Lithuanian Journal of Physics, Vol. **52**, No. 2, pp. 102–114 (2012)

WEAKLY INEFFICIENT MARKETS: STABILITY OF HIGH-FREQUENCY TRADING STRATEGIES

S. Esipov

Quant Isle LTD., Scarsdale, New York, USA
E-mail: sergei.esipov@gmail.com

Received 5 March 2012; accepted 7 June 2012

Market participants who capitalize on high-frequency price dynamics and rely on automated trading are responsible, along with market makers, for the observed level of market efficiency. The remaining *in*efficiency is usually measured as the ratio of expected P&L, derived from the price signals, to its standard deviation. Such signals are also termed *alpha* in market slang. Signals and their volatility depend on time in a different manner, leading to temporal diversification and rise of multi-step strategies. It is shown that the coexistence of small market inefficiencies, multi-step strategies, and market impact lead to price randomization. In other words, high-frequency strategies redefine prices in their attempt to amplify weak price signals, and make markets more effective. In this paper we identify and explore discrete and continuous strategies. We further demonstrate that strategies within the domain of weak inefficiency are stable when incorporated into regular risk-return framework. In the presence of market impact we show how an efficiency edge propagates towards smaller time scales.

Keywords: econophysics, financial markets

PACS: 89.65.Gh

1. Introduction

A market participant who has instantaneous access to all public market information would have an advantage over real-world traders. When the efficient market hypothesis asserts that this information, or, at least, all price-based information, is incorporated in the prices, [1, 2], it leads to the question of how exactly such incorporation is accomplished [3]. If one begins with a price process containing signals, what is the mechanism of price adjustment which can diminish or eliminate signals? Daily operations of prime broker-dealers, hedge funds [4] and major vendors [5] offering alpha-generating products provide an insight that the available level of efficiency is the result of joint action of well-capitalized and numerous market participants seeking excess returns. The shorter are the time scales where market participants operate the more challenging and demanding their infrastructure has to be [6]. While the efficiency progresses towards short times, decade by decade, the *quantitative* description of the *efficiency edge* is always of interest. There are multiple potential sources of signals at short time scales, where tick-by-tick information includes price and volume patterns, order book with its layers, and behavioral dynamics. Quantitative analysis of the efficiency edge should be able to connect the details of the market microstructure at short time scales to the conventional long-term drift-diffusion behaviour of prices. In this paper the ideal or classical efficiency is not considered. The efficiency edge is understood as a boundary between weak and strong inefficiency, or between large and short time scales, respectively.

The simplest model which enables one to study both efficient (long) and inefficient (short) time scales within a single framework is that of autocorrelated process. While hopelessly naive at short time scales, where single price is replaced by market microstructure around the bid-ask spread, the model is quite instructive in capturing the interplay of price signals, market impact, optimal strategies and their feedback on the price process on both sides of the efficiency edge. In this paper we consider a price model which uses signals of two different types. The first type is the exogenous expectation of price increments, which is usually derived from predictors other than the price itself. Its leads to a term structure of such expectations. The second type is the price self-predictability manifested in the auto-correlation of the price process.

We then introduce dynamic strategies on such a price process, and show that by decreasing time steps, in the domain of strong inefficiency, a multi-step optimized strategy leads to an inherent parametric sensitivity, where even signs of the position (whether it is long or short) at any moment in time are not determined by the signs of the price forecasts at that time, but are instead dictated by global constraints. The positions exhibit a divergent or "choppy" nature, and rely, with high sensitivity, on the details of the price forecasts and price correlations. While the used price model is too simplistic in this domain, the sign-changing solutions display behavior reminiscent of what happens at the bid-ask spread, when intermittent trades at bid and ask side are considered.

We then explore these strategies in the continuous limit, under constrains that they remain smooth in price and time, and show that some properties of the price process, the so-called transport coefficients, is what remains of signal and correlation on the efficiency edge, namely, on the weakly efficient side of it. The two different types of signals are needed since the first one survives, in modified form, the application of continuous limit (price drift), while the second one does not (autocorrelation gets encapsulated in drift and diffusion).

With market impact explicitly taken into account in Section 4 we show how smooth strategies acquire additional time scale of optimal position rebalancing, leading to non-local solutions in time. Holders of such optimal positions should be either (i) self-limited in size either due to market impact or (ii) grow until the price process is completely modified, and randomness (or effectiveness) is established at short time scales. It provides a mechanism of efficiency propagation toward smaller time scales.

2. Price-insensitive positions

While we intend to study dynamic strategies resolved in time and price simultaneously, in this Section, in order to begin the study, the positions in the underlying security depend only upon time t, and not on the price variable, S.

2.1. Price process with autocorrelation

The short-term intraday price dynamics is assumed to be non-Markovian. In discrete time setting it has the form

$$S_t = S_{t-1} + \mu_t + \sigma\xi_t\,, \tag{1}$$

where t is the time index. While the time structure of expected price increments μ_t is resolved here, these increments are assumed to be small (with respect to their standard deviations), as market is close to being effective, and therefore the term structure of price variances is disregarded in this section, with the exception of step-to-step autocorrelation. Thus, σ is the time-independent standard deviation of the period price increments, and ξ_t is the random number, different for every time step, historically distributed. (ξ is scaled to have the mean value of zero and standard deviation of one.) To establish price self-predictability the random numbers for subsequent time steps ξ_t are assumed to be correlated, $\langle\xi_t\xi_{t+1}\rangle = \rho$. The time-dependence of such correlation is disregarded here.

Suppose, further, that ϕ_t is the position we intend to establish at time t. This position could subsequently be revised, if needed, at any time step prior to reaching the time moment t, or at t, however consideration of such revisions is postponed until Section 3.6. Our intention is to determine optimal positions from the risk-return perspective.

2.2. Autocorrelation and time step. Stabilization

Since the value of the time step is *a priori* unknown, and its change should take us through the efficiency edge, it is useful to coarse-grain the model Eq. (1). If the time step is doubled, i. e. every two time steps are collapsed into one, the expected increment is then the sum of the two contributing increments, volatility gets multiplied by $2(1+\rho)$ times and autocorrelation ρ becomes ρ',

$$\rho' = \frac{\rho(1+\rho)}{2}\,, \tag{2}$$

When time steps are doubled, this map possesses a single attractor, $\rho = 0$. (The other root, $\rho = 1$, is a repeller.) By period-doubling one could suppress autocorrelation.

The inverse map has a single attractor at $\rho = 1$ if iterated with a positive initial condition. With negative initial condition it is not single-valued, and depending upon the branch taken diverges either in one step or, otherwise, in a few steps. By period-halving one quickly increases auto-correlation.

We will show below in Subsection 2.4 that the time step where $|\rho| = \frac{1}{2}$ is the efficiency edge for model (1). Regular risk-return considerations at larger correlations or smaller time scales are not applicable. If a certain time scale Δt is found to possess a seed correlation ρ then the smaller time scale $\Delta t\rho/2$ suggests the efficiency edge based on inverse map. We use the word "suggests" as the small-scale correlation structure obtained via inverted map provides only an educated guess: it should be sampled directly. However, as we shall see below, within the model (1) the optimal strategies at $|\rho| < \frac{1}{2}$ are smooth, while in the other case, $|\rho| > \frac{1}{2}$ they exhibit parametric sensitivity, and should appear random from a practical viewpoint.

