Statistical Arbitrage:
Medium Frequency Portfolio Trading

Igor Skachkov *

July 9, 2013

Contents

1 Signal Generation **3**
- 1.1 Regression on Indices and Largest Principal Components . . . 3
- 1.2 "Noise" Strategies . . . 5
- 1.3 Fractional Integration . . . 6

2 Execution **14**
- 2.1 Market impact models . . . 14
- 2.2 Multi-periods portfolio trading . . . 17
- 2.3 Inventory Management . . . 18

3 Results and discussion **20**

A Appendix **24**
- A.1 Multidimensional optimization . . . 24
  - A.1.1 Boundary conditions . . . 25
  - A.1.2 Terminal constraints . . . 26
- A.2 Multi-period liquidation . . . 27

*iskachkov@yahoo.com

Abstract

Medium frequency trading strategies include all trading activities, that do not require market microstructure analysis on one side and significantly depend on market impact on the other side. The most important difference from high frequency trading is the ability to analyze big amount of data using complex algorithms. Portfolio management in this case is the dynamic process, combination of signal (alpha) discovery and optimal execution on the level of trading scheduling. We used close price and trading volume time series for the list of S&P 500 companies that exist in an index since the beginning of 2008 at least. In this paper we present signal generation approaches as well as optimization of portfolio transactions. Formally the performances of medium frequency statistical arbitrage strategies are much better than the performance of their benchmarks, but they are very sensitive to the quality of trading engine and optimization software.

In this minor revision we added the results of out-of-sample tests and explanations of terms and methodology.

1 Signal Generation

We consider N time-series of logarithmic prices $\mathbf{S_t} = (S_{t1}, S_{t2}, \ldots, S_{tN})$ of the length T and corresponding log-returns $\mathbf{R}$

$$R_{tn} = S_{tn} - S_{t-1\,n}, \quad n = 1, 2, \ldots N, \quad t = 1, 2, \ldots T, \quad R_{0n} = 0$$

We assume a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y}$ can be price or return time-series, $\boldsymbol{c}$ is a constant vector, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients,$\boldsymbol{\varepsilon}$ are orthogonal to factors $\boldsymbol{X}$ residuals.

1.1 Regression on Indices and Largest Principal Components

The very first idea that can be exploited is the comparison of individual stocks and index time-series. We have a single not trivial factor.

The regression equations for return and price series are

$$\begin{aligned} \boldsymbol{R} &= [\mathbf{1}\ \boldsymbol{R}_{ind}]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{S} &= [\mathbf{1}\ \boldsymbol{t}\ \boldsymbol{S}_{ind}]\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \end{aligned} \tag{2}$$

$\boldsymbol{t} = [1, 2, \ldots t]'$, is the vector of constant trend and $\mathbf{1} = [1, 1, \ldots 1]'$ is the constant vector of the same length.

The residuals of regression equations (2) are our first candidates for trading strategy signals α:

$$\boldsymbol{\alpha} = -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} = \boldsymbol{M}\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{Hat})\boldsymbol{y}, \tag{3}$$

where $\boldsymbol{Hat}$ and $\boldsymbol{M}$ are symmetric, idempotent matrices.

$$\boldsymbol{Hat} = \boldsymbol{X}(\boldsymbol{X'}\boldsymbol{X})^{-1}\boldsymbol{X'} \tag{4}$$

The profit and loss (PnL) data for optimized portfolios presented on figure (1) and figure (2). The Sharpe ratios (SR) for return and price series based strategies are close, but the return with the second approach is much higher plus the first strategy has a turnover (the ratio of average absolute value of trading order to the sum of absolute values of portfolio positions) that is greater then 100%. Firstly, it means that the daily frequency is too low: we still hold the position when the sign of the signal is opposite. Secondly, such a high turnover leads to high transaction cost. We will analyze that subject in details in next section. In the meantime we just remind that with 100% turnover and transaction cost of 2 basis points (bps) the annualized return would be reduced by 5%.

The next step is to use some specific benchmarks. We can use ETFs or make more quantitative choice - as a regressor we take the eigenvector that corresponds to the largest eigenvalue or the most

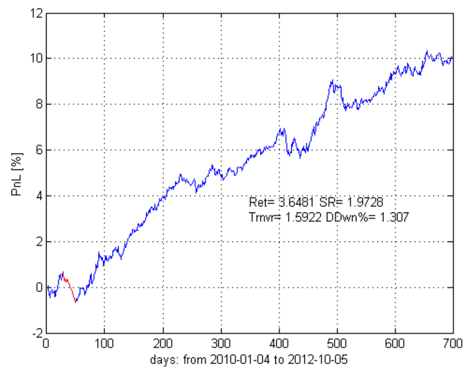

Figure 1: Regression of return time-series on SPY. Annual return equals to 3.65 %, Sharpe ratio - 1.97, turnover - 159%, max drawdown - 1.31%

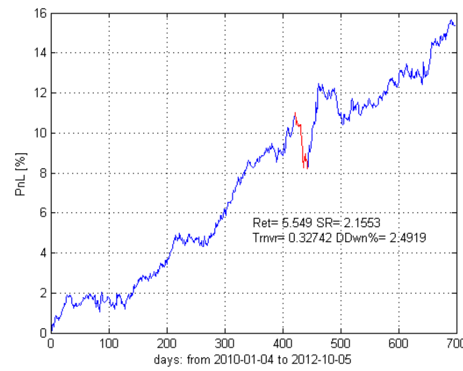

Figure 2: Regression of price time-series on SPY. Annual return equals to 5.55 %, Sharpe ratio - 2.16, turnover - 33%, max drawdown - 2.49%

principal component. S&P500 index we split to 10 Global Industry Classification Standard (GICS) sectors. The regression equation for principal components can be simplified: each time-series can be presented as a combination of principal components

$$\boldsymbol{y} = \boldsymbol{X\beta}, \tag{5}$$

X is a $T \times N$ matrix of components and $\boldsymbol{\beta}$ is a $N \times 1$ vector of coefficients. The residual of regression to the set of principal components is also a regression - to complementary set of orthogonal vectors

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2}, \tag{6}$$

$\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are $T \times k$ and $T \times (N-k)$ matrices. In matrix form

$$\boldsymbol{Y} = \boldsymbol{X_1B_1} + \boldsymbol{X_2B_2}, \tag{7}$$

where $\boldsymbol{Y}$ is a portfolio of N stock time-series, matrices $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ consists of k and $N-k$ vectors of coefficients. Let us denote matrix of eigenvectors as $\boldsymbol{V}$ and diagonal matrix of eigenvalues as $\boldsymbol{\Lambda}$. Also let the first term in eq.(7) be the main part and the second term - the residual of regression. Then

$$\varepsilon = \boldsymbol{X_2B_2} = \boldsymbol{X_2(X_2'X_2)^{-1}X_2'XV'} = \boldsymbol{X_2\Lambda_2^{-1}[\Lambda_2 0]V'} = \boldsymbol{X_2[E_2 0]V'} = \boldsymbol{YV_2V_2'} \tag{8}$$

where $\boldsymbol{E_2}$ is a $k \times k$ identity matrix, $\boldsymbol{0}$ is $k \times (N-k)$ zero matrix. Figure 3 and figure 4 show the PnL of the strategies on assumption that "fair price" of the stock follows the move of the largest principal component of the sector. The difference between correlation and covariance matrices based calculations are not significant: variance/covariance coefficients are comparatively uniform inside the sectors of large cap stocks.

