

# RISK CONTROL OF MEAN-REVERSION TIME IN STATISTICAL ARBITRAGE

Joongyeub Yeo      George Papanicolaou

December 17, 2017

## Abstract

This paper deals with the risk associated with the mis-estimation of mean-reversion of residuals in statistical arbitrage. The main idea in statistical arbitrage is to exploit short-term deviations in returns from a long-term equilibrium across several assets. This kind of strategy heavily relies on the assumption of mean-reversion of idiosyncratic returns, reverting to a long-term mean after some time. But little is known regarding the assessment of this kind of risk. In this paper, we propose a simple scheme that controls the risk associated with estimating mean-reversions by using portfolio selections and screenings. Realizing that each residual has a different mean-reversion time, the ones that are fast mean-reverting are selected to form portfolios. Further control is imposed by allowing the trading activity only when the goodness-of-fit of the estimation for trading signals is sufficiently high. We design a dynamic asset allocation strategy with market and dollar neutrality, formulated as a constrained optimization problem, which is implemented numerically. The improved reliability and robustness of this strategy is demonstrated through back-testing with real data. It is observed that its performance is robust to a variety of market conditions. We further provide some answers to the puzzle of choosing the number of factors to use, the length of estimation windows, and the role of transaction costs, which are crucial issues with direct impact on the strategy.

**Keywords:** mean-reversion time, statistical arbitrage, portfolio selection, market neutrality, principal components, factor models, residuals

JOONGYEUB YEO (CORRESPONDING AUTHOR)

INSTITUTE FOR COMPUTATIONAL AND MATHEMATICAL ENGINEERING, STANFORD UNIVERSITY

*E-mail address:* `uriyeobi@gmail.com`

*Phone:* (650)556-4130

GEORGE PAPANICOLAOU

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY

*E-mail address:* `papanico@math.stanford.edu`

*Phone:* (650)723-2081

# 1 Introduction

With the rapid evolution statistical methods for analyzing large data sets and the establishment of automated markets, statistical arbitrage has become a common investment strategy with both hedge funds and investment banks. Although there is no consensus on what is statistical arbitrage, its main idea is a trading or investment strategy that exploits short-term deviations from a long-term equilibrium across the assets. Pairs trading is one of the earliest forms of statistical arbitrage, and it is widely used. This strategy consists of buying a certain asset which is below some equilibrium price and selling another correlated asset which is above it, expecting that the two will tend to equilibrium so that the trade generates profits. In other words, it is a bet on relative positions, rather than absolute positions, in a way that the resulting portfolio is unaffected by the equilibrium itself, and is relatively insensitive to market behavior. This kind of strategy heavily relies on the belief of *mean-reversion* of idiosyncratic returns or residuals, reverting to a long-term mean within some time that can be quantified. Although residuals of highly correlated assets generally converge to each other after diverging, there is no rule that asserts this has to happen. Thus, the essential risk of statistical arbitrage lies in the reliable quantification of the mean reversion characteristics of residuals.

There are very few studies in the literature that address the issue of the mean-reversion of residuals from the statistical arbitrage point of view. Thus, the objective of this paper is to address the question:

*Can we control the risk from the mean-reversion behavior of residuals in statistical arbitrage?*

The first and the main contribution of this paper is that we consider statistical methods for the risk-control of mean-reversion. Motivated by the observation that each residual has a different rate of mean-reversion, we carry out a statistical training to assess the quality of each residual relative to its rate of mean-reversion, from which only fast mean-reverting residuals are selected to form portfolios. Regular updates of the portfolio and further screening using the goodness-of-fit in the estimation of trading signals is also imposed to boost the reliability of the strategy.

Another contribution of our work is the transformation of the asset allocation problem in statistical arbitrage, with market- and dollar-neutrality conditions, to a constrained optimization problem that can be solved numerically. In this strategy, the threshold-based rule defines the sign (long or short) of the position of each asset and the neutrality conditions determine the investment amount.

