The statistics of ‘statistical arbitrage’ in stock markets

Robert Fernholz
INTECH
One Palmer Square
Princeton, NJ 08542
bob@enhanced.com

Cary Maguire, Jr.
INTECH
One Palmer Square
Princeton, NJ 08542
camm@enhanced.com

December 20, 2006

Abstract

Hedge funds sometimes use mathematical techniques to “capture” the short-term volatility of stocks, or perhaps other types of securities. This type of strategy is sometimes referred to as *statistical arbitrage*, and we show that in the universe of large-capitalization U.S. stocks, even quite naive techniques can achieve remarkably high information ratios. Here we present a simple descriptive analysis of these strategies, and their relationship with equity trading costs.

Key words: stochastic portfolio theory, statistical arbitrage, equal-weighted portfolio.
JEL classification: G10.

1 Introduction

Market makers in financial markets generate profits by buying low and selling high over short time intervals. This occurs naturally, since as market makers they will offer a stock for sale at a higher price then they are willing to pay for it, and more urgent buyers and sellers will have to accept their terms. Market making, particularly that of New York Stock Exchange (NYSE) specialists, has been studied in the normative context of academic finance, and this approach is represented by the seminal papers of Hasbrouck and Sofianos (1993) and Madhavan and Smidt (1993).

High-speed trading strategies similar to market making have putatively been used by hedge funds in recent years, and this type of strategy has sometimes been referred to as *statistical arbitrage*, or, perhaps, “stat-arb” in the abbreviated patois of the Street. Statistical arbitrage of this nature can be studied in the context of portfolio behavior, and hence is amenable to the methods of stochastic portfolio theory, Fernholz (2002). Here we apply these methods to examine the potential profitability of such a strategy applied to large-capitalization U.S. stocks.

Equal-weighted portfolios are dynamic portfolios in which all the stocks have the same constant weights. Equal-weighted portfolios are of interest here since in such a portfolio, if a stock rises in price relative to the others, this will generate a sell trade, and if the price declines this will generate a buy. Hence, such a portfolio will sell on up-ticks and buy on down-ticks, the way a market maker would. Here we estimate the return and risk parameters of equal-weighted portfolios, and with these parameters we can determine the efficacy of statistical arbitrage in a stock market.

In the next section we begin by presenting some of the basic ideas of stochastic portfolio theory. In the subsequent sections we use these methods to analyze the behavior of large-capitalization U.S. stocks. We then propose a hedging strategy to control the risk of the high-speed trading, and we determine the information ratio of the hedged strategy. Mathematical proofs will not be included here, but can be found in Fernholz (2002), and a mathematical discussion of "trading noise" is included in an appendix.

2 An introduction to stochastic portfolio theory

Consider a market of n stocks represented by *stock price processes* $X_1, \dots, X_n$ that satisfy the stochastic differential equations

$$d \log X_i(t) = \gamma_i(t)\,dt + \sum_{\nu=1}^{d} \xi_{i\nu}(t)\,dW_\nu(t), \quad t \in [0, \infty), \tag{1}$$

for $i = 1, \dots, n$, where $(W_1, \dots, W_d)$ is d-dimensional Brownian motion with $d \geq n$, and the processes γ_i and $\xi_{i\nu}$ are measurable and adapted (which means that they do not depend on future events) and satisfy certain regularity conditions (see Fernholz (2002), Definition 1.1.1). A stochastic process of the form (1) is called a *continuous semimartingale*, and such processes are discussed in detail in Karatzas and Shreve (1991). The value $X_i(t)$ represents the price of the ith stock at time t, and we shall assume that there is a single share of stock outstanding for each company, so $X_i(t)$ represents the total capitalization of the ith company at time t. In (1), $d \log X_i(t)$ represents the *log-return* of X_i over the (instantaneous) time period dt. The process γ_i in (1) is called the *growth rate process* for X_i, and the process $\xi_{i\nu}$ measures the sensitivity of X_i to the νth source of uncertainty, W_ν.