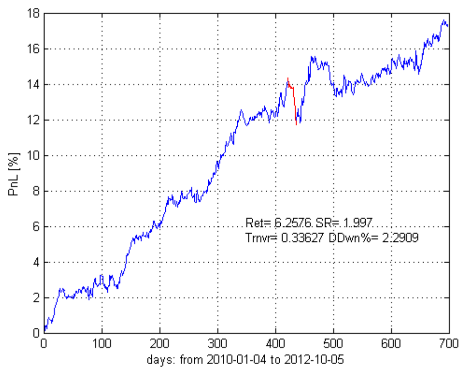

Figure 3: Regression of price time-series on the largest principal component. Covariance matrix. Annual return - 6.25 %, Sharpe ratio - 2.0, turnover - 34%, max drawdown - 2.3%

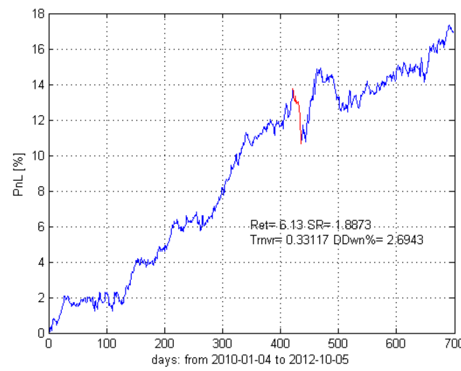

Figure 4: Regression of price time-series on the largest principal component. Correlation matrix. Annual return - 6.13 %, Sharpe ratio - 1.89, turnover - 33%, max drawdown - 2.7%

Besides correlation and covariance matrices it is possible to consider more general matrix

$$\boldsymbol{C} = \boldsymbol{diag(\sigma)}^{\gamma} \cdot \boldsymbol{Cov} \cdot \boldsymbol{diag(\sigma)}^{\gamma} \tag{9}$$

or

$$C_{i,j} = (\sigma_i \sigma_j)^{\gamma+1} \rho_{i,j} \tag{10}$$

where $\boldsymbol{Cov}$ is a covariance matrix. $\boldsymbol{\sigma}$ is a vector of standard deviations, $\rho_{i,j}$ is a correlation coefficient between stocks i and j. With $\gamma = 0, -1$ matrix $\boldsymbol{C}$ becomes covariance and correlation, but those numbers are not sacred. For example the infra-correlation $\gamma < -1$ (figure 5) and ultra-covariance $\gamma > 0$ (figure 6) matrices generate strong signals.

We can add the next largest principal component or combine two previous cases. Two components give the similar result, and eigenvector combination with index or three largest principal components make PnL lower. Figures 7 - 11 show the output of the strategy with regression on 2 principal components.

1.2 "Noise" Strategies

Bouchaud and Potters [8] showed that the density of the eigenvalues of empirical S&P500 stocks daily returns correlation matrix could be perfectly fit by theoretical curve for purely random time-series correlation matrix except for the very few highest eigenvalues. They concluded that a large part of the empirical matrices must be considered as noise. We try to use this noise for the next step in our strategy building, considering the regression on few smallest principal components as a signal. Instead of fixing the number of eigenvalues we fix the percentage of variance they cover. We sort eigenvalues

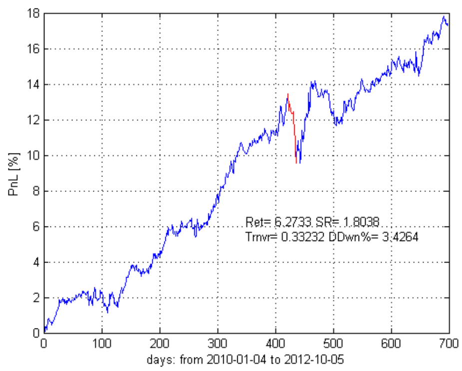

Figure 5: Regression of prices on the largest principal component. Infra-correlation matrix ($\gamma = -2$). Annual return - to 6.27 %, Sharpe ratio - 1.80, turnover - 33%, max drawdown - 3.43%

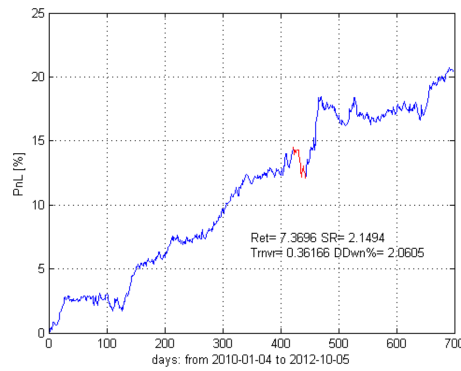

Figure 6: Regression of prices on the largest principal component. Ultra-covariance matrix ($\gamma = 2$). Annual return - to 7.37 %, Sharpe ratio - 2.15, turnover - 36%, max drawdown - 2.06%

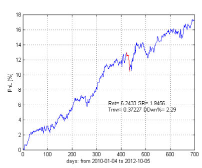

Figure 7: ($\gamma = -3$). Sharpe ratio - 1.95.

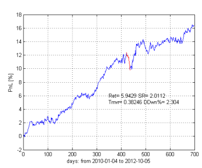

Figure 8: ($\gamma = -1$). Sharpe ratio - 2.01.

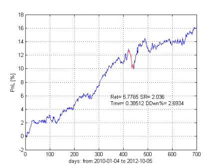

Figure 9: ($\gamma = 0$). Sharpe ratio - 2.04.

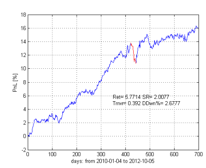

Figure 10: ($\gamma = 1$). Sharpe ratio - 2.01.

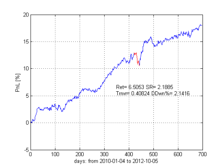

Figure 11: ($\gamma = 3$). Sharpe ratio - 2.19.

and then draw the line between "fair" price and residual when the sum of small eigenvalues exceed the given fraction of the total sum. Figures 12 - 13 show how 1% of variance works as a signal. Tables 1 - 2 give more detailed information for the results of modelling with different residual variances RV. Other parameters: PnL stands for profit and loss, TO ia a turnover and DD is a maximum drawdown. On all PnL plots maximum drawdowns indicated by red color. We used different values of additional parameters, such as aggressiveness of mean-variance optimization coefficient λ and exposure to variance γ to illustrate sensitivity to them. Analysis of volatility of output parameters in tables is the best way to estimate strategies more conservatively, avoiding overfitting.

1.3 Fractional Integration

Fractional Brownian motion (fBm) **and Fractional Gaussian noise** (fGm) Before we worked only with conventional price and return time-series. Now we consider the cases between, i.e. consider fraction-integrated time-series. Fractional integration is based on Couchy formula for repeated

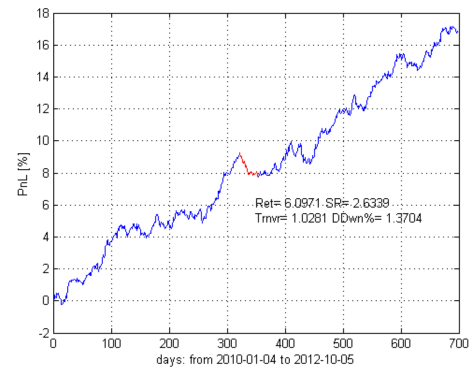

Figure 12: Regression of price time-series on the 1% of smallest principal components. Covariance matrix

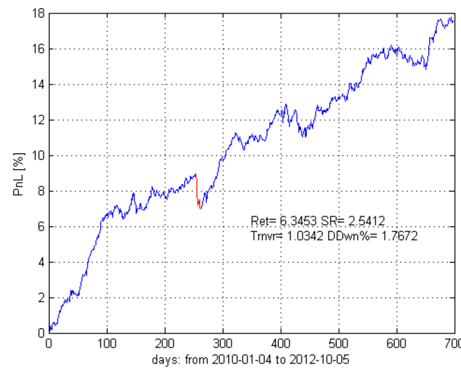

Figure 13: Regression of price time-series on the 1% of smallest principal components. Correlation matrix

$RV\%$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$PnL\%$	5.19	6.79	7.42	7.10	7.78	7.47	7.41	7.06	7.69	6.75
SR	3.0	3.26	3.28	2.85	2.98	2.77	2.66	2.52	2.65	2.23
TO	1.15	1.12	1.08	1.04	1.02	0.99	0.96	0.94	0.91	0.89
$DD\%$	1.53	1.10	1.26	1.46	1.24	1.44	1.69	2.23	1.73	2.1

Table 1: Regression on smallest principal components $\gamma = -1$; $\lambda = 0.5 \cdot 10^{-4}$, RV is a residual variance, PnL stands for profit and loss, TO ia a turnover and DD is a maximum drawdown

integration

$$I^n f(x) = \frac{1}{(n-1)!} \int_0^x (x-t)^{n-1} f(t) dt \tag{11}$$

Its generalization for real n is straightforward

$$I^\nu f(x) = \frac{1}{\Gamma(\nu)} \int_0^x (x-t)^{\nu-1} f(t) dt \tag{12}$$

The corresponding Holmgren - Riemann - Liouville fractional integral [9] defines the stochastic process

$$X^H(t) = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB_s, \tag{13}$$

dB_t is a white noise, H is a Hurst exponent. Equation (13) is different from the stochastic integral representation of fractional Brownian motion (fBm) in classic Mandelbrot and Van Ness paper [9], but they are equivalent for our purposes.