Lastly, we demonstrate the performance of our strategy with real market data. The out-of-sample results show that, compared to other strategies, the proposed risk-control method works very well, giving persistently higher Sharpe ratios in all the different market regimes encountered in the last several years, before and after the 2008 crisis.

## 1.1 Basic aspects of statistical arbitrage

Implementing a statistical arbitrage strategy involves several steps. We briefly run through the necessary ones in the context of generalized pairs trading, where each pair consists of an individual asset and its corresponding common factors.

First of all, we describe how to construct residual returns. Consider a factor model

$$R = LF + U \quad (1)$$

where  $R$  is a  $N \times T$  matrix of the data (e.g., stock returns),  $F$  is a  $p \times T$  matrix of factors,  $L$  is an  $N \times p$  matrix of factor loadings,  $U$  is an  $N \times T$  matrix of the idiosyncratic components or residuals. Usually, only  $R$  is observable, so  $L$ ,  $F$ , and  $U$  must be estimated. By its definition, the residual is determined subject to the common factors ( $LF$ , in Eq. 1). To estimate factors we use principal components.<sup>1</sup> Then determining the common factors amounts to determining the number of factors. The estimation of the number of factors in high-dimensional models has been the subject of extensive research. For simplicity, in most of this paper rather than using a time-varying number of factors, we consider several different constant values.

Next, the factor loadings ( $L$ ) are estimated by multivariate regression of  $R$  on the subspace spanned by  $F$ . Finally, the residual ( $U$ ) is the remaining part, after subtracting the estimated common factors from the original returns:

$$\hat{U} = R - \hat{L}\hat{F}. \quad (2)$$

Here the hat ( $\hat{\cdot}$ ) notation indicates estimators, since only  $R$  is observable.

The next step is modeling the residual processes, and we adopt the Ornstein-Uhlenbeck (OU) process:

$$dX_t^i = \kappa^i(m^i - X_t^i)dt + \sigma^i dW_t^i \quad (3)$$

Here  $X_t^i = \sum_{l=1}^t U_{il}$  is an integrated residual return for the  $i$ -th asset.<sup>2</sup> The parameters  $(\kappa, \sigma, m)$ <sup>3</sup> in the OU modeling are estimated by least-squares, using an associated autoregressive process in discrete time.

The last step is trading. For each asset, a trading signal is generated as a normalized level of each residual using the estimated OU parameters. Then we open or close a position whenever its signal becomes active or inactive, respectively.

---

<sup>1</sup>This not just because it is easy to do. See section 2 for further discussion.

<sup>2</sup> $X_t^i$  has units of log price.

<sup>3</sup> $\kappa$  represent the rate of mean-reverting,  $\sigma$  is for volatility,  $m$  is for the long-term mean value of  $X$ .

## 1.2 Risk control for reliable residuals

Although the strategy described above is well designed and widely used, there are several issues that must be considered.

1. First and most critically, there is no guarantee that a residual reverts to some mean. Clearly, a slow mean reversion is unfavorable in statistical arbitrage, since it increases uncertainties in spread convergence, and hence decreases the chance of getting profits.
2. Second, the parameter estimation for generating trading signals may not always be satisfactory. Signals with large estimation errors cannot be trusted.
3. Third, the estimated mean-reversion times or the correlation structure of residuals are not constant over rolling time windows. They can and do change with time.

Consequently, to improve the reliability of the statistical arbitrage strategy, we propose the following controls.

1. Strategic reliability - Residuals that have long estimated mean-reversion time should be excluded from the strategy. We only consider residuals with relatively fast mean-reversion for portfolio selection.
2. Statistical reliability - Any trading signal with large estimation error is screened out.
3. Temporal reliability - Portfolio choices based on the estimated mean-reversion time must be updated regularly.

