WHY IS THE EFFECT OF PROPORTIONAL TRANSACTION COSTS $O(\delta^{2/3})$?

L.C.G.Rogers

University of Bath

This draft: April 2000

Summary. In the classical Davis-Norman problem for optimal investment/consumption under proportional transaction costs, the form of the optimal solution results in a surprisingly wide 'no transaction' region even for very small transaction costs δ. Shreve has shown that the width of the no-transaction zone is $O(\delta^{1/3})$ for coefficients of relative risk aversion between 0 and 1, and in this paper we present a proof of this using little more than stochastic calculus and properties of one-dimensional diffusions, for the case where the coefficient of relative risk aversion is any positive number different from 1.

Abbreviated title: Why is the effect of transaction costs $O(\delta^{2/3})$?

AMS Subject Classifications: 90A09, 60G35

Keywords and phrases: transaction costs, constant relative risk aversion, optimal investment, optimal consumption

WHY IS THE EFFECT OF PROPORTIONAL TRANSACTION COSTS $O(\delta^{2/3})$?

L.C.G.Rogers

University of Bath

1 Introduction.

The short answer to the question of the title is that the solution to the problem

$$\min_{x>0}\left\{\frac{\delta}{x}+ax^2\right\} \tag{1}$$

is to take

$$x=\left(\frac{\delta}{2a}\right)^{1/3},$$

which results in the minimised value of

$$3a^{1/3}(\delta/2)^{2/3}.$$

But why is this relevant? To set the scene, we shall be discussing the classical problem of proportional transaction costs, derived from the optimal consumption problem of Merton and set forth in Constantinides (1986), Davis & Norman (1990), and various other references; see Cadenillas (1999) for a recent survey of work on transaction costs problems. In this problem, an agent invests in two assets, a risky share and a riskless bank account, and consumes from the bank account so as to maximise his expected discounted utility of consumption, where the utility has constant relative risk aversion. Moves of wealth between the two assets incur a proportional cost δ.

The solution obtained by Davis & Norman (1990) is that there is a closed interval $I=I_*$ such that while the proportion of wealth p in the share remains in I_*, the agent makes no transactions, but when the proportion reaches the ends of the interval, then just enough trading takes place to stay within I_*. The proportion of wealth in the share obeys an autonomous SDE, and so is a diffusion. Now the optimal solution to the problem represents a balancing of two effects; on the one hand, making I large reduces the amount of transacting required (and therefore reduces the transaction costs), but on the other hand, by making I large we allow the proportion invested in the share to wander a long way from the optimal value, and so we lose because the portfolio is not appreciating as rapidly as it might.

Proceeding very informally, if we constrain p to some interval I of length x, then the proportional loss per unit time due to transaction costs will be δ times the average increase of the local time at the ends of the interval I for the diffusion p; we expect

this to be of the order of δ/x, which is exactly what we would get if the invariant distribution of p were uniform (as would be the case with Brownian motion). The proportional loss per unit time due to suboptimal portfolio composition will be of the order of x^2, assuming that I contains the optimal (Merton) proportion π_*. This becomes clear if we consider the (zero-transaction-costs) Merton problem, where the agent chooses to keep a fixed proportion π of wealth in the risky asset at all times, and will consume at a rate proportional to wealth, the constant of proportionality being chosen optimally given the value of π. The payoff to the agent is a smooth concave function of π, with a unique maximum at $\pi = \pi_*$. Assuming that $\pi_* \in I$ - which seems a natural condition - the reduction in the payoff if we used π in place of π_* must be to leading order proportional to $(\pi - \pi_*)^2$ - Taylor's theorem! This justifies (weakly) our assertion concerning the losses due to suboptimal portfolio composition. Putting these two effects together gives (1).

The result of this article is not new; Shreve (1995) uses estimates on viscosity solutions to derive (for the case where the coeffiecient of relative risk aversion is in $(0, 1)$) the order of magnitude we shall demonstrate, and Fleming, Grossman, Vila & Zariphopoulou (1990) in a somewhat different context obtain the same asymptotic. Thus we should view the current paper as primarily pedagogical in purpose; we shall obtain the bounds we need in the next section using little more than a judicious application of stochastic calculus. As befits a pedagogical approach, we will make various simplifying assumptions of an inessential nature which avoid tedious complications. In the end, we find an expression bounding the loss to leading order, which we conjecture is exact.