$RV\%$	1	2	3	4	5	6	7	8	9	10
$PnL\%$	5.18	4.43	5.24	5.31	5.79	6.29	5.90	5.75	5.79	5.85
SR	2.99	2.17	2.32	2.13	2.23	2.26	1.98	1.88	1.84	1.83
TO	0.77	0.67	0.61	0.56	0.53	0.51	0.49	0.47	0.45	0.44
$DD\%$	1.34	1.64	1.83	3.23	2.35	3.15	3.27	2.86	2.54	2.57

Table 2: Regression on smallest principal components $\gamma = -0.6$; $\lambda = 2{\cdot}10^{-4}$, RV is a residual variance, PnL stands for profit and loss, TO is a turnover and DD is a maximum drawdown

$$X^H(t) = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-1/2} - (-s)^{H-1/2} \right] dB_s + \int_{0}^{t} (t-s)^{H-1/2} dB_s \right\}, \tag{14}$$

Another way to get fractionally integrated time-series is to apply a recursive algorithm for fractional differencing to integrated I^1 series.

$$(1-D)^\nu (I^1 X) = I^{1-\nu} X \tag{15}$$

$$(1-D)^\nu x_t = (1 + \frac{\nu}{1!}(-1)^1 D^1 + \frac{\nu(\nu-1)}{2!}(-1)^2 D^2 + \ldots) x_t = \sum_{i=0}^{\infty} C_i x_{t-i}, \tag{16}$$

where D is a shift operator: $Dx_n = x_{n-1}$ and coefficients C_i have obvious recursive rule:

$$C_0 = 1; C_i = C_{i-1} \frac{i-\nu-1}{i} \tag{17}$$

A normalized fractional Brownian motion is uniquely characterized by the following properties:

- $X_H(t)$ has stationary increments;
- $X_H(0) = 0$ and $\mathbb{E}(X_H(t)) = 0$ for $t \geq 0$;
- $\mathbb{E}X_H^2(t) = t^{2H}$ for $t \geq 0$;
- $X_H(t)$ has a Gaussian distribution for $t > 0$.

From the first three properties it follows that the covariance function is given by

$$\rho(s,t) = \mathbb{E}X_H(s)X_H(t) = \frac{1}{2}\left\{ t^{2H} + s^{2H} - (t-s)^{2H} \right\} \tag{18}$$

for $0 < s \leq t$. For Gaussian processes, the mean and covariance structure determine the finite-dimensional distributions uniquely.Therefore, from homogeneity of Equation (18)we conclude that fractional Brownian motion with Hurst parameter H is self-similar with Hurst parameter H.

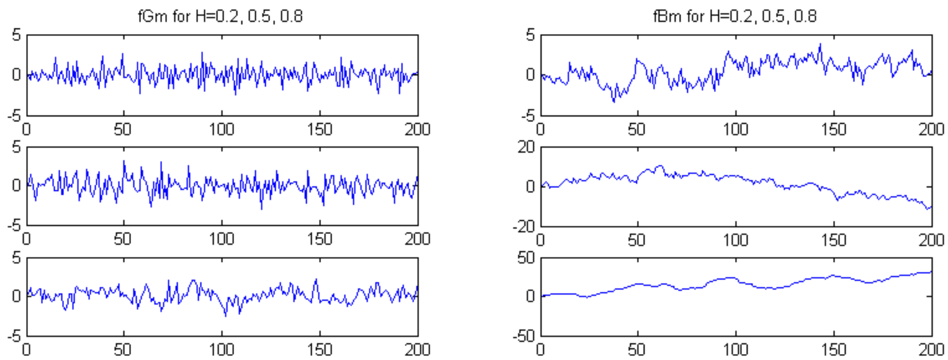

Figure 14: Fractional Gaussian noise and Fractional Brownian motion

The incremental process $R_H = X_H(k+1) - X_H(k+1), k = 0, 1 \ldots$ of fractional Brownian motion, which is called fractional Gaussian noise has a standard normal distribution for every k, but there is no independence. The corresponding autocovariance function

$$\rho_R(k) = \frac{1}{2}\Big\{|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\Big\} \tag{19}$$

is equal to zero only for $H = 1/2$. The $fBms$ corresponding to $0 < H < 1/2$, $1/2 < H < 1$, and $H = 1/2$, differ in many significant ways. If $1/2 < H < 1$, then the covariance is positive and series diverges

$$\rho_R(k) > 0; k \neq 0 \quad \sum_{k=0}^{\infty} \rho_R(k) = \infty$$

On the other hand, if $0 < H < 1/2$, the covariances are negative and the sum of their absolute values converges. We illustrate this behavior on figures (14). A positive covariance (bottom row) means that positive (negative) returns are followed by returns of the same size more frequently, and we can expect cluster phenomena. A negative covariance (top row) leads to strong intermittency or,in other words, to mean-reversion pace. A case $H = 1/2$ (middle row)is special: this is a classic Gaussian sequence of independent identically distributed variables or white noise.

Spectral Analysis For stationary stochastic processes, the spectral density is computed as follows [10], [11],[12]:

$$f(\lambda) = 2\sin(\pi H)\Gamma(2H+1)(1-\cos\lambda)\Big[\lambda^{-2H-1} + \sum \Big\{(2\pi+\lambda)^{-2H-1} + (2\pi-\lambda)^{-2H-1}\Big\}\Big] \tag{20}$$

In the article [12] the approximation for the sum $S(\lambda, H)$ at the right side of (20) was suggested

$$S(\lambda, H) = (1.0002 - 0.000134\lambda)\Big\{S_3(\lambda, H) - 2^{-7.65H-7.4}\Big\}$$

$$S_3(\lambda, H) = \Sigma_{j=1}^{3}\Big\{(a_j^+)^{-2H-1} + (a_j^-)^{-2H-1}\Big\} + \frac{(a_3^+)^{-2H} + (a_3^-)^{-2H} + (a_4^+)^{-2H} + (a_4^-)^{-2H}}{8H\pi} \tag{21}$$

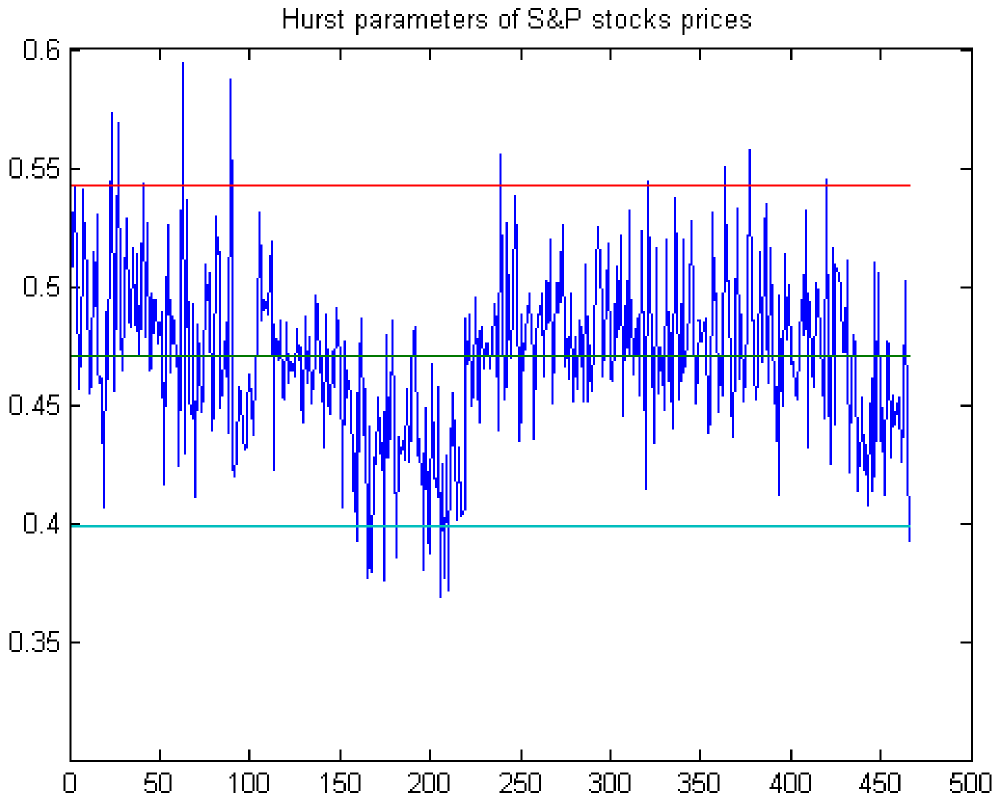

Figure 15: S&P500 Hurst parameters. *The mean* and *the mean* $\pm 2 * std$ lines

where $a_j^{\pm} = 2\pi j \pm \lambda$

Figure 15 shows the estimation of Hurst parameters for the $S\&P500$ stocks for the last 5 years. The mean is equal to 0.47 and standard deviation is equal to 0.036. Figure 16 illustrates in log-log scale different colored Gaussian noises. The color was calculated as the weighted by average amplitude of the third part of spectrum RGB mixture. Higher amplitude for higher frequency (”blue noise”) means negative covariance,strong intermittency or fast up and down alternation. This is a characteristic f hydrodynamic turbulence. At the right upper corner you can see a typical spectrum of stock return series with the Hurst parameter $H = 0.45$. With $H = 0.5$ we would have perfectly white noise. On the other side of whiteness we can see the ”pink noise” that getting more and more ”pure red” with the $H \longrightarrow 1$. It is featured by a ”strong aftereffect”, ”persistence” and ”long memory”. The parameter H between 0.5 and 1 is what H. Hurst himself observed in the fluctuations of yearly run-offs of Nile and some other rivers. Long memory was discovered in characteristics of high frequency