For the first control, we introduce a *reliability score* for each residual by the average of rates of mean-reversion ( $\sim \hat{\kappa}$ ) estimated from training periods. Then we choose only a certain portion of the highest scoring stocks, expecting that the residuals of those stocks will show good mean-reversion properties. The second control ensures that the estimation for the trading signals is as reliable as we want. Indeed, we reject a trading signal if the  $R^2$ -value from the corresponding OU-estimation is lower than a cutoff value.<sup>4</sup> The third part is to avoid relying on old information. Refreshing portfolio selection enables us to keep holding reliable residuals. These operations are all shown to significantly diminish the relevant risk and augment the stability of the strategy.

## 1.3 Optimization for asset allocations

As discussed, a trading signal determines whether to buy or sell the asset. However, how much one needs to allocate to each asset is still a non-trivial problem, since there are several constraints. In particular, market neutrality is crucial in our strategy, since it allows the portfolio to be relatively uncorrelated from market movements. Dollar neutrality and leverage constraints are also necessary for a fair assessment of strategies. In

---

<sup>4</sup>Other goodness-of-fit measures can be used, but in this paper we use  $R^2$ -value.

short, for each time  $t$ , the required conditions can be stated as follows,

$$\text{Market-neutrality : } \sum_i L_{ik} q_i = 0, \text{ for each } k=1, \dots, p \quad (4)$$

$$\text{Dollar-neutrality : } \sum_i q_i = 0 \quad (5)$$

$$\text{Total Leverage : } \sum_i \|q_i\| = I \quad (6)$$

$$\text{Long/Short : } \text{sgn}(q_i) \text{ is given by trading signals.} \quad (7)$$

Here  $L$  is the factor loading matrix in Eq.1,  $q_i$  is the investment amount on asset  $i$ ,  $p$  is the number of factors, and  $I$  is the total leverage level. Note that the first condition is derived by multiplying  $q$  on both sides of Eq.1 and letting the common factor parts be zero.

These conditions can be consolidated to formulate an optimization problem. We do not write the full expression here, but it turns it has the form:

$$\min_q \|Lq\|_1 \quad (8)$$

$$\text{s.t.} \quad Aq \preceq 0 \quad (9)$$

$$A_{eq}q = b_{eq} \quad (10)$$

$$c(q) = d \quad (11)$$

where  $A$  and  $A_{eq}$  are square matrices,  $b_{eq}$  and  $q$  are vectors,  $c(q)$  is a nonlinear function, and  $d$  is a scalar<sup>5</sup>. By solving this problem, asset allocation can be more efficient in the sense of robustness to market variations.

## 1.4 Back-tests

The performance of our method is evaluated by analyzing the cumulative profit and loss (PnL) with real data. As for market data, we use 3780 records (Jan 2, 2000 - Dec 31, 2014) of daily prices for 378<sup>6</sup> stocks of the S&P500. We use a portfolio size of 25, 50, 75, and 100 stocks, out of a total of 378 stocks. The performance is evaluated within different market regimes. We consider five regimes, each of which has two-year (504 business days) period: pre-crisis (2005-2006), in-crisis (2007-2008), post-crisis (2009-2010), afterward (2011-2012), and recent years (2013-2014). The results are compared when using different portfolios as well as with other parameter choices. As comparison groups, we first consider a portfolio where stocks are randomly chosen. In addition, portfolios where stocks are selected by capitalization, one with the highest and the other with the lowest, are also investigated.

---

<sup>5</sup>As will be discussed later, the dimensions of these quantities can change depending on the previous state of the portfolio.

<sup>6</sup>There are stocks coming in and out of the pool of S&P500. We have only picked stocks which have persisted during the 15-year period.

Figure 1 provides a snapshot of the back-testing results. We see that our portfolio selection using faster mean-reversions performs better than any other comparison portfolios: the Sharpe ratios of the controlled portfolios are much higher. By calculating the length of days of active positions as a proxy of *realized* mean-reversion time, we check that our training for the mean-reversion speeds is effective in the out-of-sample tests. Furthermore, the selective trading via the  $R^2$ -value screening boosts the overall results.<sup>7</sup> These results show that our risk controls lead to enhanced reliability in statistical arbitrage.