As mentioned in the introduction the single-valued price assumption usually becomes invalid before condition $|\rho| = \frac{1}{2}$ is reached. At small time scales the bid-ask gap is an obstacle for P&L expressions like Eq. (3), as the definition of execution price in (3) and gap-averaged price in Eq. (1) is not the same anymore. (The execution price is quantized, while the gap-averaged price doesn't have to be.) In fact, the elevated frequency of selling at the ask price and buying at bid is one of the remnants of the choppy strategies (explained in detail in Subsection 2.4) in presence of the bid-ask spread.

2.3. Two time steps

In the simple two-step setting it is already instructive to study the properties of optimal positions in presence of autocorrelation. Let the first time step to be between the times $t = 0$ and $t = 1$, and the second step to take place between the times $t = 1$ and $t = 2$. The P&L earned over the first time step is

$$X_1 = \phi_0(S_1 - S_0)\,, \tag{3}$$

and over two time steps it is

$$X_2 = \phi_0(S_1 - S_0) + \phi_1(S_2 - S_1) = \phi_0(\mu_1 + \sigma\xi_1) + \phi_1(\mu_2 + \sigma\xi_2)\,, \tag{4}$$

where we used Eq. (1). The expected P&L for the two periods together is $\bar{X}_2 = \phi_0\mu_1 + \phi_1\mu_2$, and the two-period variance is $V_2 = \sigma^2(\phi_0^2 + 2\rho\phi_0\phi_1 + \phi_1^2)$, where ρ is the correlation coefficient between ξ_1 and ξ_2. It is a common practice to have a P&L goal in terms of expected $\bar{X}_2$ and minimize standard deviation for a given expectation, in the efficient frontier framework. The search for the minimum of standard deviation could be replaced by the search of the minimum of variance, and the external condition on expectation, $\bar{X}_2 = X_0$, could be incorporated via a Lagrange multiplier, λ. The minimum of

$$L = V_2 - \lambda(\bar{X}_2 - X_0) \tag{5}$$

could further be replaced with the maximum of $\bar{X}_2 - X_0 - \lambda_1 V_2$, where $\lambda\lambda_1 = 1$ and since X_0 is a constant here, it also delivers the maximum of $\bar{X}_2 - \lambda_1 V_2$ (usually considered in utility function applications which therefore leave λ_1 unspecified).

Differentiating Lagrangian, L, in Eq. (5) with respect to ϕ_0, ϕ_1, λ, setting the derivatives to zero and solving the resulting equations one finds

$$\phi_0 = \frac{(\mu_1 - \rho\mu_2)X_0}{D}\,, \quad \phi_1 = \frac{(\mu_2 - \rho\mu_1)X_0}{D}\,, \tag{6}$$

where $D = \mu_1^2 - 2\rho\mu_1\mu_2 + \mu_2^2$. The minimal variance is given by

$$V_2 = \frac{(1-\rho^2)\sigma^2 X_0^2}{D}\,, \tag{7}$$

and the signal-to-noise (Sharpe) ratio is

$$\mathrm{Sh}_2 = \sqrt{\frac{D}{(1-\rho^2)\sigma^2}}\,. \tag{8}$$

Despite the elementary nature of this derivation one could already see rather peculiar properties of the optimal positions in presence of autocorrelation.

At zero correlation $\rho = 0$, both periods simply make their contributions to the positions via weights, $\mu_1/(\mu_1^2 + \mu_2^2)$ and $\mu_2/(\mu_1^2 + \mu_2^2)$. However at large correlations, say in the $\rho \to 1$ limit, the positions become opposite to each other in $\phi_0 \simeq X_0/(\mu_1 - \mu_2) \simeq -\phi_1$. This dependence also has a singularity at equal expected increments, $\mu_1 = \mu_2$. If the first increment is larger, $\mu_1 > \mu_2$, the position for the first step is positive regardless of the sign on the expected increment μ_1. Optimal position for the second time step is then negative, and equal to the first one by absolute value. The alternating sign of the two-period optimal positions at large positive correlations is quite different from what one could have arrived at by using step-by-step optimization. For example, at large positive correlations, with positive μ_1 and μ_2, the step-by-step optimization implies keeping a positive position,

$$\phi_0 = X_0/(2\mu_1) > 0\,, \quad \phi_1 = X_0/(2\mu_2) > 0\,. \tag{9}$$

At large negative correlations, the optimal position remains unchanged at $t = 1$, $\phi_0 \simeq \phi_1 \simeq X_0/(\mu_1 + \mu_2)$, producing, again, an answer that has nothing to do with the step-by-step optimization. The latter leads to alternating positions, as μ_1 and μ_2 are likely to have opposite signs in (9).

Both two-step optimal strategies have diverging Sharpe ratio, $\mathrm{Sh}_2 \propto (1 - |\rho|)^{-1}$, implying that the strategies take advantage of predictability offered by high correlation to greatly diminish the two-step volatility. It is this property of global versus local optimization which allows one to guess in advance that the corresponding multi-period problem ought to possess solutions with alternating signs.

2.4. Multi-step discrete setting

Let us consider a case where expectations of price increments are available for many time steps in advance. Although each time step amounts only to a small revision of expected increments, the problem with the two-step optimization is not resolved in multi-step setting as we shall see below.

The P&L and its variance are given by

$$X_n = \sum_{m=1}^{m=n} \phi_{m-1}(\mu_m + \sigma\xi_m)\,,$$

$$V_n = \sigma^2 \sum_{m=1}^{m=n} (\phi_{m-1}^2 + 2\rho\phi_{m-2}\phi_{m-1})\,. \quad (10)$$

Differentiating the Lagrangian

$$L = V_n - \lambda(\bar{X}_n - X_0) \quad (11)$$

with respect to positions $\boldsymbol{\phi}$, results in the linear system

$$\boldsymbol{M\phi} = \frac{\lambda}{2\rho\sigma}\boldsymbol{\mu}\,, \quad (12)$$

where $\boldsymbol{M}$ is a tridiagonal Toeplitz matrix, and $\boldsymbol{\mu}$ is the vector of expected increments,

$$\boldsymbol{M} = \begin{pmatrix} \rho^{-1} & 1 & 0 & & & \\ 1 & \rho^{-1} & 1 & & & \\ 0 & 1 & \rho^{-1} & 1 & & \\ & & & . & & \\ & & & & . & \\ & & & 1 & \rho^{-1} & 1 \\ & & & 0 & 1 & \rho^{-1} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ . \\ \mu_n \end{pmatrix}. \quad (13)$$

The inversion of this matrix is given by

$$M_{ij}^{-1} = (-1)^{i+j}\,\frac{U_{\min(i,j)-1}(v)U_{n-\max(i,j)}(v)}{U_n(v)}\,,$$

$$v = \frac{1}{2\rho}\,, \quad (14)$$

where $U_i(v)$ is the Chebyshev polynomial of the second kind of order i [7, 8]. At small correlations, $\rho \ll 1$, this inverse matrix behaves as

$$M_{ij}^{-1} = (-1)^{i+j}\rho^{1-|i-j|}\,, \quad (15)$$

i. e. decays exponentially away from the main diagonal, along with the contribution of forecasts of adjacent time steps. These contributions are alternating in sign, providing another hint of what happens at larger correlations ρ. Differentiating (11) over λ and setting the derivative to zero gives

$$\lambda = \frac{2\rho\sigma^2 X_0}{\sum_{i,j=1}^n M_{ij}^{-1}\mu_i\mu_j}. \quad (16)$$

and finally the multi-step optimal positions as seen at $t = 0$ are given by

$$\phi_m = \frac{\sum_{i=1}^n M_{m+1,i}^{-1}\mu_i X_0}{\sum_{i,j=1}^n M_{ij}^{-1}\mu_i\mu_j}\,. \quad (17)$$