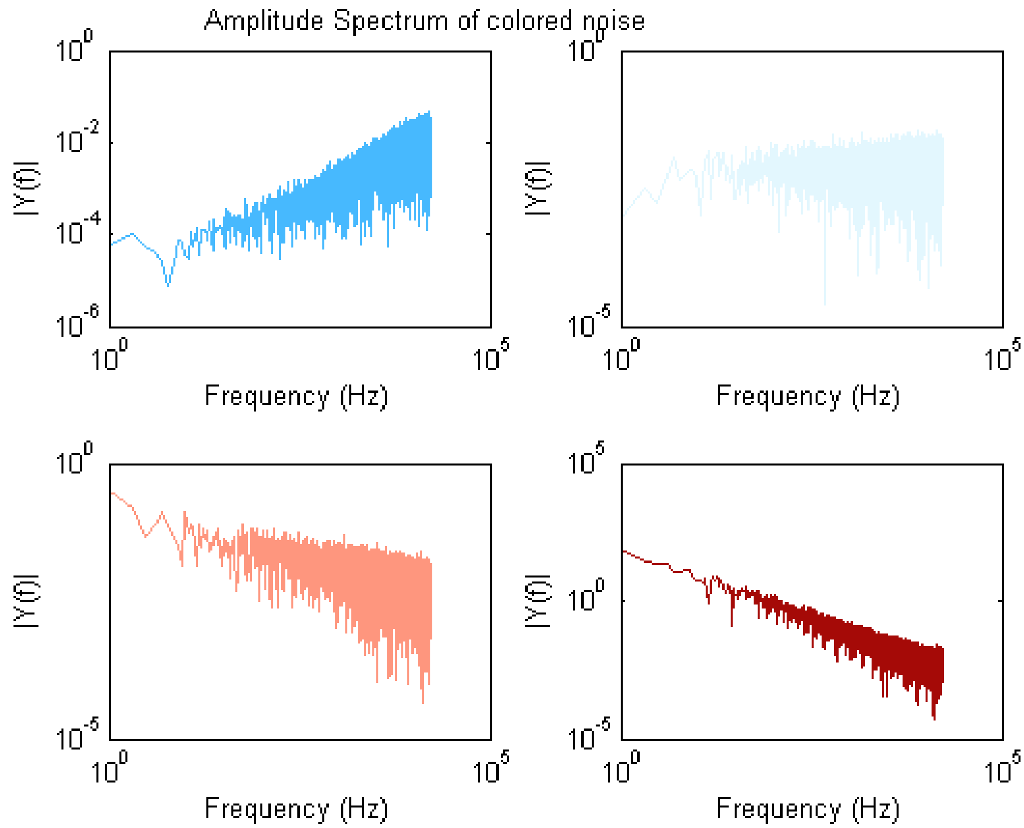

Figure 16: Fractional Gaussian noises spectrum: blue H=0.01, almost white (H=0.45), pink(0.80) and Brownian

trading, particularly in a series of signs of trades. The "Brownian noise" is the spectrum of Gaussian Brownian motion (a cumulative sum of "white noise") and a mixture of mostly red and "black noise" (zero frequency). It is different from other cases because it is not a stationary process that should have $0 < H < 1$.

Fractional Brownian motion trading strategies Figures 17 and 18 show Hurst parameters for principal components sorted by eigenvalues. At left side of the figures the axis X (logarithmic) is the cumulative relative variance in percents that cover principal components from the least volatile to the current one. In sample, with 1202 time points and 466 series we see that the least volatile principal components are stationary. If we could find stationary out of sample price time-series, we would have a perfect trading strategy. The lower row of figures 17 and 18 prove the complete dissipation of that property on the next one year interval. Fractional differentiation doesn't give any visible

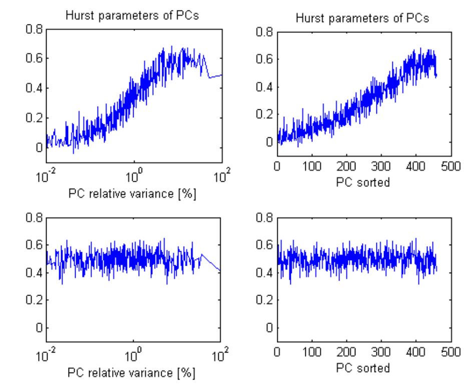

Figure 17: Hurst parameters for prices. Upper row - in sample, lower row - out of sample

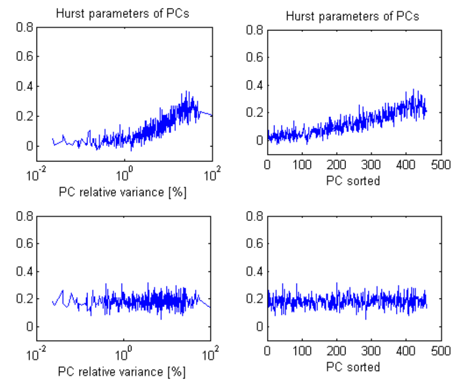

Figure 18: Hurst parameters for fraction - integrated return. Upper row - in sample, lower row - out of sample.

improvements. However, the positive "pre-tax" profit of our strategies confirms that the cointegrated systems do not disappear instantly. Figures 19 - 22 show the dynamics of trading portfolio with the Sharpe Ratio up to 4.10.

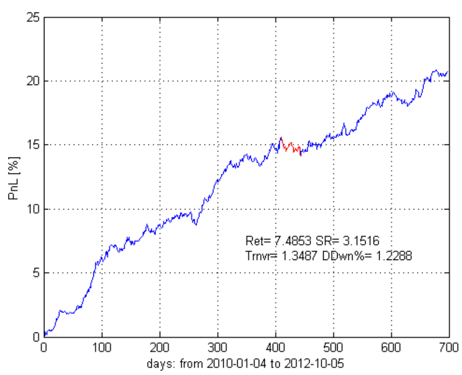

Figure 19: Regression of fractionally integrated return on the 4% of smallest principal components. Covariance matrix. $H = 1/3$

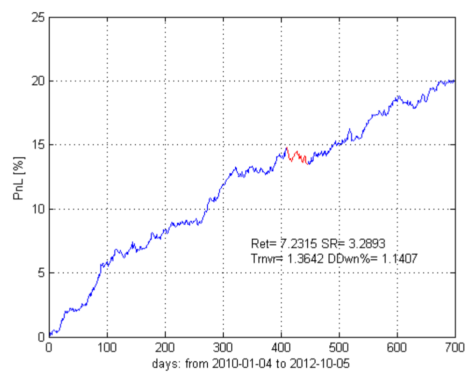

Figure 20: Regression of fractionally integrated return on the 4% of smallest principal components. $\gamma = -2/3$, $H = 1/3$

We added figures 23 - 24 to show results of out-of-sample tests for the same set of parameters as at 2 previous plots.