We also see that the investment adjustment derived from the proposed optimization problem plays an important role in enhancing the robustness of the strategy. Regardless of regimes, the Sharpe ratios are more stable than those from uniform asset allocation, which shows the importance of the optimization problem regarding market-neutrality. Lastly, we also investigate the effect of parameters such as number of factors, window lengths, and transaction costs. From the results we find that the performance with shorter windows is more sensitive to the impact of higher transaction costs, as might be expected. Furthermore, we see that in volatile markets only a small number of factors is sufficient for our strategy to work well. The issues of portfolio size and the decreasing performance in recent years are discussed as well.

## 1.5 Relevant works

Pairs trading has been studied extensively from various perspectives. A seminal work is [1], where the authors present a very simple relative value trading rule with U.S. data from 1963 to 2002. They use the Euclidean squared distance of price time series to identify pairs, and show that a simple rule to open/close a trade generates statistically and economically significant excess returns. This work has been extended by [2] and [3]. They refine the formation of portfolios by allowing securities within the 48 Fama-French industries and by measuring the number of zero crossing. Nevertheless, there are several limitations, including the fact that none of these studies take into account trading costs.

There are not many studies devoted to mean reversion in residuals in the context of statistical arbitrage.<sup>8</sup> In [9] a Gaussian linear state-space processes is suggested for modeling mean-reverting spreads arising in pairs trading. Their point of view is that the observed process must be seen as a noisy realization of an underlying hidden process, so the comparison between the two can lead to a more profitable strategy. They suggest to use the EM (expectation-maximization) algorithm for estimating model parameters. Although their model is fully tractable, its applicability in practice is debatable, due to the fact that the model restricts the long run relationship between the two stocks to one of return parity.<sup>9</sup> In [11] a framework for forecasting is developed using a co-integration method. Stocks with similar common trends are preselected, are assumed to be non-

<sup>7</sup>This is verified with various choices on the number of extracted factors and estimation windows.

<sup>8</sup>However, the notion of mean reversion of stock *prices* has received a considerable amount of attention in the financial literature ([4]; [5]; [6]; [7]; [8]).

<sup>9</sup>In the long run, the two stocks chosen must provide the same return so that any departure from it will be expected to be corrected in the future ([10]).

stationary within the framework of a common trend model as in [12] and in Arbitrage Pricing Theory of [13]. In Vidyamurthy a method is also developed for finding the optimal triggering level for each pair, by counting the number of times that each trigger level is exceeded. However, Vidyamurthy’s arguments are informal, depend on heuristics, and lack explanations for decisions that affect other aspects of the strategy and thus make the approach of questionable value for practitioners.

The Ornstein-Uhlenback process has played a significant role in modeling residuals. Closed form expressions are obtained in [14] for the mean and variance of the trade duration and the strategy return, using OU modeling. The effect of transaction costs is also considered. In [15] stochastic control theory with OU-modeling is used, and an optimal dynamic strategy for mean-reverting arbitrage opportunities is presented. But transaction costs are still missing in the paper, and so their daily rebalancing is not expected to outperform the threshold-based trading rule.

Recently, [16] carried out an extensive study of statistical arbitrage in the U.S. equities market. The authors use principal component analysis and sector ETFs to find factors, and analyze the performance of statistical arbitrage strategies using residual modeling. Our study is motivated by their work, while our main focuses is on the risks associated with mean-reversion times, portfolio selections, and the reliability of estimations. We also enforce market- and dollar- neutrality by solving a constrained optimization problem, which is also different from what they do. Our approach is especially effective when there is a restriction on the number of stocks in a portfolio. We also study in detail the effect of various parameter choices, such as the number of factors, estimation windows, capitalizations, regimes, or transaction costs.