The leading ρ correction at small $\rho \ll 1$ could be obtained from (15) and (17)

$$\phi_m = \frac{\mu_{m+1}X_0}{\sum_{i=1}^n \mu_i^2} \times \left[1 - \rho\left(\frac{\mu_{m,0} + \mu_{m+2,0}}{\mu_{m+1,0}} - \frac{2\sum_{i=1}^n \mu_i\mu_{i-1}}{\sum_{i=1}^n \mu_i^2}\right)\right]. \quad (18)$$

If one puts the normalization aside, then to first order in the correlation ρ, only the nearest neighboring time steps make contributions to a given position.

Chebyshev polynomials have zeros at $|v| < 1$ in (14), i. e. at $|\rho| > 1/2$. Similar to the two-step case we encounter here multiple divergences as correlations grow in view of oscillating nature of these polynomials.[1] While the multi-step setting was introduced to diminish the optimal position sensitivity to single-step updates in Subsection 2.3, it failed as such. The ARMA(1,1) price process used here is a reliable model, and therefore one may conclude that it is namely the step-by-step sensitive, *discrete* nature of the optimization procedure that leads to oscillating strategies with strong inter-period coupling, with signs completely different from the signs of expected price increments, $\boldsymbol{\mu}$. We call the results of this section *parametrically unstable* or parametrically sensitive, since parametrization with real market data is always noisy, while small changes in the measured values of the expected price increments μ and correlation ρ may lead to

[1] Consider the following numerical 4-period example: $\mu_1 = 0.1$, $\mu_2 = 0.11$, $\mu_3 = 0.09$, $\mu_4 = 0.1$. Using (17) one finds at $\rho = 0.62$ the following strategy in terms of the expected P&L X_0: $\phi_1 = -64.46$, $\phi_2 = 96.98$, $\phi_3 = -99.64$, $\phi_4 = -57.45$. However, at $\rho = 0.63$ the strategy is $\phi_1 = 21.32$, $\phi_2 = -23.11$, $\phi_3 = 27.17$, $\phi_4 = -10.36$. Only the last period sign happened to be the same, and it is negative, while last period signal is positive. In fact, the position sizes go through a singularity at $\rho = 0.62203$. For reference we also list the answer at $\rho = 0.1$, $\phi_1 = 2.58$, $\phi_2 = 2.67$, $\phi_3 = 2.03$, $\phi_4 = 2.65$, in agreement not only with the signs but also with relative amplitudes of the signals.

complete revision of the optimal positions. Given the parametrization uncertainty, there are no optimal positions here at all: the strategy is simply random.

3. Strategies resolved in time and price

We are now interested in the class of strategies which do not change dramatically from step to step, or from price to price, as a result of global optimization, unless it is warranted by the local signal, which is also assumed to be smooth. The corresponding limit of the master equation governing the stochastic process (1) is that of the drift-diffusion (aka continuous) approximation. Presence of autocorrelation makes the derivation of this approximation somewhat involved.

3.1. Lattice model and the continuous approximation

In order to derive the drift- diffusion approximation we first introduce a lattice model on (t, S). On a binary recombining tree (aka tilted square lattice) consider a node at a certain time and price, $(t + \Delta t, S)$, which is connected with two preceding nodes $(t, S - \Delta S)$ and $(t, S + \Delta S)$, and two subsequent nodes, $(t + 2\Delta t, S - \Delta S)$ and $(t + 2\Delta t, S + \Delta S)$.

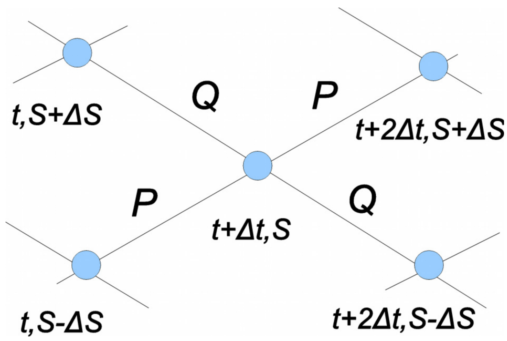

Fig. 1. Lattice.

We now introduce a probabilistic description which naturally accounts for the price drift and autocorrelation of the ARMA(1,1) process (1). Let the probability of passing through the node (t, S) *and* node $(t + \Delta t, S + \Delta S)$ to be denoted as $P(t, S)$, while the probability of passing through the node (t, S) *and* node $(t + \Delta t, S - \Delta S)$ to be denoted as $Q(t, S)$.

Now, the price trajectories which arrive at the node $(t + \Delta t, S)$ from above, i. e. from $(t, S + \Delta S)$ are described by the probability $Q(t, S + \Delta S)$. Suppose a fraction q of these trajectories continues to go down, and contribute to $Q(t+\Delta t, S)$, while the fraction $1-q$ reverts up, and contributes to $P(t + \Delta t, S)$. Similarly, the price trajectories which arrive at $(t + \Delta t, S)$ from below, i. e. from $(t, S - \Delta S)$ are described by the probability $P(t, S - \Delta S)$. Suppose a fraction p of these trajectories continues to go up, and contribute to $P(t+\Delta t, S)$, while the fraction $1-p$ reverts down, and contributes to $(t + \Delta t, S)$. Thus, the master equations on the lattice read

$$\begin{cases} P(t+\Delta t, S) = pP(t, S-\Delta S) \\ \qquad\qquad + (1-q)Q(t, S+\Delta S)\,, \\ Q(t+\Delta t, S) = qQ(t, S+\Delta S) \\ \qquad\qquad + (1-p)P(t, S-\Delta S)\,. \end{cases} \tag{19}$$

These probabilities are normalized by summation over price index S,

$$\sum_S [P(t, S) + Q(t, S)] = 1\,. \tag{20}$$

Assuming detailed balance,

$$(1-q)Q \simeq (1-p)P\,, \tag{21}$$

and expanding these equations to the second order in ΔS and to the first order in Δt one finds

$$\partial_t R = A\partial_S R + \tfrac{1}{2}B^2\partial_S^2 R\,,$$

$$A = \frac{(q-p)\Delta S}{(2-p-q)\Delta t}\,, \quad B^2 = \frac{(q+p)(\Delta S)^2}{(2-p-q)\Delta t}\,, \tag{22}$$

where R stands for probability density corresponding in the continuous setting to either P or Q, or to most of their linear combinations, including the combination $P+Q$, the usual probability of visiting the point (t, S) regardless of the subsequent direction taken. This is the definition of R which is used below. To make the connection with the previous section one must assume that

$$\mu = \frac{(p-q)\Delta S}{1-\rho}\,, \quad \rho = p+q-1\,. \tag{23}$$

Dependencies of the drift coefficient A and diffusion coefficient $\frac{1}{2}B^2$ on the microscopic lattice probabilities p and q are as follows. The sign of the drift coefficient is, obviously, that of the difference $p-q$. In the case of strong "momentum", $p+q \to 2$ or $\rho \to 1$, the drift term diverges, along with the diffusion coefficient (for the diffusion approximation to be valid one requires progressively larger times). In the limit of strong "mean-reversion", $p + q \to 0$ or $\rho \to -1$, the drift term vanishes, $A \simeq (q-p)\Delta S/2\Delta t$, also along with the diffusion coefficient, $\frac{1}{2}B^2 \simeq (q+p)(\Delta S)^2/4\Delta t$, in agreement with what one would expect.

