and the Avellaneda and Paras(1994) generalization of the Leland approach an equal $k$ is applied to both long and short gamma. A closer examination of the dynamics of order execution shows a difference in the transaction costs between a long and short gamma position. Section II remedies such point.

**Assumption 1 (efficiency):** The most efficient dynamic hedger is the marginal operator with the most efficient cost structure.

Such a cost structure, we will see, depends on the internal organization of the operation and the access to markets.

**Assumption 2 (homogeneity):** $\sigma'$, the base variance of the process (outside of the effect of the transaction costs adjustments) is constant.

Assumption 2 imparts sufficient homogeneity to satisfy the criteria for a representative agent framework for analysis. Given Assumption 2, the market maker in the option will not make a bet based on implied volatility differential between his expectation and that of the market for the duration of the option. The pricing at the margin will be solely dictated by the expectation of the transaction costs during the life of the instrument in his inventory.

The consideration of competitive efficiency of the dynamic hedger imparts another homogeneity: that of transaction costs.

In the industry the MEDH is best known as the market maker who specializes in a particular option. Without going into the microstructure of the
Transaction Costs: Much Smaller than We Thought

product, it suffices to mention that the business of MEDH has evolved from that of a craft in the 1970s at the Chicago Board of Options Exchange in Chicago to competitive integrated groups acting across floors and financial instruments. Likewise the total dollar-volume of option transactions has shifted in the 1990s to close to 95% to the “upstairs” (for a discussion see Grossman, 1992, for stylized facts see (Taleb 1997a) or the practitioner literature such as Natenberg, 1995, Baird, 1995).

Assumption 3 (fixed costs) : On the margin we assume that the MEDH incurs no other transaction costs than the bid-offer spread of the market maker in the underlying security.

Thanks to Assumptions 1 to 3, the class of the MEDH can be analytically sketched as a representative agent making no profit. In other words we can use standard economic arguments to assume that the competition between dynamic hedgers will be such that the valuation will be that of a single break-even operator.

Assumption 4 (winner’s curse): No size impact by the dynamic hedger on the underlying process. In other words, there is no particular bias causing a “winner’s curse”.

Define a limit order in the marketplace as an order to sell above the current price or to buy below the current price, enforced in such a way that the market will not trade higher or lower than the limit order without the satisfaction of the order being guaranteed. Take a stack of (assumed equal) orders $o_1$ through $o_n$ at the limit price $S^*$, assumed higher than current price $S_t$. Take $\tau$ the stopping time at $S^*$. A positive priority fill for $o_i$ means that the order is filled first, or that $E[S_{t^*}] \leq S^* e^{\mu(t_j-\tau)}$ where $t_j$ is the stopping time of the next available price. A neutral priority fill means that $E[S_{t^*}] = S^* e^{\mu(t_j-\tau)}$. Equivalently, a neutral priority fill assumes an i. i. d. process for the underlying security whose process is examined between the fills, (i.e. no autocorrelation at the order of the intervention). Note that the general literature on limit orders (see the more recent analysis by Handa and Schwartz, 1997) focuses on the investor’s access to markets and ignores the more competitive market participants (like derivatives traders), a flaw that can be explained by the fact that their studies are focused on the more minor individual equity markets.

Define $\sigma_t$ as the adjusted volatility of the process under a given policy.

Proposition 1-3 A positive gamma MEDH incurs no transaction costs ($\sigma_t = \sigma$).

It is first necessary, for the proof, to establish that the positive gamma MEDH will elect the $\Delta S / S$ rule, and that it is more efficient. Assume that the MEDH will not be disadvantaged by a negative priority fill. Take $S_t(1+v)$ as a limit offer and $S_t(1-\nu)$ as a bid. The trader leaving the order will be guaranteed to be

---

10 An argument here against doubts about the feasibility of such priority neutrality is the fact that the MEDH should, tautologically, be the one that has at least such neutrality.
Transaction Costs: Much Smaller than We Thought

filled upon the trading of $S_t(1+\nu)$ or $S_t(1-\nu)$. Given neutral priority we assume that subsequent prices are a semi-martingale. By ignoring the subsequent drift over too small a period, $E[S_{\tau}|F_{\tau}] = S_\tau$, $\tau$ being the time when the barrier is reached. Thus he will not incur a bid-offer expenditure.