## 1.6 Outline

The remainder of the paper is organized as follows. In section 2, we first recall the factor model framework and explain how the residuals are generated. Then the estimation of mean-reversion times is discussed. In addition, we introduce a risk control method to search for reliable residuals. Section 3 explains the trading rule and the required balance conditions for the strategy. We propose an optimization problem for the asset allocation. The numerical results and back-testings are presented in section 4. We compare results with different parameter choices and with various other portfolios. We first investigate the effects of our control methods and the proposed optimization problem for asset allocations. We explore further other issues, such as capitalization, market regimes, number of factors, training window lengths, transaction costs, and portfolio size. An issue in recent years’ market and the results from time-varying number of factors are also addressed. Finally, we conclude in section 5.

## 2 Risk-control of mean-reversion times of residual processes

In this section we explain how residuals are constructed, calibrated, and controlled.



## 2.1 Construction of residuals

Most trading strategies in statistical arbitrage only concern the residual returns of assets. Thus, the way of determining residuals directly affects its performance. By definition of residuals in factor models, the problem of determining residuals is equivalent to that of determining common factors.

Let us consider a set of  $N$  securities (e.g., stocks) each represented by a time series of returns and let  $T$  be the number of time series observations. For  $i = 1, \dots, N$  and  $t = 1, \dots, T$ , we write a factor model as

$$\underset{n \times 1}{r_t} = \underset{N \times p}{L} \underset{p \times 1}{f_t} + \underset{n \times 1}{u_t} \quad (12)$$

where  $r_t$  is a time series of original returns,  $f_t$  represents factors,  $L$  represents factor loadings, and  $u_t$  represents idiosyncratic returns. Note that the rationale of this model is to decompose the original signal into systemic components (factors) and idiosyncratic components (residuals). Collecting the above vectors, we can rewrite the above factor model in a matrix form

$$\underset{N \times T}{R} = \underset{N \times p}{L} \underset{p \times T}{F} + \underset{N \times T}{U} \quad (13)$$

In this paper, we adopt statistical factor models in which factors are unobservable and so extracted from asset returns. That is, one needs to determine factors, their loadings, the number of factors, and residuals, based on the information taken from market returns. We follow [16] for constructing factors. The number of factors is also an important quantity to be specified. One simple and common way is to choose  $p$  factors that explain a certain level of variance. In this study, we mainly consider constant number of factors.<sup>10</sup>

From the data matrix  $R$ , a correlation matrix can be obtained according to the formula:

$$C = \frac{1}{T} M M^T \quad (14)$$

where  $M$  is  $N \times T$  matrix of the normalized returns derived from  $R$ .<sup>11</sup> Each element of  $C$  is obviously the Pearson correlation coefficient  $C_{i,j}$  between a pair of stocks  $i$  and  $j$ . The correlation matrix can be diagonalized by solving the eigenvalue problem

$$C v_i = \lambda_i v_i, \quad i = 1, \dots, N \quad (15)$$

From investment theory perspective, each eigenvector  $v_i$  can be considered as a realization of an  $N$ -security portfolio with weights equal to the eigenvector components  $v_i^k$ , where  $k = 1, \dots, N$ . However, in order to

---

<sup>10</sup>In section 4.9, we also consider the dynamic number of factors by variance explanation ratios. The number of factors should also take into account the correlation structure as well as variance. We developed a new method to deal with this issue in another paper.