Equation (22) is one of the central equations of this study, it describes evolution of probability densities which are spatially (in S) and temporary (in t) smoothed out at distances exceeding the corresponding steps, and thus free of the step-by-step oscillations which were abundant in the previous section. There is nothing novel in the observation that at large times an ARMA(1,1) process obeys a central limit theorem and its probability evolves in agreement with the Fokker-Planck equation (22). However, the transition from the ARMA(1,1) stochastic description of Eq. (1) to the SDE $dS = Adt + BdW$, where W is the white noise (aka Wiener process), with the coefficients of (22) is far from being straightforward. It is important for this study that optimization based on Eq. (22) is free from oscillating divergences.

3.2. Equations for moments and optimization

Following closely our previous work [9, 10] one could introduce a more detailed $R(t,X,S)$, where X is the account balance accumulated at the point (t,S) as a result of establishing a dynamic position $\phi(t,X,S)$ in the security with the price S. In view of the deterministic relationship $dX = \phi dS$ the derivatives with respect to S could be simply replaced as $\partial_S \to \partial_S + \phi\partial_X$, so Eq. (22) becomes

$$\partial_t R = A(\partial_S + \phi\partial_X)R + \tfrac{1}{2}B^2(\partial_S^2 + 2\phi\partial_S\partial_X + \partial_X^2)R\,. \tag{24}$$

Unlike in the theory of derivatives this is a forward-in-time PDE. Closed equations for the expected P&L and its variance,

$$\bar{X} = \int dX\, X R(t,X,S)\,,$$
$$V = \int dX\,(X-\bar{X})^2 P(t,X,S) \tag{25}$$

could be obtained from (24) using integration by parts, provided that the dynamic strategy is independent on P&L, $\phi(t,S)$. These equations read

$$\begin{cases}\partial_t\bar{X} - A\partial_S\bar{X} - \frac{1}{2}B^2\partial_S^2\bar{X} = -A\phi\,,\\ \partial_t V - A\partial_S V - \frac{1}{2}B^2\partial_S^2 V = B^2\left(\phi - \partial_S\bar{X}\right)^2\,.\end{cases} \tag{26}$$

Solutions of these equations depend on future time and price (t,S), and the forward-in-time optimization leads to strategies which are contingent on these, unknown future arguments. The corresponding backward problem with the terminal conditions $\bar{F}(t,S) = 0$, $V(t,S) = 0$,

$$\begin{cases}\partial_{t_0}\bar{F} - A\partial_{S_0}\bar{F} + \frac{1}{2}B^2\partial_{S_0}^2\bar{F} = A\phi\,,\\ \partial_{t_0}V - A\partial_{S_0}V + \frac{1}{2}B^2\partial_{S_0}^2 V = -B^2\left(\phi + \partial_{S_0}\bar{F}\right)^2\,,\end{cases} \tag{27}$$

is free from such inconvenience, and optimization could easily be performed. Before we proceed it is worth noting that the Lagrangian $L = V - \lambda(F - X_0)$ satisfies a similar equation

$$\partial_{t_0}L - A\partial_{S_0}L + \tfrac{1}{2}B^2\partial_{S_0}^2 L = -B^2\left(\phi + \partial_{S_0}\bar{F}\right)^2 - \lambda A\phi\,, \tag{28}$$

and the *locally* optimal solution is simply given by

$$\phi_l = -\partial_{S_0}\bar{F} - \frac{\lambda A}{2B^2}\,. \tag{29}$$

Substituting this result into the first of Eqs. (27), and assuming constant coefficients A and B one finds $\bar{F}(t,S) = \lambda A^2(T-t_0)/2B^2$, and at $t=0$, $S=S_0$ this gives X_0 so that $\lambda = 2X_0B^2/A^2T$. The second of Eqs. (27) with (29) gives then the variance

$$V(t,S) = \frac{B^2X_0^2(T-t)}{A^2T^2}\,. \tag{30}$$

We will see below that the global optimum (39) is lower.

Returning to Eqs. (27) one may simplify the problem by shifting the strategy, $\phi = -\partial_{S_0}\bar{F} + \phi_1$. The equations become

$$\begin{cases}\partial_{t_0}\bar{F} + \frac{1}{2}B^2\partial_{S_0}^2\bar{F} = A\phi_1\,,\\ \partial_{t_0}V - A\partial_{S_0}V + \frac{1}{2}B^2\partial_{S_0}^2 V = -B^2\phi_1^2\,,\end{cases} \tag{31}$$

Let us denote as $G_A(t-t_0, S-S_0)$ the Green function of the left-hand-side operators. It satisfies the equation

$$\partial_{t_0}G_A - A\partial_{S_0}G_A + \tfrac{1}{2}B^2\partial_{S_0}^2 G_A = \delta(t-t_0)\delta(S-S_0)\,, \tag{32}$$

where index A is the drift. Solutions for the cumulants are given by means of this Green function,

$$\begin{cases}\bar{F}(t_0,S_0) = \\ \qquad \int\limits_{t_0}^{T} dt \int\limits_{-\infty}^{\infty} dS\, G_0(t-t_0, S-S_0)A\phi_1(t,S)\,,\\ V(t_0,S_0) = \\ \qquad \int\limits_{t_0}^{T} dt \int\limits_{-\infty}^{\infty} dS\, G_A(t-t_0, S-S_0)B^2\phi_1^2(t,S)\,.\end{cases} \tag{33}$$

The Euler equation which minimizes Lagrangian $\delta L/\delta\phi(t,S) = 0$ reads

$$2B^2 G_A(t, S-S_0)\phi_1(t,S) + \lambda A G_0(t, S-S_0) = 0\,, \tag{34}$$

at $t_0 = 0$, and allows one to determine the strategy ϕ_1 up to an overall multiplier. Up to this moment the derivation is independent on whether the coefficients are constant or exhibit dependence on time and price.

If the coefficients are indeed constant, then substituting ϕ_1 back into $\bar{F}(0,S_0) = X_0$, using (33), and integrating, one finds

$$\bar{F}(0,S_0) = \frac{\lambda}{2}\left[\exp\frac{A^2T}{B^2} - 1\right] = X_0\,. \tag{35}$$

Determining λ from Eq. (35) and substituting it into (34) we obtain ϕ_1

$$\phi_1 = -\frac{AX_0}{B^2}\frac{\exp\left[\frac{A}{B^2}(S-S_0) + \frac{A^2t}{2B^2}\right]}{\exp\frac{A^2T}{B^2} - 1}\,, \tag{36}$$

The intermediate average unearned P&L and unrealized variance could both be determined from (33) with this ϕ_1:

$$\bar{F}(t,S) = X_0\exp\left[\frac{A(S-S_0)}{B^2} - \frac{A^2t}{2B^2}\right] \times \frac{\exp\frac{A^2T}{B^2} - \exp\frac{A^2t}{B^2}}{\exp\frac{A^2T}{B^2} - 1}\,, \tag{37}$$

and optimal future variance is

$$V(t,S) = X_0^2\exp\left[\frac{2A(S-S_0)}{B^2}\right] \times \frac{\exp\frac{A^2T}{B^2} - \exp\frac{A^2t}{B^2}}{\left[\exp\frac{A^2T}{B^2} - 1\right]^2}\,, \tag{38}$$

which at $t = 0$, $S = S_0$ reduces to

$$V = \frac{X_0^2}{\exp\frac{A^2T}{B^2} - 1}\,. \tag{39}$$

As one could see this global minimum is lower than the local minimum (30).