The *covariance process* for X_i and X_j is given by

$$\sigma_{ij}(t) = \sum_{\nu=1}^{d} \xi_{i\nu}(t)\xi_{j\nu}(t), \quad t \in [0, \infty),$$

with the notation $\sigma_i^2(t) = \sigma_{ii}(t)$ for the *variance* processes. It is commonly assumed that the matrix $(\sigma_{ij}(t))$ is nonsingular, and we shall do so here. It would not be difficult to include dividend processes in our model, but for simplicity we shall assume here that stocks pay no dividends.

Equation (1) is the *logarithmic representation* of the stock price processes, and this representation is further developed in Section 1.1 of Fernholz (2002). With the logarithmic representation we consider the growth rate γ_i, which can be interpreted as the expected rate of change of the logarithm of the stock price at time t. The logarithmic representation is equivalent to the usual arithmetic representation commonly used in mathematical finance (see, e.g., Karatzas and Shreve (1998)), which in this case will be

$$\frac{dX_i(t)}{X_i(t)} = \alpha_i(t)\,dt + \sum_{\nu=1}^{d} \xi_{i\nu}(t)\,dW_\nu(t), \quad t \in [0, \infty), \tag{2}$$

where the process α_i is called the *rate of return*, and is related to the growth rate by

$$\alpha_i(t) = \gamma_i(t) + \frac{\sigma_i^2(t)}{2}, \quad t \in [0, \infty), \quad \text{a.s.}$$

This equation follows from Itô's rule (see Karatzas and Shreve (1991)), and is discussed in Section 1.1 of Fernholz (2002). The expression on the left of (2) is sometimes referred to as the *instantaneous return* on X_i at time t. Our use of the logarithmic representation makes no assumption that we wish to maximize logarithmic utility, or, for that matter, any utility function at all. The reason that we use the logarithmic representation is because it brings to light certain aspects of portfolio behavior that remain obscure with the conventional arithmetic representation. Let us now consider portfolios, and their growth rates and variances.

We represent a portfolio π by its *weight processes* $\pi_1, \ldots, \pi_n$ where $\pi_i(t)$ is the proportion of the portfolio invested in X_i at time t. The weight processes are assumed to be adapted and bounded, and must sum to one, so we have $\pi_1(t) + \cdots + \pi_n(t) = 1$, a.s., for all t. A negative value of $\pi_i(t)$ indicates a short sale in X_i.

If we let $Z_\pi(t) > 0$ represent the value of π at time t, then, with the initial value $Z_\pi(0)$, Z_π will satisfy the stochastic differential equation

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \sum_{i=1}^n \pi_i(t) \frac{dX_i(t)}{X_i(t)}, \quad t \in [0, \infty). \tag{3}$$

(Please note that with this definition of the portfolio value process, the portfolio will be *self-financing* in the sense of, e.g., Duffie (1992).) Equation (3) shows that the instantaneous return on the portfolio π is the weighted average of the instantaneous returns of the stocks in the portfolio.

We can apply Itô's rule (see Fernholz (2002), Corollary 1.1.6) to express (3) in logarithmic form as

$$d \log Z_\pi(t) = \sum_{i=1}^n \pi_i(t)\, d \log X_i(t) + \gamma_\pi^*(t)\, dt, \quad t \in [0, \infty), \quad \text{a.s.}, \tag{4}$$

where

$$\gamma_\pi^*(t) = \frac{1}{2}\Big(\sum_{i=1}^n \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^n \pi_i(t)\pi_j(t)\sigma_{ij}(t) \Big) \tag{5}$$

is called the *excess growth rate* process. The last summation on (5) represents the portfolio variance rate

$$\sigma_{\pi\pi}(t) = \sum_{i,j=1}^n \pi_i(t)\pi_j(t)\sigma_{ij}(t), \quad t \in [0, \infty),$$

and this expression is identical to the portfolio variance rate in the arithmetic representation. We can also define the covariance process of the ith stock with the portfolio π by

$$\sigma_{i\pi}(t) = \sum_{j=1}^n \pi_j(t)\sigma_{ij}(t), \quad t \in [0, \infty).$$