<sup>11</sup> $M_{it} := (R_{it} - \bar{R}_i)/\sigma_i$ , where  $\bar{R}_i$  and  $\sigma_i$  are the mean and standard deviations, respectively, of  $R_{it}$ .

estimate each factor, we use the following *eigensignals*:

$$\hat{F}_m = \langle q_m, r_t \rangle, \quad m = 1, \dots, p \quad (16)$$

where  $i$ -th element of a vector  $q_m$  is defined by  $q_m^i := v_m^i / \sigma_i$  with  $\sigma_i$  the standard derivation of stock  $i$ , and  $\langle \cdot \rangle$  is the inner product, and  $p$  is the number of factors. The division by  $\sigma_i$  is to take into account the fact that higher capitalized firms are less volatile as discussed in [16]. Once  $\hat{F}$  is estimated, we use multivariate least squares regressions of  $R$  on  $\hat{F}$  to estimate the factor loading coefficients,  $\hat{L}$ . Finally, the residuals are obtained by the difference between the original returns and common factors.

$$\hat{U} = R - \hat{L}\hat{F}. \quad (17)$$

Note that this calculation is defined within a given time window  $T$ .<sup>12</sup>

## 2.2 Calibration of residuals

For the estimation of mean-reversion time of residual processes, we employ the Ornstein-Uhlenbeck (OU) process.<sup>13</sup>

Rewriting the factor model, we have

$$R_{it} = \sum_{j=1}^p L_{ij} F_{jt} + U_{it} \quad (18)$$

It is natural to assume that the integrated residual process

$$X_{it} = \sum_{l=1}^t U_{il} \quad (19)$$

shows mean-reverting behaviors. The OU-modeling for stock  $i$  is given as

$$dX_{it} = \kappa_i(m_i - X_{it})dt + \sigma_i dW_{it}, \quad \kappa_i > 0. \quad (20)$$

Note that the OU model is a continuous stochastic differential equation, but we only have discrete time observations<sup>14</sup>. One way to deal with this is to discretize the OU process to the autoregressive process of order 1, and use the standard estimation method, the least-squares (LS), to estimate parameters. That is, we first

---

<sup>12</sup>Two kinds of time windows are mentioned in this paper. One is for estimating the sample correlation matrix and PCA, and the other is for parameter estimation of residual modeling. For simplicity and consistency, we fix the same length for both.

<sup>13</sup>There are many possible ways to model the residuals. However, the OU modeling is the most frequently used for mean-reverting processes and is convenient for parameter estimation.

<sup>14</sup>There have been numerous studies made on estimating continuous time models based on discrete time observations. We refer to the following literature: [17]; [18]; [19]; [20].

estimate the following model

$$X_{k+1} = a + bX_k + \epsilon_k, \quad (21)$$

where  $\epsilon_k \sim N(0, \nu^2)$ , using LS.<sup>15</sup> Then it is straightforward to relate the estimated parameter sets  $\{\hat{a}, \hat{b}, \hat{\nu}\}$  to  $\{\hat{\kappa}, \hat{m}, \hat{\sigma}\}$ . The details of this estimation method are described in an appendix of [16].<sup>16</sup>

## 2.3 Risk-control methods

### 2.3.1 Ranks by mean-reverting speed

The mean-reversion time of residuals is hard to analyze and control, since it reflects idiosyncratic features of individual asset, which may not be easily interpreted with economic indicators. For example, the mean-reversion dynamics of residuals does not have to be similar to that of the original returns. Obviously, there is no fundamental theory for this. However, our approach starts from the premise that some stocks have a more distinguishable correlation structure than others. Then their residuals, the remainders after removing factors, are more well-decoupled from the market and hence are more reliable for market-neutral strategies. In other words, these residuals are less trending or faster mean-reverting. Since there is no *structural model* for this mechanism, we develop empirical schemes using market data.

Note that in OU modeling, the parameter  $\kappa$  represents the mean-reverting speed<sup>17</sup>. In order to find “good” residuals, we first filter out slow mean-reversions. This is because if a residual is not reverting for a long time after opening a position on it, it will be hard to find an optimal time to close the position. Hence, we accept residuals which have relatively faster mean reversion speed.