2 Bounds on the payoff.

We shall be considering the dynamics of the pair (x_t, y_t) in the form

$$dx_t = (rx_t - c_t)dt + (1-\delta)dM_t - dL_t \tag{2}$$
$$dy_t = y_t(\sigma dW_t + \mu dt) - dM_t + (1-\delta)dL_t, \tag{3}$$

where x_t is the value of the agent's holding in the riskless asset (bank account), bearing interest at constant rate r, and y_t is the value of the agent's holding in the risky asset (share). The increasing processes L (respectively, M) measure the cumulative amounts of money moved from bank account to share (respectively, from share to bank account), and the agent's problem is to choose these, along with the non-negative consumption process c is such a way as to keep the pair (x, y) always in the *solvency region*

$$\mathcal{S} \equiv \{(x, y) : x + (1-\delta)y \geq 0, y + (1-\delta)x \geq 0\},$$

while at the same time maximising the payoff

$$\Pi(c, L, M; x, y) \equiv E\Big[\int_0^\infty e^{-\rho t}U(c_t)dt | x_0 = x, y_0 = y\Big]. \tag{4}$$

The parameters σ, ρ, δ and μ are all constant, and are strictly positive (except for μ).

We shall assume that the utility function U has constant relative risk aversion:

$$U(x) = \frac{x^{1-R}}{1-R}$$

for some $R > 0$ different from 1. This assumption simplifies the solution considerably: defining the value function

$$V(x, y) \equiv \sup_{c,L,M} \Pi(c, L, M; x, y),$$

it is easy to show that for any positive λ

$$V(\lambda x, \lambda y) = \lambda^{1-R} V(x, y),$$

which thereby reduces the problem to one dimension. The solution of Davis & Norman (1990) is in terms of an interval $I_* \equiv [\pi_-, \pi_+]$ such that while the proportion $p_t \equiv y_t/(x_t+y_t)$ of wealth in the share remains in I_*, the agent makes no transactions, but when the proportion reaches the ends of the interval, then just enough trading takes place to keep p within I_*. No explicit expression can be found for the values π_-, π_+. The optimal (Merton) proportion

$$\pi_* \equiv \frac{\mu - r}{\sigma^2 R} \tag{5}$$

is often in the interval I_*, but examples can be constructed where it is not; see Shreve & Soner (1994).

The description of the optimal consumption process is not quite so simple; all one can say is that the optimal c is of the form

$$c_t^* = w_t g(p_t),$$

and Davis & Norman give a characterisation the function g. In his paper, Constantinides (1986) restricted attention to policies where the function g was assumed constant, and thereby derived lower bounds for the value. Although this assumption is substantive, it allowed him to deduce many of the key features of the solution, and to bound the size of the transaction cost effects, which he found to be be quite small.

We shall follow Constantinides in assuming that the consumption process is of the form $c_t = \gamma w_t$ for some constant γ, so that the wealth process $w_t \equiv x_t + y_t$ now satisfies the dynamics

$$dw_t = (r - \gamma) w_t dt + p_t w_t (\sigma dW_t + (\mu - r) dt) - \delta dA_t. \tag{6}$$

Here, $dA \equiv dL + dM$. Since we are interested in proving a lower bound for the payoff, we may (and shall) assume that the value of γ is the value which is optimal for the Merton (no-transaction-costs) problem:

$$\gamma = \frac{\rho + (R-1)(r + \frac{R}{2}\,\sigma^2\pi_*^2)}{R} \tag{7}$$

If $R > 1$, this is always positive, but if $R \in (0,1)$ we require ρ to be large enough that this is positive, else the optimisation problem is ill-posed.