From (5) we see that the excess growth rate is one half the difference of the weighted average of the stock variances minus the portfolio variance, and it is shown in (8) below that the excess growth rate $\gamma_\pi^*(t)$ is always positive for a portfolio with no short sales, at least if the portfolio holds more than a single stock. In this sense the excess growth rate measures the efficacy of diversification in reducing the portfolio risk. However, the excess growth rate measures more than the reduction of portfolio risk. From (1) and (4), we see that the *portfolio growth rate* γ_π will satisfy

$$\gamma_\pi(t) = \sum_{i=1}^n \pi_i(t)\gamma_i(t) + \gamma_\pi^*(t), \quad t \in [0, \infty), \quad \text{a.s.},$$

so at a given time t, π will have a higher growth rate than the weighted average of the growth rates of its component stocks. Hence, superior diversification not only reduces risk, but also increases the growth rate of a portfolio.

Let us suppose that η is a given portfolio. We can define the *relative covariances* of the stocks with respect to η by

$$\tau^{\eta}_{ij}(t) = \sigma_{ij}(t) - \sigma_{i\eta}(t) - \sigma_{j\eta}(t) + \sigma_{\eta\eta}(t), \quad t \in [0,\infty),$$

for $i, j = 1, \ldots, n$. For any portfolio π, we also have the relative variance and covariances

$$\tau^{\eta}_{\pi\pi}(t) = \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\tau^{\eta}_{ij}(t) \quad \text{and} \quad \tau^{\eta}_{i\pi}(t) = \sum_{j=1}^{n} \pi_j(t)\tau^{\eta}_{ij}(t), \quad t \in [0,\infty),$$

for $i = 1, \ldots, n$. It is not difficult to verify that

$$\tau^{\eta}_{\eta\eta}(t) = 0, \quad t \in [0,\infty), \quad \text{a.s.,} \tag{6}$$

and that

$$\gamma^*_{\pi}(t) = \frac{1}{2}\Big(\sum_{i=1}^{n} \pi_i(t)\tau^{\eta}_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\tau^{\eta}_{ij}(t)\Big), \quad t \in [0,\infty), \quad \text{a.s.} \tag{7}$$

This last equation demonstrates the *numeraire invariance* of the excess growth rate, i.e., that it is the same whether measured absolutely or measured relative to an arbitrary portfolio η. In the particular case that the portfolio π is the same as the portfolio η, (7) becomes

$$\gamma^*_{\pi}(t) = \frac{1}{2}\sum_{i=1}^{n} \pi_i(t)\tau^{\pi}_{ii}(t), \quad t \in [0,\infty), \quad \text{a.s.,} \tag{8}$$

since by (6) the relative portfolio variance term in (7) vanishes. If π holds more than a single stock, then the nonsingularity of $(\sigma_{ij}(t))$ implies that all the relative variances $\tau^{\pi}_{ii}(t)$ are positive, so $\gamma^*_{\pi}(t)$ must also be positive.

3 Portfolios with constant weights

Let us consider a *constant-weighted portfolio* π in which $\pi_i(t) = p_i$ for all t and $i = 1, \ldots, n$, where the p_i are positive constants such that $p_1 + \cdots + p_n = 1$. In this case, (4) becomes

$$\begin{aligned} d\log Z_{\pi}(t) &= \sum_{i=1}^{n} p_i \, d\log X_i(t) + \gamma^*_{\pi}(t)\,dt \\ &= d\log\big(X_1^{p_1}(t)\cdots X_n^{p_n}(t)\big) + \gamma^*_{\pi}(t)\,dt, \quad t \in [0,\infty), \quad \text{a.s.} \end{aligned} \tag{9}$$