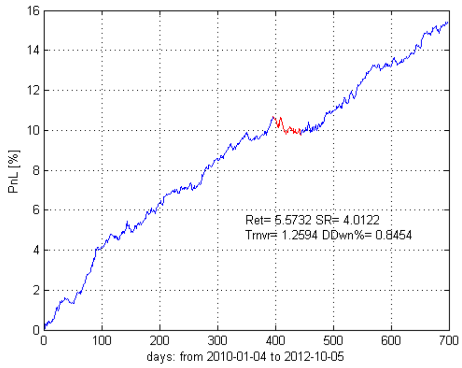

Figure 21: Regression of fractionally integrated return on the 4% of smallest principal components. $\gamma = -0.7$, $H = 0.3$

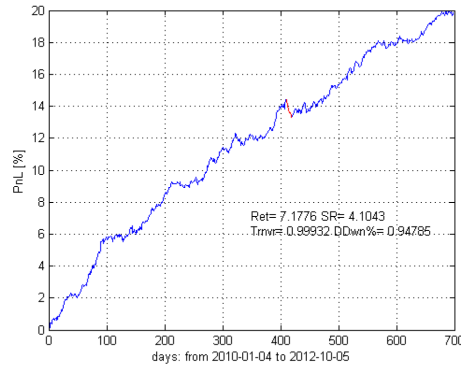

Figure 22: Regression of fractionally integrated return on the 0.3% of smallest principal components. $\gamma = -0.6$, $H = 0.3$

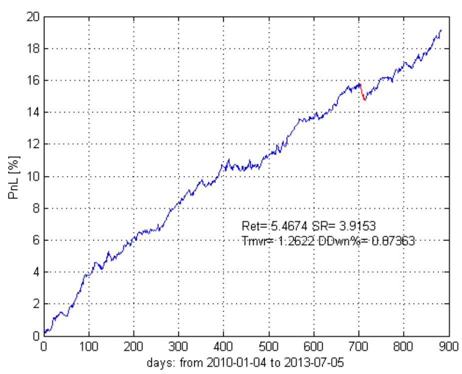

Figure 23: Regression of fractionally integrated return on the 4% of smallest principal components. $\gamma = -0.7$, $H = 0.3$. Out-of-sample test.

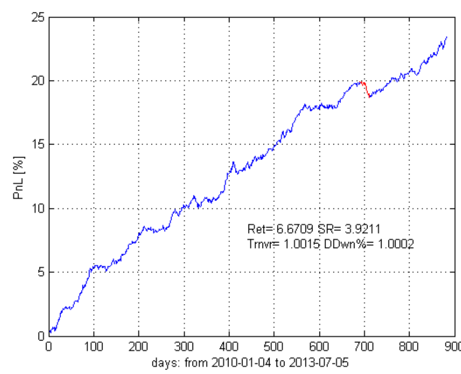

Figure 24: Regression of fractionally integrated return on the 0.3% of smallest principal components. $\gamma = -0.6$, $H = 0.3$. Out-of-sample test.

2 Execution

R. Grinold and R. Kahn [4] and R.Almgren and N.Chriss [1, 3] independently pioneered application of calculus of variation to the problem of portfolio liquidation. In the meantime most modern trading engines use some modifications of that approach. They found the maximum of mean-variance utility function Mean-variance utility is given by

$$\Phi = \int_0^t (E(R) - \lambda Var(R))dt = \int_0^T (\alpha x + h[x]\dot{x} - \lambda(Sx\sigma)^2)dt \tag{22}$$

where S is a stock price

R is an absolute asset price return in \$

Var is a variance of that return

$x(t)$ is a current position in stock

$h[x]$ is a market impact functional

T is a time horizon

λ is a risk-averse parameter.

$\alpha(t)$ is a signal or expected return (drift) $\alpha = E(R)$

Traditionally impact is divided by two parts: permanent and temporary. Permanent impact is a linear function of the amount traded only, because otherwise we could move the value of particular asset cyclically buying and selling the same number of shares but with different speed. We don't include this permanent impact in our analysis for two reasons: first - doesn't depend on trading trajectories and second - typically the sizes of proprietary traders transactions don't exceed few percent of average daily volume (ADV) and correspondingly are negligible fractions of capitalization.

2.1 Market impact models

Execution price is path-dependent in general through the affect by previous trading history. The difference between arrival price and execution price a.k.a implementation shortfall is supposed to be of opposite sign to direction of trade: we pay more to increase our position and get less to unwind it.

$$h[x] = -\int_0^t f(\dot{x}(\tau))K(t,\tau)dt \tag{23}$$

Market impact kernel $K(\tau, t)$ is assumed homogenous in volume time $K(\tau, t) = K(t - \tau)$. Initial time is set to 0. Market impact is assumed to be linear functional of trading rate

$$\begin{aligned} h(t, \dot{x}) &= -\int_0^t (\tilde{\eta}K(t-\tau)\dot{x}(\tau))d\tau \\ \tilde{\eta} &= \eta\frac{\sigma}{ADV\,S_0} \end{aligned} \tag{24}$$

G.K.A.C. model The simplest form of the convolution integral kernel was proposed by Grinold and Kahn [4] and analyzed in details by Almgren and Chriss [1, 2] (G.K.A.C.)

$$K(t-\tau) = \delta(0), \qquad Dirac's\ delta\ function$$

Utility function (22) is given by

$$\begin{aligned} \Phi[x] &= -\int_0^T \Psi(x,\dot{x},t)\,dt \rightarrow max \\ \Psi(x,\dot{x},t) &= -\alpha x + \eta\dot{x}^2 + \lambda(xS\sigma)^2 \end{aligned} \tag{25}$$

Subject to initial and terminal conditions:

$$x(0) = X_0,\ x(T) = X_T \tag{26}$$

Maximization of (25) is a standard calculus of variation problem. Applying The Euler equation

$$\frac{\partial\Psi}{dx} - \frac{d}{dt}\frac{\partial\Psi}{\partial\dot{x}} = 0 \tag{27}$$

we obtain

$$\ddot{x} - k^2 x = -\tilde{\alpha} \tag{28}$$

Our solution is a simple generalization of Almgren & Chriss [2] to the drift with exponential decay

$$\begin{aligned} x\,(t) &= (X_0 + a(0))\,\frac{\sinh\left(k(T-t)\right)}{\sinh\left(kT\right)} + (X_T + a(T))\,\frac{\sinh\left(kt\right)}{\sinh\left(kT\right)} - a(t) \\ \dot{x}\,(t) &= \frac{dx}{dt} = -\,(X_0 + a(0))\,\frac{k\cosh\left(k(T-t\right)}{\sinh\left(kT\right)} + (X_T + a(T))\,\frac{k\cosh\left(kt\right)}{\sinh\left(kT\right)} - \dot{a}(t) \end{aligned} \tag{29}$$

where

$\alpha(t) = \alpha_0 \exp(-\gamma t)$
$k^2 = \lambda/\tilde{\eta}$
$\tilde{\alpha} = \alpha/(2\tilde{\eta})$
$a(t)$ is a particular solution of equation (28)
$a(t) = -\tilde{\alpha}(t))/\gamma^2 - k^2)$

Exponential kernel The solution (29) is a decent approximation for optimal trajectories on practice, but it cannot describe the market impact of a single discrete trade (it is a delta-function). The instantaneous recovery assumption is unrealistic and inconsistent with calibration procedures. To resolve those problems we have to replace delta function with some smooth kernel. The function under integral in (25) is given by:

$$\Psi(x,\dot{x},t) = -\alpha x + \eta\int_0^t \dot{x}(\tau)K(t-\tau)d\tau + \lambda(xS\sigma)^2 \tag{30}$$

Calculating variations [1] we obtain the following equation

$$2\tilde{\lambda}x = \dot{F} = \frac{d}{dt}\int_0^T \dot{x}(\tau)K(|t-\tau|)d\tau \tag{31}$$

Assuming that the kernel is twice differentiable gives

$$2\tilde{\lambda}\ddot{x} = \frac{d}{dt}(\int_0^T \dot{x}(\tau)\ddot{K}(|t-\tau|)d\tau + 2\dot{x}\dot{K}_0(0))$$

Now we plug the exponential kernel

$$K(t-\tau) = \beta \cdot exp(-\beta(t-\tau))$$

into generic equations and get the system:

$$\begin{aligned} &\ddot{x} - \frac{\lambda\beta^2}{\lambda+\beta^2}x = \frac{\beta^2\alpha + \ddot{\alpha}}{2(\lambda+\beta^2)} \\ &\ddot{F} - \beta^2 F = -2\beta^2\dot{x} \end{aligned} \tag{32}$$

We presented a solution of the system in simple and familiar form [6]:

$$\begin{aligned} &x\left(t\right) = \left(X_0 + a\right) B\frac{\sinh\left(\left(k(T-t\right) + A\right)}{\sinh\left(kT + 2A\right)} + \left(X_T + a\right) B\frac{\sinh\left(kt\right)}{\sinh\left(kT + 2A\right)} - a \\ &where \\ &A = \ln\sqrt{\frac{\beta+k}{\beta-k}}, \quad B = \frac{k}{\sqrt{\lambda}} \end{aligned} \tag{33}$$

For risk-neutral traders ($\lambda \to 0$) and $\alpha = 0$ the optimal schedule under exponential impact relaxation is a combination of two jumps and straight line between them.

$$\begin{aligned} &\lim_{\lambda\to 0} x = (X_0 - \Delta X_0)\frac{T-t}{T} + (X_T + \Delta X_T)\frac{t}{T} \\ &\Delta X_0 = \Delta X_T = \frac{X_0 - X_T}{\beta T + 2} \end{aligned} \tag{34}$$

Optimal trading strategy with the exponential kernel was the subject of A.Obizhaeva and J.Wang, [7]. They probably were the first who pointed out to discontinuity of optimal paths at the ends of time interval and derived equation (34) for risk-neutral traders. Both discrete and continuous regimes of trading were considered using dynamic-programming approach. They also obtained a solution in quadratures for more general case, but it is much more complex then our analytical formula.