The full optimal position could be obtained from ϕ_1 by adding the $-\partial_S\bar{F}(t,S)$ shift

$$\phi(t,S) = -\frac{AX_0}{B^2}\frac{\exp\left[\frac{A}{B^2}(S-S_0) + \frac{A^2(2T-t)}{2B^2}\right]}{\exp\frac{A^2T}{B^2} - 1}\,. \tag{40}$$

If the security price grows, $p > q$, and $A < 0$ according to (22), the optimal strategy (40) has contrarian behavior: it decreases the position as price grows and vice versa. This contrarian behavior is more pronounced at the beginning of the time interval, and diminishes thereafter.

At the end we also list the solution of the first of Eqs. (26) with the strategy (36)

$$\bar{X}(t,S) = X_0\exp\left[\frac{A(S-S_0)}{B^2} + \frac{A^2t}{2B^2}\right] \times \frac{\exp\frac{A^2t}{B^2} - 1}{\exp\frac{A^2T}{B^2} - 1}\,. \tag{41}$$

If integrated over S with $G_A(T,0,S,S_0)$ is gives X_0 at $t = T$.

3.3. *Time-dependent transport coefficients*

At the beginning of this study, when strategies did not depend on price, it was the time-dependence of the price increments that lead to extreme parameter sensitivity. Now, in the drift-diffusion approximation, it is of interest to find out what is the status of this problem. If the drift rate, A, and diffusivity $B^2/2$ are functions of time, the Green function of Eq. (32) is the same functional form, only drift and diffusion lengths are time-averaged as in (43) below. The solution for the optimal trajectory now reads

$$\phi_1(t,S) = -\frac{A(t)X_0}{B^2(t)} \times \left[\int_0^T dt'\,\frac{A^2(t')}{B^2(t')}\exp\left(\frac{\langle A(t')\rangle^2}{\langle B^2(t')\rangle}\right)\right]^{-1} \times \exp\left[\frac{(S-S_0)\langle A(t)\rangle}{2\langle B^2(t)\rangle} + \frac{\langle A(t)\rangle^2}{\langle B^2(t)\rangle}\right]\,, \tag{42}$$

where

$$\langle C(t)\rangle \equiv \int_0^t dt'\,C(t'). \tag{43}$$

As one can see the strategy is smooth, and its sign is dictated by local signal, $A(t)$.

3.4. The case of small signals

In the small signal limit, where

$$(S-S_0)\langle A(t)\rangle \ll \langle B^2(t)\rangle, \langle A(t)\rangle^2 \ll \langle B^2(t)\rangle\,, \tag{44}$$

there is no difference between ϕ_1 and ϕ. The optimal strategy follows from (42)

$$\phi_1(t,S)=\phi(t,S)=-\frac{A(t)X_0}{B^2(t)}\left[\int\limits_0^T dt'\,\frac{A^2(t')}{B^2(t')}\right]^{-1}, \tag{45}$$

which is the continuous analog of Eq. (17) if one suppresses the time dependence of diffusivity (such time dependence was disregarded in the derivation of (17)). In the continuous limit the one-step lead-lag dependence is not needed: formula (45) corresponds to the uncorrelated limit of (18) at $\rho=0$. This approximation could be called “local”, as this result could be obtained directly from (29).

Thus, in the small signal limit, *if* the signal is price-independent, then there is no need to consider price-dependent strategies.

Similar calculation could be carried out in the case where the drift coefficient is price-dependent, $A(t,S)$ but the diffusivity is not, $B(t)$, which is the most relevant case from practical viewpoint. The optimal strategy is simply

$$\phi(t,S)=-\frac{A(t,S)X_0}{B^2(t)}\Bigg[\int\limits_0^T\frac{dt'}{B^2(t')\sqrt{2\pi\langle B^2(t')\rangle}}$$
$$\times\int\limits_{-\infty}^{\infty}dS'\,A^2(t',S')\exp\left(-\frac{(S-S_0)^2}{2\langle B^2(t')\rangle}\right)\Bigg]^{-1}. \tag{46}$$

While this strategy is price-dependent, it is also local, and also could be obtained directly from (29).

We have seen above that after expanding the exact discrete Eqs. (19) in the continuous limit (the drift-diffusion approximation) the auto-correlation between time intervals is encapsulated in the transport coefficients, A,B. In other words, in the class of smooth optimal strategies the residual auto-correlation is irrelevant as long as drift and volatility are properly determined.

Finally, Eq. (46) provides an answer for the question of what happens if the small signal is price-dependent. As one could see, things do not change much in the local approximation: the optimal position is still proportional to the local signal. It is normalization, or Lagrange multiplier λ, that gets modified to take into account possible price trajectories with their appropriate weights given by the Green function.

3.5. Multiple securities

For a price vector $\boldsymbol{S}=(S_1,...,S_n)$, and positions vector $\boldsymbol{\phi}=(\phi_1,...,\phi_n)$ the stochastic variables in the drift-diffusion approximation are driven by coupled Ito SDEs

$$\begin{cases} dS_k = A_k dt + B_k dW_k\,, \quad k=[1,n]\,; \\ dX=\phi_m dS_m\,, \end{cases} \tag{47}$$

where summation over repeating indices is understood, $a_n b_m \equiv \sum_{k=1}^{k=n} a_k b_k$. The PDE for the probability distribution of having an unearned P&L, $R(F,\boldsymbol{S},t)$, analogous to forward Eq. (24), is given by

$$\partial_t R = A_k(\partial_k+\phi_k\partial_X)R+\tfrac{1}{2}B_kB_m r_{km}$$
$$\times\left(\partial_k\partial_m+2\phi_m\partial_k\partial_X+\phi_k\phi,\partial_X^2\right)R, \tag{48}$$

with the convention $\partial/\partial S_k=\partial_k$. The backward equations for the first two cumulants, generalizing Eqs. (27), are

$$\begin{cases} \partial_t\bar{F}-A_k\partial_k\bar{F}+\tfrac{1}{2}B_kB_m r_{km}\partial_k\partial_m\bar{F}= \\ \qquad A_k\phi_k\,, \\ \partial_t V - A_k\partial_k V+\tfrac{1}{2}B_kB_m r_{km}\partial_k\partial_m V= \\ \qquad -B_kB_m\left(\phi_k-\partial_k\bar{F}\right)\left(\phi_m-\partial_m\bar{F}\right)\,. \end{cases} \tag{49}$$

The local optimizer is

$$\phi_k=-\partial_m\bar{F}-\tfrac{\lambda}{2}N_{km}^{-1}A_m,\quad N_{km}=B_kB_m r_{km}\,, \tag{50}$$

and after normalization (determining λ) it reads

$$\phi_k(t,\boldsymbol{S})=-X_0N_{km}^{-1}A_m(t,\boldsymbol{S})\Bigg\{\int\limits_0^T\frac{dt'}{(2\pi)^{n/2}}\times$$
$$\times\frac{1}{\langle\det\boldsymbol{N}^{1/2}\rangle}\int\limits_{-\infty...-\infty}^{\infty...\infty}d\boldsymbol{S}\,A_l(t',\boldsymbol{S})N_{lj}^{-1}(t')A_j(t',\boldsymbol{S})$$
$$\times\exp\left[-\tfrac{1}{2}\left(S_i-S_{i0}\right)\langle N_{ip}^{-1}(t')\rangle\left(S_p-S_{p0}\right)\right]\Bigg\}^{-1},$$

generalizing Eq. (46).