By its very nature, the constant-weighted portfolio π will sell a stock when its price rises relative to the rest of the portfolio, and buy when the price declines. From (9) we see that over the period $[0, t]$ the log-return on π will be

$$\log\big(Z_{\pi}(t)/Z_{\pi}(0)\big) = \log\big(X_1^{p_1}(t)\cdots X_n^{p_n}(t)\big) - \log\big(X_1^{p_1}(0)\cdots X_n^{p_n}(0)\big) + \int_0^t \gamma^*_{\pi}(s)\,ds, \quad \text{a.s.} \tag{10}$$

The particular type of constant-weighted portfolio that interests us here is the case in which π is *equal-weighted*, i.e., with $\pi_1(t) = \cdots = \pi_n(t) = 1/n$ for all t. For an equal-weighted portfolio π, (10) becomes

$$\log\big(Z_{\pi}(t)/Z_{\pi}(0)\big) = \log \sqrt[n]{X_1(t)\cdots X_n(t)} - \log \sqrt[n]{X_1(0)\cdots X_n(0)} + \int_0^t \gamma^*_{\pi}(s)\,ds, \quad \text{a.s.} \tag{11}$$

Hence, the log-return over the period $[0, t]$ depends only on the change in the geometric mean of the stock prices (i.e., capitalizations) and the accumulated excess growth.

Our plan is to use an equal-weighted portfolio to act as a market maker, and thus exploit the short-term volatility generated by the difference in price between buy and sell trades. In order to determine the efficacy this strategy, we must estimate the value of $\gamma_\pi^*(t)$ in (8), and we shall do so in the next section.

It is important to remember that real stock portfolios cannot have precisely constant weights because it is impossible to trade the stocks back to the weights as fast as the stock prices move. Instead, real portfolios are rebalanced at discrete time intervals to bring the portfolio weights back to weights that are nearly constant. Nevertheless, it is widely known and accepted in mathematical finance that stochastic integrals of the form (10) and (11) provide a close approximation to discrete trading strategies for real portfolios as long as the trading intervals are short enough (see Fouque et al. (2000)). In practice, portfolio strategies with trading intervals of a month or less can be closely approximated by continuous models of the form (10) and (11) (see Fernholz (2002)). However, it is essential for the parameters in a continuous model to be compatible with the trading intervals used in the strategy, and we shall develop this idea in the next section.

4 The estimation of excess growth rates

In this section we shall estimate the excess growth rate for an equal-weighted portfolio of large-capitalization U.S. stocks, specifically, those stocks included in either the S&P 500 Index or the Russell 1000 Index. These stocks are traded on the NYSE or NASDAQ, and the data we use are the individual transaction prices, or "prints", of the trades that occur on these exchanges each day. From (8) it is apparent that the estimation of excess growth depends on the estimation of the (relative) variance, since for an equal-weighted portfolio (8) shows us that γ_π^* is simply one half the average of the relative variances τ_{ii}^π of the stocks in the portfolio. To estimate these relative variances, it is necessary to sample the stock-price time series. To sample a time series, a particular length of sampling interval must be chosen, and the value of the variance estimate may depend on this choice (see, e.g., Fouque et al. (2000) for a discussion of this). This is particularly likely if there is trading noise that may be caused by differences in price between buy and sell trades. Under these circumstances, estimates of daily stock variance will depend on the sampling interval used, with shorter sampling intervals usually generating higher estimates of daily variance than longer sampling intervals (this is discussed in the appendix).