Figure 1 Limit Orders Execution. The positive gamma market maker puts stop orders in the market at barriers $S_u$ and $S_l$ without knowing which period it will hit. The market hits $S_u$ at period $t_3$ and the trader is filled at $S_u$ without transaction costs provided his order did not impact the market.

![Figure 1](image1.png)

Figure 2 Negative gamma order triggered at $t_3$ but filled 2 periods later. The negative gamma market maker puts limit orders in the market at barriers $S_u$ and $S_l$ without knowing which period it will hit. The market hits $S_u$ at period $t_3$ and the trader’s buy order hits the market at $t_4$, only to be filled during period $t_5$. The trader will incur transaction costs since his order is not anticipating. There is the additional uncertainty owing to the gap between the time the order hits the market and the time when it is filled.

![Figure 2](image2.png)

Figure 1 and Figure 2 show the contrast between a long gamma and short gamma position revised using a price rule. We assume the market trades discretely, by periods $t_1$ through $t_4$. A long gamma will revised exactly at the intended price.
Assume $S$ reaches the price $S_u$, the revision price level at period $t_3$. The short gamma operator, provided he has an order in place in the market at that price, does not obtain $S_u$. A short gamma position will trigger an execution order at the next available price, which means that it will depend on $t_4$, conditional on $S_u$ having traded. Thus it requires a market order activated during $t_3$, thus obtaining $S_{t_4}$ plus proportional transaction costs\(^{11}\), here $S_{t_4} (1+k)$.

**1.6. An Order Flow Model for the Option.**

The representative agent MEDH will incur a long or short options that will be the exact inverse image to that of the static users ($q = -\sum_{i=1}^{n} w_i$, where $w_i$ are the holdings of the various non-dynamic hedgers).

**Assumption 5:** the dynamics of the underlying security are entirely independent from that of the option order flow. Hence the supply or demand for options is not linked to the price level in the underlying, which here is solely specified by its own variance $\sigma^2$.

We assume here for simplification only one option trading in the market, with one strike price and one single expiration $T$. The dynamics of the order flow for the option are

$$dq_t = m_{ord} \delta t + \sigma_{ord} dz$$

(7)

where $q_t$ is the accumulated net inventory in the option by the dynamic hedger. $\sigma_{ord}$ is the standard deviation of the order flow, $m_{ord}$ its expected drift, $z$ a standard Brownian motion.

The MEDH agents will have to carry in their inventory $q_t$ and the static hedger community the mirror quantity-$q_t$. Thus there will be a flow of quantities of options between both parties until the expiration of the option at period $T$. There is no need here for an equilibrium approach owing to assumption 5.

The MEDH starts with an inventory, by definition, $q_t$. He is asked to quote an option for expiration $T$. His bid (in implied volatility terms) on the option will be necessarily at the last $\sigma$ trade – as it would reduce his existing inventory and lower his transaction costs. He will sell the option at an equivalent $\sigma$ plus a markup that corresponds to the expectation of the reduction in inventory for that option during the life of the option. Figure 3 shows the “slices”, where (for clarity of the graph) it is assumed that the market trades in portions $q_t$. The dynamic hedger has 9 units and is asked to quote one additional one. It becomes natural to

\[^{11}\text{We also note a bias in the variance measurement as orders are left in the market at time } t_n \text{ triggered by price } S_n \text{ from observations that strictly precede } t_n. \text{ This will cause the variance of the process from a price based policy to be slightly biased upwards—but does not affect the mean.}\]
assume a LIFO system as the inventory \( q_1 \) will be liquidated first, and the inventory just acquired will be liquidated after that one. The quantity just acquired will come last, thus the expectation of keeping it until expiration is highest.