We shall also suppose that two values p_- and p_+ have been picked so that no transactions occur while p_t is in the interior of $I \equiv [p_-, p_+]$, and trading occurs at the boundaries to keep the proportion p_t within I. We may (and shall) suppose that $p_- = \pi_* - \varepsilon$, $p_+ = \pi_* + \varepsilon$ for some (small) positive ε. We shall make a choice of the starting point $p_0 \in I$, and for simplicity of exposition we shall assume that

$$p_0 = \pi_*.$$

We can write down the solution to (6) explicitly:

$$w_t = w_0 \exp[\int_0^t \sigma p_s dW_s + \int_0^t \{(\mu - r)p_s - \frac{1}{2}\sigma^2 p_s^2\}ds + (r-\gamma)t - \delta a_t],$$

where $da_t = dA_t/w_t$. This therefore allows us to express the payoff as

$$\begin{aligned}
\Pi &= U(\gamma w_0)\; E\Big[\int_0^\infty \exp\{(1-R)\int_0^t \sigma p_s dW_s + (1-R)\int_0^t \{(\mu-r)p_s - \frac{1}{2}\sigma^2 p_s^2\}ds \\
&\qquad\qquad +(1-R)(r-\gamma)t - (1-R)\delta a_t\} e^{-\rho t} dt\Big] \\
&= U(\gamma w_0)\; E\Big[\int_0^\infty Z_t \exp\{(1-R)\int_0^t \{(\mu-r)p_s - \frac{R}{2}\sigma^2 p_s^2\}ds \\
&\qquad\qquad +(1-R)(r-\gamma)t - (1-R)\delta a_t\} e^{-\rho t} dt\Big] \\
&\geq U(\gamma w_0)\; E\Big[\int_0^\infty Z_t \exp\{-\alpha t - (1-R)\delta a_t\} dt\Big]
\end{aligned} \tag{8}$$

where

$$Z_t \equiv \exp((1-R)\int_0^t \sigma p_s dW_s - \frac{1}{2}\int_0^t (1-R)^2\sigma^2 p_s^2 ds)$$

is a positive martingale, and

$$\begin{aligned}
-\alpha &\equiv -\rho + (1-R)\min_{z\in I}\{r - \gamma + (\mu - r)z - \frac{R}{2}\sigma^2 z^2\}. \\
&= -\rho + (1-R)\Big[r - \gamma + \frac{\sigma^2 R}{2}\,(\pi_*^2 - \varepsilon^2)\Big] \\
&\equiv -\alpha_0 - (1-R)\frac{\sigma^2 R}{2}\,\varepsilon^2.
\end{aligned} \tag{9}$$

Now we know that the effect of the martingale Z is to change the measure with respect to which we take expectation; if P^* is the new measure, defined by

$$\left.\frac{dP^*}{dP}\right|_{\mathcal{F}_t} = Z_t,$$

then under P^*,

$$dW_t^* \equiv dW_t - (1-R)\sigma p_t dt \tag{10}$$

is a Brownian motion. Thus the bound(8) can be re-expressed as

$$\Pi \geq U(w_0)\; E^*\left[\int_0^\infty \exp\{-\alpha t - (1-R)\delta a_t\}dt\right], \tag{11}$$

and what remains is to understand the term involving a_t in (11). For that we consider the process p, which can be shown to solve the stochastic differential equation

$$\begin{aligned} dp &= \sigma p(1-p)\; dW + p\{(\mu-r)(1-p) + \gamma - \sigma^2 p(1-p)\}\; dt + \kappa_- dL/w - \kappa_+ dM/w \\ &= \sigma p(1-p)\; dW^* + p\{(\mu-r)(1-p) + \gamma - \sigma^2 p(1-p) + (1-R)\sigma\}\; dt \\ &\qquad +\kappa_- dL/w - \kappa_+ dM/w \\ &= \sigma p(1-p)\; dW^* + b(p)dt + \kappa_- dL/w - \kappa_+ dM/w \end{aligned} \tag{12}$$

where

$$\begin{aligned} \kappa_+ &= 1 - p_+\delta \\ \kappa_- &= 1 - (1-p_-)\delta \\ b(p) &\equiv p\{(\mu-r)(1-p) + \gamma - \sigma^2 p(1-p) + (1-R)\sigma\} \end{aligned}$$