Suppose that we let V_T represent the average daily relative variance of the stocks in the equal-weighted portfolio, estimated by sampling at intervals of length T the time series for the relative return of each stock compared to the equal-weighted portfolio. The graph of V_T versus T is a form of *variogram*, and a discussion of this type of analytical methodology can be found in Fouque et al. (2000). In Figure 1 we see a variogram for the year 2005, and it shows that the estimate of daily variance V_T decreases from $V_{1.5} = .000273$ for a 1.5-minute sampling interval to $V_{390} = .000169$ for a 390-minute interval, which is the total number of minutes the exchanges in the U.S. are open each business day. By (8), these variances correspond to daily excess growth rates of about .0137% for a sampling interval of 1.5 minutes, and about .0085% for a sampling interval of 390 minutes. To annualize these numbers we must multiply by 250, the number of trading days in 2005. With this multiple, we have annual excess growth rates of about 3.41% for the 1.5-minute sampling interval and about 2.11% for the 390-minute sampling interval. Figure 1 indicates that sampling intervals between 1.5 and 390 minutes will produce estimated excess growth rates between the two values we have seen here.

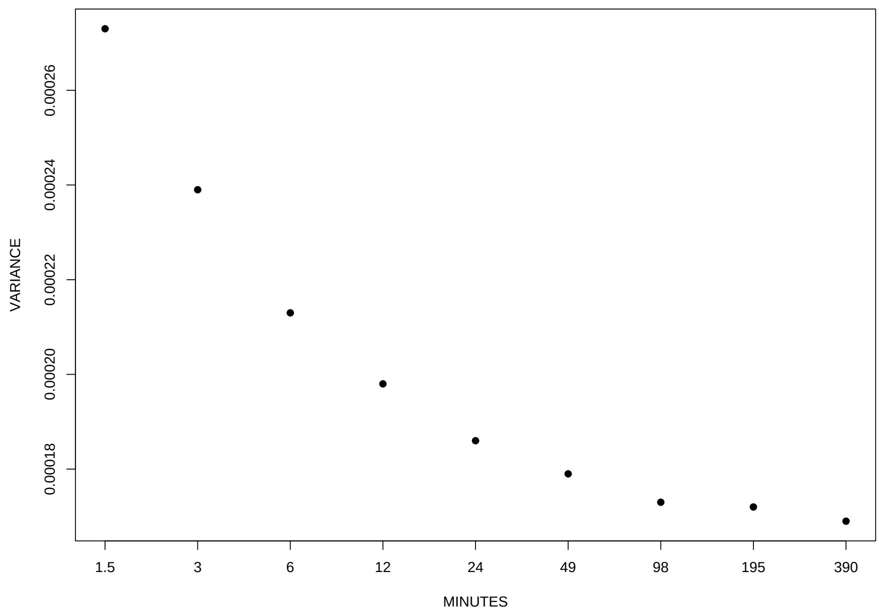

Figure 1: Estimated average daily V_T for the year 2005, with sampling intervals of $T = 1.5, 3, 6, \ldots, 390$ minutes.

With estimates of annual excess growth that range from 2.11% to 3.41%, which value is appropriate for use in (11)? The appropriate estimate corresponds to the sampling interval closest to the interval at which the portfolio π will be rebalanced back to equal weights. Hence, for a high-speed trader who trades continuously, the 1.5 minute interval is appropriate, while for a trader who trades only at the open and close of the market, the 390 minute interval should be used. It follows that the high-speed trader will generate 3.41% of excess growth a year, while the open/close trader will generate only 2.11%. More frequent trading allows the portfolio to capture more volatility.

5 A hedged strategy

An investment in an equal-weighted portfolio can be risky, because although the γ_π^* term in (11) is monotonically increasing in t, the log $\sqrt[n]{X_1(t)\cdots X_n(t)}$ term can vary quite a bit. However, if we consider two equal-weighted portfolios, each rebalanced at different intervals, the log $\sqrt[n]{X_1(t)\cdots X_n(t)}$ terms should be about equal, whereas we have seen above that there can be a systematic difference in the γ_π^* terms.