3.6. Signal updates

At time $t = 1$ new expectations regarding second-period price increment are formed, $\mu_{11} \neq \mu_{10}$, as new information becomes available, not necessarily accounted for by the correlation coefficient ρ, as the latter only accounts for (auto)correlation of the price, and the price forecast may have other predictors.

The revised second period position could be determined in more than one way. If the ultimate goal is to achieve a certain P&L, then $\phi_{11} = (X_0 - X_{11})/\mu_{21}$, where $X_{11} = \phi_{00}(S_{11} - S_{00})$ is the realized P&L for the first time step as assessed at $t = 1$ (the second subscript here refers to the time when assessment is made). Regardless of whether the updated second period expected price increment, μ_{21}, has anything to do with the its previous estimate μ_{20}, the updated position differs arbitrarily from the initially proposed value in (6).

If, instead of keeping the P&L goal, the two-step optimization procedure is simply rolled over, then the updated optimal position

$$\phi_{11} = \frac{(\mu_{21} - \rho\mu_{31})X_0}{\mu_{21}^2 - 2\rho\mu_{21}\mu_{31} + \mu_{31}^2}\,, \tag{51}$$

also differs arbitrarily from the initially proposed value in (6). Since the value of μ_{21} is unknown at the moment when first position ϕ_{00} is established one cannot in advance determine whether the two-step optimization is to be chosen.

The situation is simplified in the case of small signals where standard deviation of the mean value of the signal is smaller than period price volatility, $\sigma_\mu \ll \sigma$. Explicit account of the second-step signal volatility is then dwarfed by that of the price and the original model is recovered. Thus, there is no need to modify the optimization procedure in view of the signal volatility. This simplification comes under condition that only the truly "average" part of the signal is left, i. e. the parametrization procedure for determining expectations of future period price increments, such as μ_{20}, retains only that portion of the signal which, on expected basis, survives the future signal updates.

The price process (1), by itself, is stable with respect to the deterioration of the initial forecast μ_{20} over time (which manifests itself in renormalization of the inter-period correlation coefficient). Consider, for example, a model where the forecast update is contingent on the realized price increment. If such deterioration is small over one time step (or could be made small by decreasing the time step), then to the first approximation $\mu_{21} - \mu_{20} \simeq \gamma(S_{11} - S_{00}) = \gamma(\mu_{20} + \sigma\xi_{11}) \simeq \gamma\sigma\xi_{11}$, where ξ_{11} is the realized "random" number for the first period, as seen from $t = 1$. The constant term is absent here, as there couldn't be any systematic price-independent increase or decrease of increments. In the model, the second price increment acquires dependence on the previous noise term, ξ_{10} :

$$S_{20} = S_{10} + \mu_{20} + \sigma(\gamma\xi_{10} + \xi_{20})\,. \tag{52}$$

Since the term in the parentheses here is again a zero-mean normal number with the variance slightly increased in $1 + 2\gamma\rho + \gamma^2$ times, and renormalized correlation coefficient is $(\gamma + \rho)/\sqrt{1 + 2\gamma\rho + \gamma^2} \simeq \gamma + \rho$, the model of Eq. (1) is adequate for describing autocorrelation and deterioration of expected increments at the same time. The processes (1) and (52) both belong to the ARMA(1,1) category, and the latter form shows both noise terms explicitly. We preferred to work with the form (1).

4. Market impact

Dynamic optimal strategies obtained above follow signals regardless of the position size and trading rate. In practice, both position size and trading rates are constrained. Price responds to trading rate and number of shares traded, this response is known as market impact [11]. Detailed modeling of market impact and its implications for portfolio rebalancing is a developed topic [12, 13]. Possession of insider information regarding expected future price impact does not contadict even ideal efficiency, as the trader may determine (from historical experience and from quantitative analysis) the price implications for a given trading trajectory, ϕ. Analysis of the market impact given below combines it with weak market inefficiency. In a simple *permanent* impact model we consider a contribution to the drift term which is an odd function of the local rate of portfolio rebalancing, $A_\phi = A + a(\partial_t\phi)$, and the equations for the cumulants (27) acquire strategy-dependent drift A_ϕ. While the general (t, S)-dependent search for optimal strategies becomes analytically intractable past Eqs. (33), since the Green function G_A now explicitly depends on strategy through the drift term, the progress could be made in the small-signal small-impact approximation, or, separately, for S-independent strategies.

4.1. Small-signal small-impact approximation

While the small-signal limit, along with the available level of efficiency, are both maintained by joint action of market participants, the impact magnitude of rebalancing someone's portfolio is entirely in the hands of portfolio managers. Assuming that it is in the best interest of the managers to keep the impact small while taking advantage of the signal (this assumption is relaxed in Subsection 4.3), one could study the local approximation. As explained above, in the local approximation the difference between the Green function with the drift and without is negligible, and so is the difference between ϕ and δ-corrected ϕ_1. In this approximation Eq. (28) has the following solution for the Lagrangian, L,

$$L = \int_0^T dt \int_{-\infty}^{\infty} dS\, G_0(T-t, S-S_0) \times \left\{ B^2\phi^2 + \lambda\phi\left[A + a(\partial_t\phi)\right]\right\} . \qquad (53)$$

If there is no S-dependence in the drift and diffusion coefficients, the averaging over prices is reduced to normalization condition and taking derivative with respect to $\phi(t)$ results in the Euler equation

$$2B^2\phi + \lambda\left[A + a(\dot\phi)\right] = \lambda\frac{d}{dt}\left[\phi a'(\dot\phi)\right] , \qquad (54)$$

where $\dot\phi = d\phi/dt$. To give an example of what this Euler equation leads to, consider a family of power-law impacts, $a(x) = a_0x^\nu$ with $\nu > 1$. Then Eq. (54) is satisfied, for instance, by

$$B^2\phi(t) = -\lambda A + \left\{ \left[B^2\phi(0) + \lambda A\right]^{\frac{\nu-1}{\nu}} - B^2t\left[\lambda a_0(\nu-1)\right]^{-\frac{1}{\nu}} \right\}^{\frac{\nu}{\nu-1}} \qquad (55)$$