[1]

$$\delta_{\dot{x}}\Psi = \int_0^T [\delta(\dot{x}(t))\int_0^t \dot{x}(\tau)K(t-\tau)d\tau + \dot{x}(t)\int_0^t \delta(\dot{x}(\tau))K(t-\tau)d\tau]dt$$

We change the order of integration for the second integral and get

$$\delta_{\dot{x}}\Psi = \int_0^T \int_0^t \dot{x}(\tau)K(|t-\tau|)d\tau\delta(\dot{x}(t))dt$$

2.2 Multi-periods portfolio trading

We assume that the asset can be traded for unknown number of periods with the breaks between them. During the breaks the trading rate is zero. The length of overnight (lunch) breaks Δt is calculated as the ratio of close-to-open (lunch) variance to day (half day) variance. Inside each active trading interval we have Almgren & Chriss trajectory:

$$
\begin{aligned}
x\left(t\right) &= \left(X_{n-1} + \alpha^*\right) \frac{\sinh\left(k(T-t)\right)}{\sinh\left(kT\right)} + \left(X_n + \alpha^*\right) \frac{\sinh\left(kt\right)}{\sinh\left(kT\right)} - \alpha^* \\
\dot{x}\left(t\right) &= \frac{dx}{dt} = -\left(X_{n-1} + \alpha^*\right) \frac{k\cosh\left(k(T-t\right)}{\sinh\left(kT\right)} + \left(X_n + \alpha^*\right) \frac{k\cosh\left(kt\right)}{\sinh\left(kT\right)}
\end{aligned}
\tag{35}
$$

We assume that the sign of position is held inside time intervals. We consider the simple case when the α is constant and exists only for the first active trading period.

To define $N+1$ of unknown initial and terminal values of our position X_n we have two boundary equations at the very beginning and the very end

$$
\begin{aligned}
x\left(0\right) &= X_0 \\
x\left(T\right) &= 0
\end{aligned}
\tag{36}
$$

and $N-1$ transversality conditions

$$
\left.\frac{\partial \Psi}{\partial \dot{x}}\right|_{t_n - 0}^{t_n + 0} + \left.\frac{\partial \Psi}{\partial \dot{x}}\right|_{t_n + \Delta t - 0}^{t_n + \Delta t + 0} + \frac{dG}{dx_n} = 0 \tag{37}
$$

where $G(x_n, t_n) = \int_{t_n}^{t_n + \Delta t} \Psi(x, q, t)\, dt$

$$
\begin{pmatrix}
b & c & & & \\
a & b & c & & \\
 & a & b & c & \\
\vdots & \vdots & \ddots & \ddots & \\
 & & & a & b
\end{pmatrix}
\cdot
\begin{pmatrix}
x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1}
\end{pmatrix}
=
\begin{pmatrix}
r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_{N-1}
\end{pmatrix}
\tag{38}
$$

Assuming equal time intervals $t_n = t_1$ we get:

$$
\begin{aligned}
& a = c = 1 \\
& b = -2(\Delta t \cdot k \cdot sinh(kt_1) + cosh(kt_1)) \\
& r_n = -\alpha \cdot (2 \cdot (1 - cosh(kt_1)) - \Delta t \cdot sinh(kt_1)/k) \\
& r_1 = r_1 - x_0
\end{aligned}
\tag{39}
$$

We presented the fast and simple method of solution in appendix (A.2). We extend the execution time iteratively until the position or direction of trade change their signs. In both cases it means that the initial position was liquidated on the last time interval. The optimal time of execution t^* can be found applying additional boundary condition for the rate of trade.

$$\begin{aligned} x\left(t^*\right) &= 0 \\ \dot{x}\left(t^*\right) &= 0 \end{aligned} \tag{40}$$

Substituting solution 35 for the last interval into (40) we get $t^* = \frac{1}{k} acosh((x_{n-1} + \alpha)/\alpha)$

2.3 Inventory Management

For portfolio trading the approach will be same, but instead of one equation we will have a system of $ODEs$ (A.1), that after the orthogonalization gives for each eigenvector the solution of the form (29).

$$z\left(t\right) = Z_{n-1} \frac{\sinh\left(k(T-t)\right)}{\sinh\left(kT\right)} + Z_n \frac{\sinh\left(kt\right)}{\sinh\left(kT\right)} \tag{41}$$

For simplification we dropped the price drift term.

Constraints Soft constraints (additional quadratic terms to objective functions) is a flexible tool to modify the optimal trajectory without changing complexity of the problem. We consider cash and market neutrality constraints on portfolio level and a combination of cash neutrality and relative variance on sector level.

$$C_{ij} = C_{ij} + \lambda_{cash} + \lambda_{market} \cdot \beta_i \beta_j \tag{42}$$

Usually it is not a good idea to apply hard equality constraints to the asset positions as functions of time. Those constraints would require immediate move regardless of cost of trade. At the same time the goals like to have cash, market or even sector neutral portfolio at the end of the day are not hard to achieve.

Naïve and not so naïve strategies It is always a temptation to apply well known from adjacent areas results to a new problem. In our case the source of solutions is obvious - the Portfolio Management. However this natural approach has limitations. The classic mean-variance investment portfolio optimization is similar but not identical to the dynamic problem of trading portfolio management.

$$\Phi(x) = x^t C x \rightarrow min \tag{43}$$

Solving the system above we are looking for the cash-neutral minimal variance portfolio given the maximum market participation rate. This strategy is naive or shortsighted because we don't consider neither different transaction cost to obtain this intermediate portfolio nor the risk and market impact during following liquidation. We can improve that approach making strong but reasonable assumptions about unwinding/hedging trajectories. The optimal time of execution is infinite and a trajectory for a single stock or independent part of portfolio (eigenvector) is an exponential decay.

$$\Phi(x) = (X-x)^t \eta (X-x) + \lambda(x^t C x + z^t \frac{e^{2k}}{k^2} z) \to min \tag{44}$$

3 Results and discussion

Wide universe. In this section we present more up-to-date, more realistic and on wider universe results. The methodology we used is similar to the best approaches in Section 1, but the parameter values and other details of trading strategies are intentionally hidden.

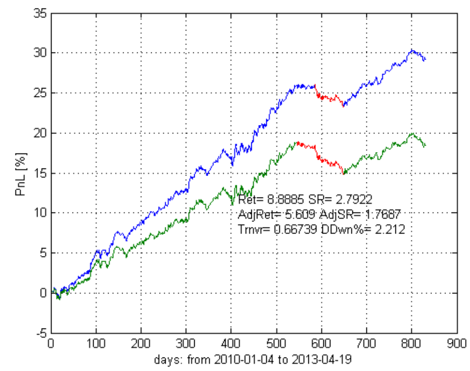

Figure 25: Regression of price time-series. Uniform capital allocation

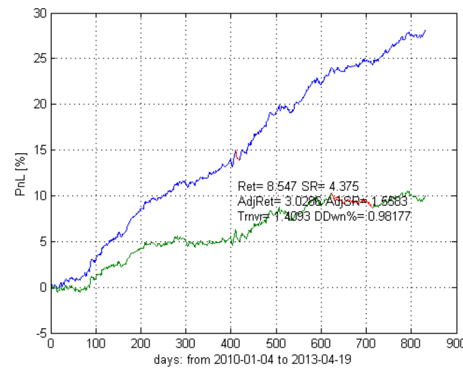

Figure 26: Regression of fractionally integrated return time-series. Uniform capital allocation