Consider two investments, one in an equal-weighted portfolio π that is rebalanced at intervals of about 1.5 minutes, and one in an equal-weighted portfolio η that is rebalanced only at the open and close of the market, i.e., at an interval of 390 minutes. Suppose that we buy \$100 of π and sell \$100 of η short. This long/short combination of portfolios is similar to the position a market maker takes, since a market maker need not hold a large inventory of the stocks he trades, and will probably want to close out his positions at the end of the day. The portfolio π will buy on downticks and sell on upticks, but over the period of a trading day, it will not move far from the short portfolio η. This will keep the total long/short position close to market-neutral during the trading day, and at the close, both π and η are rebalanced back to equal weights, so their positions are neutralized.

The value of each of these investments can be approximated by an equation of the form (11),

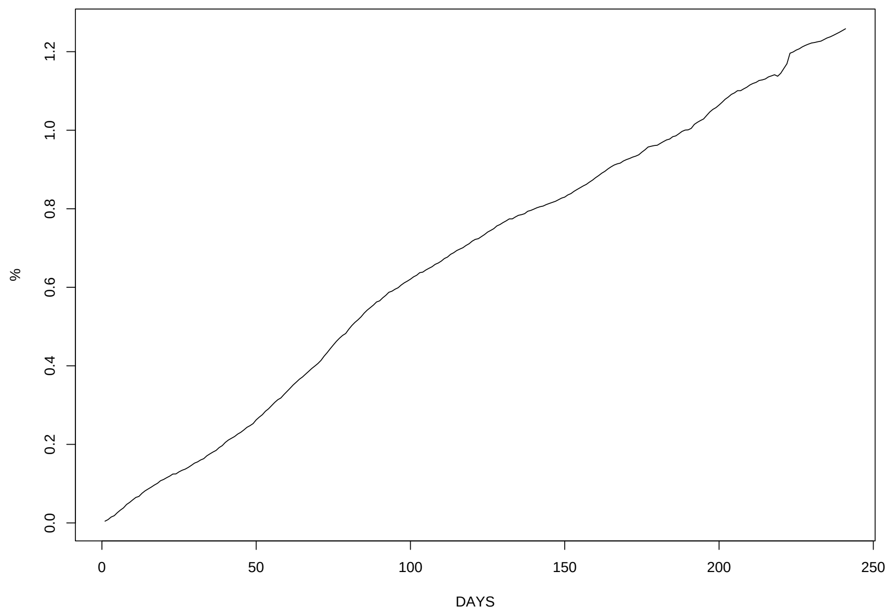

Figure 2: Simulated value of $\log\big(Z_\pi(t)/Z_\pi(0)\big) - \log\big(Z_\eta(t)/Z_\eta(0)\big)$ for the year 2005.

where the value of $\gamma^*_\pi(t)$ will correspond to a variance estimate based on a sampling interval of 1.5 minutes, and $\gamma^*_\eta(t)$ will correspond to a variance estimate based on a sampling interval of 390 minutes. In this case we have $\gamma^*_\pi(t) \cong 3.41\%$, while $\gamma^*_\eta(t) \cong 2.11\%$, so over a year the hedged portfolio, which is long \$100 of π and short \$100 of η, will accumulate about \$1.30.

In Figure 2 we see the log-return of a simulated continually-rebalanced equal-weighted portfolio versus an equal-weighted portfolio which is rebalanced only at the open and the close of the market each day. The simulation was run over the year 2005, the same year as we used for estimating the excess growth rates. (Figure 2 has only 241 days of performance because no tick data were collected for the nine days from October 24 through November 3 due to office closure resulting from Hurricane Wilma.) We see from Figure 2 that relative log-return was about 1.26%, which for the reduced 241-trading-day year is consistent with the 1.30% annual log-return calculated in the previous paragraph. The mean and standard deviation for the daily log-returns in Figure 2 were about .0052% and .0026%, respectively, for a daily information ratio of about 2.0. This implies an annual information ratio of about 32 for the long/short combination. In real trading, it might be difficult for a high-speed trader to actually trade every minute and a half, so a correction may be needed to adjust for attainable order flow, and trading commissions could come into play if the high-speed trader was not a broker/dealer or specialist. However, a trading strategy with an annual information ratio of just 3 would produce a negative year only about once every seven centuries, at least if log-returns followed a normal distribution. (No wonder specialists' seats on the NYSE were hereditary!)