(there exist other solutions). In presence of signal A this trading trajectory delivers a time-dependent optimizer for rebalancing portfolio from $\phi(0) \geq \phi(t)$ down to the local optimizer $-\lambda A/B^2$ which we encountered in (29). The rebalancing takes finite time, t_r, and it is over when the curly bracket in (55) reaches zero,

$$t_r = B^{-2}\left[\lambda a_0(\nu-1)\right]^{\frac{1}{\nu}}\left[B^2\phi(0) + \lambda A\right]^{\frac{\nu-1}{\nu}} , \qquad (56)$$

If the signal changes considerably over this time scale, the strategy enters a strong-coupling regime where rebalancing is not finished by the time a given forecast expires, and market impact begins to limit the position trajectory at all times. Clearly, the rebalancing time, t_r increases with the position size, and any growing strategy always becomes impact-limited. When this happens the position sign may again lose direct relation to the local signal, and in this sense large portfolios may act as price randomizers, redefining price volatility and other refined properties. Indeed, the price process $dS = [A + a(\dot\phi)]dt + BdW$ is contingent on the strategy $\phi(t)$ above. [2]

On the other side, the rebalancing time decreases as impact elasticity parameter ν approaches 1 from above. At $\nu = 1$ impact is trajectory-independent, and at $\nu < 1$ instant rebalancing is favored, this leads to *block trading*, and it is not considered here.

If the drift is time-dependent, $A(t)$, Eq. (54) could be integrated numerically. If drift A and therefore ϕ exhibit joint (t, S)-dependence, the Euler equation reads

$$G_0(T-t, S-S_0)\left\{2B^2\phi + \lambda\left[A + a(\dot\phi)\right]\right\} = \lambda\partial_t\left[\phi G_0(T-t, S-S_0)\frac{da(\dot\phi)}{d\dot\phi}\right] . \qquad (57)$$

or

$$2B^2\phi + \lambda(A + a) = \lambda\partial_t(\phi a') + \lambda\phi a' h(t, S) ,$$
$$h(t, S) = \frac{(S-S_0)^2}{2B^2(T-t)^2} - \frac{1}{2(T-t)} , \qquad (58)$$

and also requires numerical integration for different S. Here $\dot\phi = \partial_t\phi$. Finally, if price volatility depends on time, a substitution $B^2(T-t) \to \int_t^T dt' B^2(t')$ should be done in the Green functions G_0 entering Eq. (57).

4.2. Multiple securities revisited

Market impact for multiple securities is a relatively less understood issue as compared to single-security case. Assuming for simplicity that each security is influenced by its own trading, and disregarding cross-terms, one could write the solution for multiple

[2] Even in the highly idealistic case when there are (i) only two market participants, and (ii) their views regarding the signals are identical, and (iii) they make use of the optimization procedure outlined above, and (iv) they employ the same capital, each participant should factor in the presence of the other. Otherwise, a term proportional to $[a(2\dot\phi) - a(\dot\phi)]$ would have to be assigned to the volatility, thus slightly redefining B. Since none of the assumptions above holds true in real markets, there is no quantitative methodology that may separate joint market impact of trading and that of the news which are exogenous to trading.

securities Lagrangian in the small-signal small-impact approximation

$$L = \int_0^T dt \int d\boldsymbol{S}\, G_0(T-t, \boldsymbol{S}-\boldsymbol{S}_0)$$

$$\times \{B_k B_m \phi_k \phi_m + \lambda \phi_k [A_k + a_k(\partial_t \phi_k)]\} . \quad (59)$$

The system of coupled non-linear Euler ODEs reads

$$G_0(T-t, \boldsymbol{S}-\boldsymbol{S}_0)$$

$$\times \{2B_k B_m \phi_m + \lambda [A_k(t,\boldsymbol{S}) + a_k(\dot{\phi}_k)]\} = \quad (60)$$

$$\lambda \partial_t \left[\phi_k G_0(T-t, \boldsymbol{S}-\boldsymbol{S}_0) \frac{d a_k(\dot{\phi}_k)}{d(\dot{\phi}_k)} \right] , \; k \in [1, N] , \quad (61)$$

or

$$2B_k B_m \phi_m + \lambda (A_k(t,S) + a_k) =$$

$$\lambda \partial_t (\phi_k a_k') + \lambda \phi_k a_k' H(t, \boldsymbol{S}) ,$$

$$H(t,\boldsymbol{S}) = \frac{\partial_t G_0(T-t, \boldsymbol{S}-\boldsymbol{S}_0)}{G_0(T-t, \boldsymbol{S}-\boldsymbol{S}_0)} . \quad (62)$$

This system requires a numerical solution.

4.3. Market impact at two discrete timesteps

The trading-dependent non-linearity that market impact introduces raises two fundamental questions. The first question is whether the "roundtrip" strategy could be made profitable? In other words, could a position be opened and then closed with different expected impacts? The second question is how does market impact limit the position size as one increases the value of expected profit X_0?

Let us return to the simple setting of Subsection 2.3, where, in addition, we assume first return μ_1 is positive and exceeds the second return μ_2 by absolute value. Further suppose that a long position ϕ is uniformly opened during the first time step and uniformly closed during the second. Then the rate of position change is $\phi/\Delta t$, and the market impact is $\pm a_0(\phi/\Delta t)^\nu = \pm a_1 \phi^\nu$, where plus sign corresponds to the first step when the position is opened, and minus sign corresponds to the second time step when the positon is closed. The total random P&L is

$$F_2 = \frac{\phi(\mu_1 + a_1\phi^\nu)}{2} + \frac{\phi(\mu_2 - a_2\phi^\nu)}{2} + \frac{\sigma\phi(\xi_1+\xi_2)}{3} , \quad (63)$$

and the minmization results in

$$\begin{cases} \sigma^2\phi(1+\rho) = \mu_1 + \mu_2 + (a_1 - a_2)(\nu+1)\phi^\nu , \\ \phi[\mu_1 + \mu_2 + (a_1 - a_2)\phi^\nu] = 2X_0 . \end{cases} \quad (64)$$

These equations are instructive. First of all $a_1 \neq a_2$.[3] At $a_1 < a_2$, by increasing X_0 in the second equation (64) one finds that there exists a critical value of $X_{0c} = \phi_c[\mu_1 + \mu_2 + (a_1 - a_2)\phi_c^\nu]/2$ corresponding to the *largest* possible position size, $\phi_c^\nu = (\mu_1+\mu_2)/(a_2 - a_1)(\nu+1)$. At yet larger values of X_0 there is no positive solution (and could be no solution at all, depending on ν): the negative impact from positon closure dominates both the signal and the positive impact of position opening, thus making the goal X_0 unfeasable. Depending on value of ν (e. g. at $\nu = 3$) more than one real root may exist, including negative roots, taking advantage of the impact properties.

At $a_1 > a_2$ in the limit of large X_0 and $\nu > 1$ (see Subsection 4.1) the signals $\mu_{1,2}$ become irrelevant, $\phi \simeq [2X_0/(a_1 - a_2)]^{1/(\nu+1)}$, and profit is made based on the position "roundtrip", due to the impact alone. The parameter relation $a_1 > a_2$ implies that opening positions generates more impact than closing the position. Moreover, one can see that the position size *after* the market impact taken into account is *unlimited* in the model, and therefore the corresponding inefficiency will be discovered, the impact will grow with the position sizes assumed by the participants, and the price process will be surely modified, potentially invalidating the data series used to perform parametrization, on which the condition $a_1 > a_2$ was based. Since correlation ρ is not present in the last two paragraphs, the consideration remains valid on both sides of the efficiency edge.