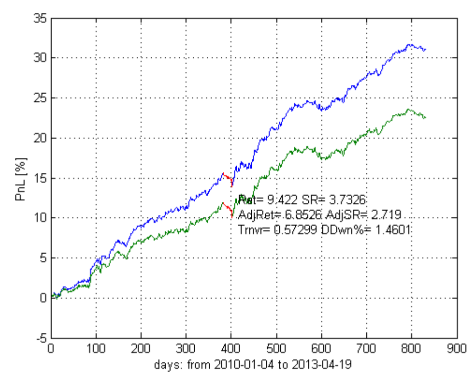

Figure 27: Regression of price time-series. Optimal solution

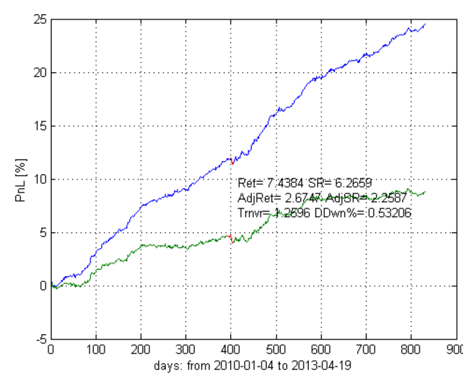

Figure 28: Regression of fractionally integrated return time-series. Optimal solution

The strategies shown on Figures 19,20,21,22 look attractive, but we showed the results in a perfect world without friction, i.e. market impact, fees and other chargers. We simulated perfect world (blue lines) and adjusted to fees and market impact (green lines) returns on Figures 25,26,27,28 The results at lower level (Figures 27,28) are better then at upper one (Figures 25,26). The results at upper level were obtained by allocation of 1% of notional capital (in our simulation - \$10 *million*) into stocks with 50 highest positive and 50 highest negative signals. Lower level charts proves that the portfolio optimization is a powerful method to increase return and reduce risk. Optimization looks as an unconditionally better choice in strategies implementation, but it is not. The special financial optimization software can be efficient and comparatively friendly for traders but they are very expensive, e.g. $Axioma\ Portfolio\ Optimizer$. It Is easy to find free or almost free general optimization packages, but the development and maintenance of trading engines with them requires more advanced and professional people. From left (Figures 25, 27) to right (Figures 26, 28) the theoretical values of Sharpe ratio grow but adjusted ones slightly drop. The reason is in much higher turnover for fractional integrated time-series. Actually turnover greater then 100% means that the portfolio rebalancing was overdue: signals have shorter life then intervals between trades. It is an indication that the more frequent data is required for analysis and, correspondingly, intraday portfolio rebalancing should be considered.

Out of the box. In the section 1.2 I used the unconventional names "ultra-covariance" matrix and "infra-correlation" matrix. Rather than to coin new terms and "to multiplied entities beyond necessity" my intention was to to point out that something useful can exist out of the box - even if we never saw or heard it. Covariance matrix is the natural measure for the relations between time-series of similar stocks like the stocks of the same sector of $S\&P500$ index. Largest principal component (the component with the highest variance) is a good indicator of the sector behavior and can be chosen as a next generation ETF. Note, that the volatilities of $S\&P500$ stocks are not very different inside their sectors. For wider universe using covariance matrix for PCA means giving more weights to most volatile small cap stocks. Correlation (normalized covariance) matrix eliminates that premium. To give the premium to most solid and "flat" stocks we need to multiply correlation coefficients by standard deviations in negative power. The step in opposite direction also can be explained (good theoretician should be able to explain everything!): more volatile stocks are more sensitive to trends and therefore serve as better indicators of future returns.

There are many other combinations of time-series and methods of their processing and filtering. The most obvious one is a combination with the trading volume. Using changes in market value (returns times volume traded) sequences seems pretty logical. But the opposite operation - calculating volume time returns, i.e. dividing returns by trading volume looks logical as well and very popular. Volume traded is better measure of events happened than calendar time. Actually both ideas are far from perfect as is, because of very high volatility of daily volumes. Short time (including daily and weekly) volume time-series would completely overshadow the most important price moves. Some smoothing or different out-of-the-box tricks required to incorporate that additional information.

In-sample or out-of-sample? All time-series combinations and especially data mining such as smoothing and filtering are significantly nonlinear procedures, all of them make the models less transparent and therefore make proper formal statistic tests hard to design and even harder to believe.

Pure out-of-sample tests are easy to run for high frequency strategies and almost mentally impossible for daily trading strategies. Ideally the researcher should tune all parameters in sample leaving at least a hundred points (~ 5 months) intact. After single out-of-sample run any further changes in working version should be postponed for the next 5 months. The possibility to use more sophisticated constructions that could provide some warranty against overfitting, for example, bootstrapping is very limited because of subtle nature of portfolio level signal generation. A.Inoue and L.Kilian [14] questioned and "overturned the conventional wisdom that out-of-sample test results are more reliable than in-sample test results". Long time horizon and smooth and wide maximums of PnL functions (see Table 1 and Table 2) of the most sensitive parameters are more important than unverified virginity of relatively small samples. Anyway, the electronic form of this publication allow me to respond promptly to the questions about in-sample - out-of-sample tests and add updated results - see Figures 23 and 24. The results presented are not "the best" - I didn't try to maximize the PnL for all static parameters in in-sample tests or make them dynamic (changing day by day during simulation). Also I didn't treat sectors separatively.

Even the most primitive of statistical arbitrage trading strategies, as we saw, can give us formally the performance that is much better than the historical performance of indices. And this free lunch is paid by other expenses: investments in smart trading engines and dynamic liquidity adjusted optimization software.

I don't discuss *Execution* section here. There are many fundamental and practical problems in market impact and optimal execution theories that are not covered in literature. I am going to address some of them in near future.

References

[1] Almgren R. and N. Chriss (2001) Optimal Execution of Portfolio Transactions, *Journal of Risk*, **3(2)**, 5–40.

[2] Almgren, Robert, 2003, Optimal execution with nonlinear impact functions and trading enhanced risk, *Applied Mathematical Finance* **10**, 118.

[3] Almgren, Robert, and Neil Chriss, 1999, Value under liquidation, *Risk*, **12** (12).

[4] Grinold Richard, and Kahn Ronald, 1999, Active Portfolio Management, McGraw-Hill, New York.

[5] BARRA, Market impact model handbook (Berkeley, California, Barra, 1997).

[6] Skachkov Igor, 2010 Optimal Execution: Linear Market Impact with Exponential Decay. *http://ssrn.com/abstract=2283027*

[7] Obizhaeva, Anna and Wang, Jiang , 2005 Optimal trading strategy and supply/demand dynamics. *MIT working paper*

[8] Bouchaud, Jean-Philippe and Potters, Marc, 2000, Theory of Financial Risks, Cambridge U. Press, Cambridge, UKAC1999

[9] Mandelbrot, B.B. and Van Ness J.W. Fractional Brownian Motion, fractional Noises and Applications. SIAM Review, 10:422-437, 1968

[10] Shiryaev, A.. 1999, Essential of stochastic finance: Facts, Models, Theory, World Scientific,Singapore

[11] Dieker,Ton. 2004, Simulation of fractional Brownian motion, Master Thesis, Amsterdam

[12] Paxson, V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic, Computer Communication Review, 27 (1997), pp. 5-18.

[13] Gatheral J., Schied A. (2013) Dynamical models of market impact and algorithms for order execution. Preprint. **URL:** *http://ssrn.com/abstract=2034178*

[14] Inoue, Atsushi and Kilian, Lutz, In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? (November 2002). ECB Working Paper No. 195. Available at SSRN: http://ssrn.com/abstract=358500

A Appendix

A.1 Multidimensional optimization

Unconstrained N-dimensional optimization could be divided to N one-dimensional problems AC (2000). We rewrite equations (25) in matrix form

$$
\begin{aligned}
&\Phi[x] = -\int_0^T \Psi(x, \dot{x}, t)\, dt \to max \\
&\Psi(x, \dot{x}, t) = -\boldsymbol{\alpha^t} x + \dot{x}^t \boldsymbol{\eta} \dot{x} + \lambda \cdot x^t \Omega x
\end{aligned}
\tag{45}
$$

where

α is a vector of expected returns

$\boldsymbol{\eta}$ is diagonal matrix of market impact coefficients.