We would like to estimate the total amount high-speed trading in large-capitalization stocks that the U.S. markets could support. For a stock with a daily relative variance of .000273 the corresponding 1.5-minute relative standard deviation will be about $\sigma = .00103$. With this value of σ, the expected relative log-price change over a minute and a half will be about $\pm\sqrt{2/\pi}\sigma = \pm.000824$. In a trading day there are about 256 intervals of a minute and a half, so the expected cumulative absolute change in the log-prices corresponding to these trades will be about 21.1%. Hence, the daily trading on a \$100 equal-weighted portfolio that is traded about every minute and a half will amount to approximately \$21.10.

Let us say that current daily volume on the NYSE and NASDAQ is around 2 billion shares in each market and that the average share price is \$30, so about \$120 billion of stocks are traded a day, and let us suppose that of this \$120 billion of stocks traded, \$100 billion is in large-capitalization stocks. Hence, if a \$100-dollar equal-weighted portfolio trades about \$21.10 a day, and if high-speed traders with equal-weighted portfolios were on one side or the other of every trade involving a large-capitalization stock, then the markets could support about \$474 billion of such portfolios. (Of course, the total holdings for long/short portfolio combinations would be far less than \$474 billion, due to cancellation between the long positions and the short positions.) At the rate of \$1.30 a year for each \$100, the total annual income generated by these high-speed long/short portfolios would be about \$6.2 billion.

The trading cost due to market impact (i.e., excluding commissions) for U.S. stocks is estimated to be about 11 basis points in Plexus (2006) (and is referred to as "Broker Impact" in that publication). If we assume \$100 billion volume a day over 250 trading days in the year, at 11 basis points the total trading cost due to market impact will be about \$27.5 billion. Why is this number so much larger than the \$6.2 billion we calculated for our high-speed trader above? First, since our trader had to pay 2.11% for the hedging portfolio, he collected only 1.30% of the 3.41% a year generated by his long portfolio. However, the 2.11% hedging expense of a high-speed market maker does not go to the market maker's trading counterparties, so they will pay the full 3.41% as trading cost, which means that the total market trading cost will be about \$16.1 billion rather than just the \$6.2 billion profit generated by high-speed market makers. Second, if the stock price fails to revert back to its previous value the high-speed traders will earn nothing, but all the price changes involved would be considered trading cost by the counterparties, whether or not the price reverts. Since short-term price reversion probably occurs more than half the time (considering Figure 1), but not always, trading costs will be somewhere between \$16.1 billion and \$32.2 billion, which is consistent with the \$27.5 billion we estimated above.

For simplicity, we have used equal-weighted portfolios in our example, and even with this naive strategy, the information ratio indicates that a high degree of leverage would be possible. Similar analysis can be carried out for any constant-weighted portfolio, and in this case some type of optimization could be applied to the weights. For example, weights on more volatile stocks might be increased to generate greater γ_π^*, or decreased to control portfolio risk, and the weights could be varied from day to day to adjust for changing market conditions. Optimization of the long/short portfolio could perhaps eliminate much of the 2.11% hedging expense, and perhaps the mean-reversion evident in Figure 1 could also be exploited in some manner (see appendix). The discussion in the previous paragraph indicates that with appropriate optimization, high-speed trading strategies might be able to generate several times our naive estimate of \$6.2 billion a year.

6 Conclusion

We have shown that dynamic portfolios in which the portfolio holdings are systematically rebalanced to maintain almost constant weights will capture in a natural way the volatility that is present in stock markets. These portfolios act as market makers by selling on upticks and buying on downticks. We have shown that, in the absence of trading commissions, such portfolios can generate remarkably high information ratios.

Acknowledgements: The authors thank Jason Greene for his many helpful suggestions and thank Joseph Runnels for supplying essential data on trading costs.