4.4. Market impact randomizes the price process

The simple two-step example of the previous Subsection allows one to consider how the price process is modified. For this let's suppose that $a_1 > a_2$, and the value of expected profil X_0 is large enough, so that the impact is the major contributor to P&L. When the position is opened, at the end of first time step, with accumulated position, $\phi = [2X_0/(a_1 - a_2)]^{1/(\nu+1)}$, the price acquires an additional positive increment

$$\Delta S_{\phi,1} = a_1[2X_0/(a_1 - a_2)]^{\nu/(\nu+1)} . \quad (65)$$

[3] Even if these coefficients are made equal, the equality will be violated by changing the time step durations.

Similarly, during the second period one gets an additional negative increment,

$$\Delta S_{\phi,2} = -(a_2/a_1)\Delta S_{\phi,1}\,. \tag{66}$$

Single-step observed volatility is then renormalized

$$\sigma^2 \to \sigma^2 + \frac{(a_1+a_2)^2}{4a_1^2}\Delta S_{\phi,1}^2 =$$

$$\sigma^2 + \frac{(a_1+a_2)^2}{4}[2X_0/(a_1-a_2)]^{2\nu/(\nu+1)}\,, \tag{67}$$

i. e. inreases with X_0, assuming everything else being constant. The limit $a_1 \to a_2$, where the market-impact"roundtrip" P&L vanishes, is singular in terms of volatility renormalization. This singularity is the result of removing signal $\mu_{1,2}$ from consideration, and it is saturated at $a_1 - a_2 \sim (\mu_1+\mu_2)^{\nu+1}(2X_0)^{-\nu}$. Using the dependence $\sigma \propto X_0^{\nu/(\nu+1)}$, and terms structure of volatility at different time scales, T, (on top of $\sqrt{T}$) one could estimate the amount of capital employed at different time scales.

Step-to-step correlation is similarly renormalized,

$$\rho \to \frac{\rho - \frac{(a_1-a_2)^2}{4a_1^2\sigma^2}\Delta S_{\phi,1}^2}{1 + \frac{(a_1+a_2)^2}{4a_1^2\sigma^2}\Delta S_{\phi,1}^2}\,, \tag{68}$$

i. e. decreases with X_0 while market impact could still be considered as a perturbation. The growth of volatility is the major manifestations of price randomization caused by market impact. [4]

5. Conclusion

We have studied optimal dynamic investment strategies within a range of timescales in a price model with external and internal signals (auto-correlation). At short timescales as determined by temporal correlation of the underlying security price the optimal strategies are noisy and parametrically unstable. The signs and amplitudes of the optimal positions are not determined by the local properties of the signal, these positions exhibit strong coupling and are essentially global in their properties. Small change of the correlation assumption or of the price signal term structure may completely modify these strategies. In other words the parameter space $(\rho > 1/2, \boldsymbol{\mu})$ is broken into an infinitely large number of sub-domains containing quickly varying term structures of positions. If such large correlations were in existence in a world with single-valued price, then price at these time scales would have become increasingly chaotic not because the price process is assumed random, but because any attempts to optimize the high-frequency trading under these assumptions lead to practically random strategies, and randomize the price process with market impact taken into account. The bid-ask spread trading is the remnant of these oscillations. The net effect of this process would be propagation of the efficiency edge towards shorter time scales.

At large timescales correlations enter transport coefficients of drift and diffusion, and the market is only weakly ineffective. The optimal dynamic strategies depending on time and underlying security prices have been studied globally here for large time-scales in the continuous or drift-diffusion approximation. They were found to be contrarian in nature and more so at the beginning of the time domain rather then at the end. Extensions to time and price-dependent transport coefficients and to multiple securities are given in the so-called small-signal approximation, which is also termed "local". In this approximation the sign and amplitude of positions are simply proportional to the signal (barring the non-local signal-dependent normalization factor if Lagrange multiplier is left unspecified).

Presence of market impact modifies optimal strategies at large time scales, and numerical computations are required to determine the positions by solving systems of non-linear ODEs even in the local approximation. Market impact introduces its own time scale, the so-called optimal rebalancing time. If this time exceeds the time step the problem again exhibits strong coupling due to the delays caused by slow trading in view of market impact. With increasing position size one finds that a price randomization mechanism operates at large time scales as well.

We have shown that high-frequency strategies in presence of market impact provide a mechanism for efficiency propagation towards smaller time scales. The larger are the positions the stronger is the randomization effect. A number of solutions (or equations to be solved numerically) is given for optimal strategies, in the risk-return framework, under the joint effect of inefficiency and market impact.

[4] The decrease of correlation is only helpful if the seed correlation was positive. If the seed correlation were negative it would have become even more negative. The negative contribution to autocorrelation is simply the result of position being opened and closed in this example. As all kinds of strategies are explored the impacted price will exhibit all kinds of trajectories.

References

[1] E. Fama, The behavior of stock market prices, Journal of Business **38**, 34–105 (1965), http://www.jstor.org/stable/2350752

[2] P. Samuelson, Proof that properly anticipated prices fluctuate randomly, Industrial Management Review **6**, 41–49 (1965)

[3] E. Fama, Efficient capital markets: A review of theory and empirical work, Journal of Finance **25**(2), 383–417 (1970), http://dx.doi.org/10.1111/j.1540-6261.1970.tb00518.x

[4] A.W. Lo, *Hedge Funds: an Analytic Perspective* (Princeton University Press, Princeton and Oxford, 2008)

[5] On its web site Thomson Reuters advertises a set of data and tools to help their clients in developing and deploying strategies for generating alpha and controlling risk (Thomson Reuters, 2010), http://thomsonreuters.com/products_services/financial/financial_products/quantitave_research_trading

[6] J. Keehner, Milliseconds are focus in algorithmic trades (2007), http://www.reuters.com/article/ExchangesandTrading07/idUSN1046529820070511

[7] T. Yamamoto and Y. Ikebe, Inversion of band matrices, Linear Algebra and its Applications **24**, 105–111 (1979), http://dx.doi.org/10.1016/0024-3795(79)90151-4

[8] G. Meurant, A review on the inverse of symmetric tridiagonal and block tridiagonal matrices, SIAM Journal on Matrix Analysis and Applications **13**(3), 707–728 (1992), http://dx.doi.org/10.1137/0613045

[9] S. Esipov and I. Vaysburd, On the profit and loss distribution of dynamic hedging strategies, International Journal of Theoretical and Applied Finance **2**, 131–153 (1999)

[10] S. Esipov and I. Vaysburd, Dynamic investment strategies: Portfolio insurance versus efficient frontier, available at SSRN (1999), http://ssrn.com/abstract=170574 or http://dx.doi.org/10.2139/ssrn.170574

[11] A. Kyle, Continuous auctions and insider trading, Econometrica **53**(6), 1315–1335(1985), http://www.jstor.org/stable/1913210

[12] R. Almgren and N. Chriss, Optimal execution of portfolio transactions, Journal of Risk **3**, 5–39 (Winter 2000/2001)

[13] R. Almgren, Optimal execution with nonlinear impact functions and trading-enhanced risk, Applied Mathematical Finance **10**, 1–18 (2003), http://dx.doi.org/10.1080/135048602100056