Ω is a covariance matrix

The chain of linear transformations

$$
\begin{aligned}
&\dot{x}^t \boldsymbol{\eta} \dot{x} = \dot{x}^t \eta^{1/2} \eta^{1/2} \dot{x} = (\eta^{1/2}\dot{x})^t \cdot (\eta^{1/2}\dot{x}) \\
&\lambda \cdot x^t \Omega x = (\eta^{1/2} x)^t (\lambda \eta^{-1/2} \Omega \eta^{-1/2})(\eta^{1/2} x) = \\
&(\eta^{1/2} x)^t V^t U V (\eta^{1/2} x) = (V \eta^{1/2} x)^t U (V \eta^{1/2} x) = z^t U z \\
&(\eta^{1/2}\dot{x})^t \cdot (\eta^{1/2}\dot{x}) = (\eta^{1/2}\dot{x})^t V^{-1} V (\eta^{1/2}\dot{x}) = (\eta^{1/2}\dot{x})^t V^t V (\eta^{1/2}\dot{x}) = \dot{z}^t \dot{z} \\
&\alpha^t x = \alpha^t (\eta^{-1/2} V^t)(V \eta^{1/2} x) = ((\mathcal{T}^{-1})^t \alpha)^t (\mathcal{T} x) = \alpha_z^t z \\
&where \\
&V \text{ is a matrix of eigenvectors} \\
&U \text{ is a diagonal matrix of eigenvalues} \{k_1^2, k_2^2, \ldots, k_N^2\} \\
&\mathcal{T} \text{ is the transformation matrix: } z = \mathcal{T} x
\end{aligned}
$$

makes the problem split.

$$
\begin{aligned}
&\Phi[z] = -\int_0^T \Psi(z, \dot{z}, t)\, dt \to max \\
&\Psi(z, \dot{z}, t) = -\boldsymbol{\alpha_z^t} z + \dot{z}^t \dot{z} + z^t U z
\end{aligned}
\tag{46}
$$

Applying Euler equation to (46) we get the system of N independent differential equations:

$$
\ddot{z}_i - k_i^2 z = -\alpha_{zi} \quad i = 1, 2, \ldots, N \tag{47}
$$

A.1.1 Boundary conditions

For the first boundary problem:

$$x(0) = X_0, \quad x(T) = X_T \tag{48}$$

we present the solution of (47) in the form:

$z(t) = \mathcal{S}_1(t)C_1 + \mathcal{S}_2(t)C_2$

where

$\mathcal{S}_1(t)$ *and* $\mathcal{S}_2(t)$are two sets of independent solutions of homogeneous Euler equations

or explicitly

$$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} \sinh(k_1(T-t)) & 0 & \dots & 0 \\ 0 & \sinh(k_2(T-t)) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sinh(k_N(T-t)) \end{pmatrix} \begin{pmatrix} C_{11} \\ C_{12} \\ \vdots \\ C_{1N} \end{pmatrix} + \begin{pmatrix} \sinh(k_1 t) & 0 & \dots & 0 \\ 0 & \sinh(k_2 t) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sinh(k_N t) \end{pmatrix} \begin{pmatrix} C_{21} \\ C_{22} \\ \vdots \\ C_{2N} \end{pmatrix}$$

Multiplying boundary conditions (48) by transformation matrix we get

$$\begin{pmatrix} \mathcal{S}_{10} & \mathcal{S}_{20} \\ \mathcal{S}_{1T} & \mathcal{S}_{2T} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} Z_0 \\ Z_T \end{pmatrix}$$

where

$$\mathcal{S}_0 = \mathcal{S}(0) \quad \mathcal{S}_T = \mathcal{S}(T)$$

With our choice for functional form of solution equations to define C_1 and C_2 would be divided

$$C_1 = \mathcal{S}_{10}^{-1} Z_0 \quad C_2 = \mathcal{S}_{2T}^{-1} Z_T$$

If all left or all right (or both) conditions are defined for trading rates instead of initial positions the procedure is very similar. We just need to replace sinh(.) by cosh(.) in corresponding set of solutions. Those homogeneous conditions are typical on practice, but if we have to calculate marginal cost of particular position we should consider more special case:

$$(x_1, x_2, \dots, \dot{x}_k, \dots, x_N)^t|_{t=0} = (X_1, X_2, \dots, Q_k, \dots, X_N)^t$$

We denote a vector on the right side as $\tilde{X}$. The matrix equation for arbitrary constants C_1 is

$$(\mathcal{D}_k \mathcal{T}^{-1} \mathcal{S}_1)|_{t=0} C_1 = \tilde{X} \tag{49}$$

where operator

$$\mathcal{D}_k = \begin{pmatrix} E & \vdots & \mathbf{0} \\ \cdots & \frac{d}{dt} & \cdots \\ \mathbf{0} & \vdots & E \end{pmatrix}$$

differentiate the $k-th$ row of the matrix at right and leave other elements intact.

A.1.2 Terminal constraints

Usually it is required that at the end of the day portfolio will satisfy to the set of hard constraints, e.g. cash and β neutrality (total or sector wise), budget constraint and possible fixed exposure to to some particular stocks or groups of stocks.The problem is described by lagrangian:

$$\begin{aligned} \Phi[x] &= -\int_0^T \Psi(x, \dot{x}, t)\, dt + G(T, X) \rightarrow max \\ G(T, X) &= \lambda^t (\mathcal{B}^t X - c) = \sum_{i=1}^{M} \lambda_i (\beta_i^t X - c_i) \end{aligned} \tag{50}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_M)$ is a vector of Lagrange coefficients, $\mathcal{B}$ is a $(N \times M)$ matrix. For example if the only constraint is cash neutrality then

$$\mathcal{B} = (1, \ldots, 1)^t \text{ is a N-dimensional unity vector}$$

As before we change the variable to make liquidity adjusted covariance matrix orthogonal. Additional term in target function doesn't depend on internal points and therefore doesn't change the Euler equations. What it does change is boundary condition at the right end of time interval. It gives rise to the transversality condition with terminal value:

$$\begin{aligned} &\Psi_{\dot{z}}|_{t=T} + G_Z = 0 \\ &\Psi_{\dot{z}} = \frac{d\Psi}{d\dot{z}} \qquad G_Z = \frac{dG}{dZ} \\ &G(T, Z) = \lambda^t (\mathcal{B}^t \mathcal{T}^{-1} Z - c) = \lambda^t (\tilde{\mathcal{B}}^t Z - c) \end{aligned} \tag{51}$$

$$z(0) = Z_0 \quad \dot{z}(T) + \tilde{\mathcal{B}} \lambda = 0 \tag{52}$$

We present the solution in the form that satisfies initial condition:

$$z(t) = \frac{\cosh(k(T-t))}{\cosh(kT)} Z_0 + \frac{\sinh(kt)}{k\cosh(kT)} C \tag{53}$$

For the constants C we have the system of $N+M$ equations:

$$\begin{aligned} &C + \tilde{\mathcal{B}}\lambda = 0 \\ &\tilde{\mathcal{B}}^t(Z_0 + S_{2T}C) - c = 0 \\ &S_{2T} = \frac{1}{k}\tanh(kT) \end{aligned} \tag{54}$$

It follows that

$$\begin{aligned} &\lambda = \frac{\tilde{\mathcal{B}}^t Z_0 - c}{\tilde{\mathcal{B}}^t S_{2T} \tilde{\mathcal{B}}} \\ &C = -\tilde{\mathcal{B}}\lambda \end{aligned} \tag{55}$$

A.2 Multi-period liquidation

The standard approach to solve the tridiagonal system of linear equations is *progonka* a.k.a Thompson's algorithm. This algorithm perfectly fits our goal to find the optimal strategy with unknown terminal time. At first stage of *progonka* called decomposition and forward substitution we sequentially obtain the values of variables x_k under implicit assumption that the next term x_k+1 is equal to zero. Because the risks factors always work in direction of decreasing the absolute value of a position - the interval where the trajectory cross the line $x=0$ would be the last. The algorithm is presented by the following pseudo-code:

$$\begin{aligned} &\gamma_1 = b, \;\; \beta = b, \;\; n = 1, \;\; x_n = r_1/\beta, \;\; \dot{x} = -1 \\ &while(\dot{x} < 0 \;\&\; x_n > 0) \;\; \{ \\ &\quad n = n+1 \;\; Decomposition\ and\ forvard\ substitution \\ &\quad \gamma_n = 1/\beta \\ &\quad \beta = b - \gamma_n \\ &\quad x_n = (r - x_n)/\beta \\ &\quad \dot{x} = \frac{k}{sinh(kt_1)}((\alpha + x_n)\cdot cosh(kt_1) - (x_{n-1} + \alpha)) \\ &\} \\ &x_n = 0; \\ &for\ (j=2)\ to\ (n-1) \quad Backsubstitution \\ &\quad x_{n-j} = x_{n-j} - \gamma_{n-j+1}\cdot x_{n-j+1} \end{aligned} \tag{56}$$