Thus under P^* the process p is an autonomous diffusion with bounded drift, reflected at the ends of the interval I, and the local time processes at the endpoints are

$$dl_-(t) = \kappa_- dL(t)/w(t), \quad dl_+(t) = \kappa_+ dM(t)/w(t).$$

If $\mathcal{G}$ denotes the generator of the diffusion p, and ψ_α^+ (respectively, ψ_α^-) is the increasing (respectively, decreasing) solution to

$$(\alpha - \mathcal{G})f = 0,$$

then the expectation in (11) can be expressed as

$$\begin{aligned} F(x;\alpha) &\equiv E^*\left[\int_0^\infty \exp\{-\alpha t - (1-R)\delta a_t\}dt | p_0 = x\right] \\ &= E^*\left[\int_0^\infty \exp\{-\alpha t - \beta_- l_-(t) - \beta_+ l_+(t))\}dt | p_0 = x\right] \\ &= \alpha^{-1} + \gamma_-\psi_\alpha^+ + \gamma_+\psi_\alpha^-, \end{aligned} \tag{13}$$

where $\beta_\pm = (1-R)\delta/\kappa_\pm$ and the constants $\gamma_\pm$ are determined from the differential equation satisfied by F:

$$\begin{aligned} \mathcal{G}F - \alpha F &= -1, && (14)\\ F'(p_-) &= \beta_- F(p_-), && (15)\\ F'(p_+) &= -\beta_+ F(p_+). && (16) \end{aligned}$$

While these can be solved explicitly for $\gamma_\pm$, it is not really necessary to write out the solution explicitly. Notice that in the case $R > 1$, for large enough δ the expectation in (11) will be infinite. However, there is always an interval of δ-values around 0 in which the expectation is finite, which is sufficient for our purposes since we wish to look at limiting behaviour as $\delta \downarrow 0$. We end up with an expression for $F(\pi_*)$ which depends explicitly on the small parameters ε and δ, and we may determine to first order how this expression varies with those small parameters. The partial derivative with respect to α of $F(\pi_*, \alpha)$ at $\delta = 0$ and $\alpha = \alpha_0$ is $-1/\alpha_0^2$, of course. The partial derivative with respect to δ of $F(\pi_*, \alpha)$ at $\delta = 0$ and $\alpha = \alpha_0$ is less obvious, but is quickly obtained using Maple (the worksheet used to do the calculations of this paper is available from the author on request); it is also quickly expanded in powers of ε to give to leading order

$$-\frac{\sigma^2 p^2 (1-p)^2 (1-R)}{2\alpha_0^2 \varepsilon} \qquad (17)$$

after some rearrangement. Thus to leading order in (ε, δ), the lower bound (11) for the payoff is

$$U(\gamma w_0)\left[\frac{1}{\alpha_0} - \frac{(1-R)\sigma^2 R\varepsilon^2}{2\alpha_0^2} - \frac{\sigma^2 \pi_*^2 (1-\pi_*)^2 (1-R)\delta}{2\alpha_0^2\,\varepsilon}\right],$$

using (9) and (17). The loss is thus

$$(\gamma w_0)^{1-R}\frac{\sigma^2}{2\alpha_0^2}\left[R\varepsilon^2 + \frac{\pi_*^2(1-\pi_*)^2\delta}{\varepsilon}\right],$$

which was the form we declared at (1).

References

CADENILLAS, A., 2000, Consumption-investment problems with transaction costs: survey and open problems, *Mathematical Methods of Operations Research* **51**, 43–68.

CONSTANTINIDES, G. M., 1996, Capital market equilibrium with transaction costs, *Journal of Political Economy* **94**, 842–864.

DAVIS, M. H. A., and NORMAN, A., 1990, Portfolio selection with transaction costs, *SIAM Journal of Control and Optimisation* **31**, 470–493.

FLEMING, W. H., GROSSMAN, S. G., VILA, J.-L., and ZARIPHOPOULOU, T., 1990, Optimal portfolio rebalancing with transaction costs, Preprint.

SHREVE, S. E. and SONER, H. M., 1994, Optimal investment and consumption with transaction costs, *Annals of Applied Probability* **4**, 609–692

SHREVE, S. E., 1995, Liquidity premium for capital asset pricing with transaction costs, *Mathematical Finance IMA Volume 65*, M. H. A. Davis, D. Duffie, W. H. Fleming & S. E. Shreve (editors), Springer, New York.

University of Bath, School of Mathematical Sciences, Bath BA2 7AY, Great Britain (e-mail: lcgr@maths.bath.ac.uk)