Appendix: Variance estimation with trading noise

The model for a stock price X with trading noise that we propose here is

$$\log X(t) = Y(t) + Z(t),$$

where Y is a random walk that satisfies

$$dY(t) = \sigma\, dW_1(t),$$

with $\sigma > 0$ constant and W_1 a Brownian motion, and where Z is an *Ornstein-Uhlenbeck noise process* (see Karatzas and Shreve (1991)) that satisfies

$$dZ(t) = -\alpha Z(t)\, dt + \tau\, dW_2(t), \tag{12}$$

with a, τ positive constants and W_2 a Brownian motion independent of W_1. In this case, the representation (1) becomes

$$d \log X(t) = -\alpha Z(t)\, dt + \sigma\, dW_1(t) + \tau\, dW_2(t), \tag{13}$$

and this price process satisfies the conditions for (1). The noise process Z incorporates a "restoring force" $-\alpha Z(t)$ that returns the process to the origin. The steady-state distribution for Y is normal with mean 0, and the covariance of $Z(s)$ and $Z(t)$ is $(\tau^2/2\alpha)e^{-\alpha|t-s|}$, for $s, t \geq 0$ (see Karatzas and Shreve (1991)).

In (13) the variance rate for X is seen to be $\sigma^2 + \tau^2$, but to estimate this rate, it is necessary to sample the time series for $\log X(t)$ at regular intervals. Let us assume that at time 0 we start the process Y at the origin and the process Z at its steady-state distribution. If for $0 \leq s < t$ we sample the time series for $\log X$ at s and t, then

$$\begin{aligned}
\mathrm{Var}\big(\log X(t) - \log X(s)\big) &= \mathrm{E}\big(\log X(t) - \log X(s)\big)^2 \\
&= \mathrm{E}\big(Y(t) - Y(s)\big)^2 + \mathrm{E}\big(Z(t) - Z(s)\big)^2 \\
&= \sigma^2(t-s) + \frac{\tau^2}{\alpha}\big(1 - e^{-\alpha(t-s)}\big).
\end{aligned}$$

Hence, the expected value of the estimated variance parameter for $\log X$ using sampling intervals of length $T > 0$ will be

$$V_T = \frac{1}{T}\Big(\sigma^2 T + \frac{\tau^2}{\alpha}\big(1 - e^{-\alpha T}\big)\Big) = \sigma^2 + \tau^2\Big(\frac{1 - e^{-\alpha T}}{\alpha T}\Big),$$

with the limiting value $V_0 = \lim_{T \to 0} V_T = \sigma^2 + \tau^2$, and decreasing asymptotically toward σ^2 as T increases toward infinity.

Trading noise that can be modeled by a system of the form (12,13) is likely to be amenable to methods of prediction, in particular, the use of Kalman filters (see Lewis (1986)). With prediction, it might be possible to improve upon the strategy outlined in Section 5.

References

Duffie, D. (1992). *Dynamic Asset Pricing Theory*. Princeton, NJ: Princeton University Press.

Fernholz, R. (2002). *Stochastic Portfolio Theory*. New York: Springer-Verlag.

Fouque, J.-P., G. Papanicolaou, and K. R. Sircar (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge: Cambridge University Press.

Hasbrouck, J. and G. Sofianos (1993). The trades of market makers: an empirical analysis of NYSE specialists. *Journal of Finance 48*, 1565–1593.

Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

Karatzas, I. and S. E. Shreve (1998). *Methods of Mathematical Finance*. New York: Springer-Verlag.

Lewis, F. L. (1986). *Optimal Estimation*. New York: John Wiley & Sons.

Madhavan, A. and S. Smidt (1993). An analysis of changes in specialist inventories and quotations. *Journal of Finance 48*, 1595–1628.

Plexus (Q2 2006). Plexus Group Alpha Capture Analysis. Technical report, Plexus Group, Santa Monica , CA.