*Figure 3 “Slices” in Portfolio. We assume that options trade by blocks of equal size. The market maker just added \( Q \) to his portfolio; he will therefore have to dynamically hedge it until slices \( q_1 \) through \( q_9 \) are liquidated.*

The issue is to find the ratio \( \lambda \) of the markup over \( \sigma_A \), the adjustment policy under any rule.

*Proposition 1-4 Under a LIFO inventory policy and assumptions 1 to 4, the ratio of effective transaction costs for the dynamic hedger over the Leland adjustment is*

\[
\lambda = \frac{1}{kT} E\left( \int_0^T \min(q_s + q_s^*, 0) \, ds \right)
\]

*where*

\[
q_s^* = \sup\{q_s : 0 < \tau < s, q_{\tau} > q_0 \}
\]

More precisely the ratio \( \lambda \) can be calculated as follows:

\[
\lambda = \int_0^T \int_{q_0}^{q_s} k + q_s \varphi(q_s) \, dq_s \, ds
\]

*where \( \varphi(q_s) \) is the density of \( q_s^* \).*

The proof becomes simple once it is established that the operator is only interested in those future trades that contribute to the reduction of the inventory. A market maker, currently short options, will make in the future a series of bids and offers for options. The offer is no present concern in the computation of the future transaction costs for the existing inventory. It would increase the inventory and will be computed according to the same expected method, using the \( \lambda \) that will be calculated based on his inventory then. As to the bid, it will be placed at the break-even point. It is the expectation of the quantity that the trader will receive on the bid (until satisfaction, where the inventory becomes 0) during the life of the option that is of sole concern. Thus we are interested in the dynamics of \( q_s^* \), not \( q_s \).
Transaction Costs: Much Smaller than We Thought

The $\lambda$ can be applied to any revision policy provided the adjusted $\sigma$ is extracted from it. Thus we will price $\lambda$ of the option as $C(\sigma_A)$. The remaining $(1-\lambda)$ will be priced at $C(\sigma)$.

$$C(\sigma) + \lambda (\sigma_A - \sigma) \frac{\partial C}{\partial \sigma}$$

Figure 4 shows a simulation of the quantities of an option held by a dynamic hedger reducing his holding. Figure 5 illustrates the ratio over the Leland number.


Figure 4: Part to hedge. The dynamic hedger starts short -25 units of the option expiring in 30 days. According to the Leland adjustment he would need to dynamically hedge the -25 units all the time (the bottom dashed line). Here $q$, the accumulated order flow, follows a sample path that allows for the reduction of inventory every time it reaches a new high over the period.


Figure 5 The ratio $\lambda$. The vertical axis shows $Q$, the quantity of options held in the market maker inventory. The horizontal axis shows time. The part that would be hedged according to the Leland adjustment corresponds to the area of the rectangle $qT$.


12 Note that an option that is away from the money may require an additional second order term, $\frac{1}{2} \frac{\partial^2 C}{\partial \sigma^2} (\sigma_A - \sigma)^2$, owing to its convexity with respect to the volatility.
1.7. Concluding Comments

In this article we combined the Leland modification of the Black-Scholes equation with an economic approach. We based the analysis on three major distinctions: first, between specialized and non specialized efficient dynamic hedgers; second, between long and short gamma dynamic hedging; and third, between a time-based and a price-based adjustment policy.

We found that the relaxation of the infinitesimal revision assumption created a spate of microstructure issues, leading to a lack of asymmetry between long and short options. The extension of the length of time between revisions Δt can cause bankruptcy of a short option operator. The fact that a long option operator can use limit orders while a short option operator needs to use market orders is of a significant effect on both the optimal and feasible trading policies.

Finally, we showed that the dynamic hedger does not need to incur transaction costs for the entire period between the transaction and the expiration of the contract; instead he will only incur such costs 1) over the net inventory and 2) for a period of time possibly shorter than the nominal life of the contract. The importance of the dynamics of the order flow are larger than anticipated: it would be indeed rare for a dealer to only occasionally sell an option. The full Leland modification to the Black-Scholes equation need only to apply to the rare cases of option traders only selling options that are held until expiration.