To quantify the mean-reversion speed of each residual, we take a simple average of mean-reverting speed ( $\hat{\kappa}_{it}$ ) that are estimated for each stock  $i$  and on  $n$ -th day of the training period. A quality score is defined as

$$QS_i = \frac{1}{T_{train}} \sum_{n=1}^{T_{train}} \hat{\kappa}_{in}, \quad (22)$$

$\kappa_{in}$  is the estimated mean-reversion speed of asset  $i$  at time  $n$  within  $T_{est}$  and  $T_{train}$  and  $T_{est}$  are the length of training and estimation window, respectively<sup>18</sup>. This measure is motivated from the in-sample tests where

<sup>15</sup>The maximum likelihood (ML) is also frequently used. It can be shown that the LS estimator is equivalent to the ML estimator, and we focus LS estimator in the rest of the paper.

<sup>16</sup>One problem with continuous OU model is estimation bias for the mean-reversion estimator. Standard estimation methods, such as least squares, maximum likelihood or generalized method of moments, are known to produce biased estimators for the mean reversion parameter ([21]). Researchers have developed several methods to reduce this bias. For example, [22] adopts jackknife method to resolve this issue.

<sup>17</sup>The mean-reversion time can be defined by its inverse, scaled by the interval of discrete observations:  
 $\tau \equiv \frac{1}{\kappa \Delta t}.$

<sup>18</sup>The estimated mean-reversion parameter  $\kappa$  generally depends on the length estimation window. Thus, the mean-reversion speed must be normalized by the estimation window. However, since our selection step is only

we observe that  $\kappa$ 's are widely distributed. Since a higher score of  $QS_i$  implies a higher (on average) mean-reversion speed of stock  $i$ 's residual, those stocks are selected at the end of the training. The number of selected stocks in a portfolio can vary. In this paper, we consider 25, 50, 75, and 100 out of total 378 stocks.<sup>19</sup>

### 2.3.2 Selective trading via $R^2$ -value screening

As we discussed, the estimated OU parameters are used for generating trading signals. Poor estimation cannot provide reliable signals. In order to improve statistical reliability, we accept estimation results only with high enough goodness-of-fit. As a measure of goodness-of-fit, we adopt  $R^2$ -value<sup>20</sup> and set a threshold for this:

$$R^2 \text{ value} > \eta \quad (23)$$

where  $R^2$ -value is calculated for the OU parameters estimation and  $\eta \in (0, 1)$  is the cutoff level. This selective trading reduces the frequency of inefficient trading as well as their associated unnecessary transaction costs. Empirical distributions of estimated mean-reversion time and  $R^2$ -value are shown in Figure 12.

### 2.3.3 Updating portfolio selection

The mean-reversion dynamics of residuals are not stationary, but change in time. The validity of residuals' ranks based on the training within a time window can deteriorate with time. Thus, the portfolio selection must be updated regularly. The updating interval is set to be same as the length of estimation window, for consistency of market information. By doing this, the current portfolio maintains up-to-date information of mean-reversion times.

## 3 Robust trading algorithm and optimization

In this section, we give an overview of the trading strategy we implement. First, a threshold-based rule is explained. We then focus on developing an investment strategy that incorporates several constraints that are needed to make it robust.

Note that the strategy involves three main steps.

1. Residual generation: do PCA for historical data from  $t-T$  to  $t$ , where  $t$  is the current time, and generate residuals from factor models.

---

conducted in a fixed window, it is actually not necessarily at least here.

<sup>19</sup>If the number selected is too small, then the solution of optimization problem for market-neutral condition becomes inaccurate, since finding a feasible combination of dollar quantities that cancelled out can be hard.

<sup>20</sup>The  $R^2$  value is defined as  $R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}}$ , where  $SS_{res}$  and  $SS_{tot}$  are the residual and total sum of squares, respectively.

2. Portfolio selection: estimate parameters with OU-modeling of residuals and select stocks using confidence criteria.
3. Trading: dynamically adjust the portfolio by opening and closing the long/short positions during the trading horizon.<sup>21</sup> The investment amount for each asset is determined by solving a constrained optimization problem.

### 3.1 Trading signals

From the estimated OU parameters, we generate dimensionless trading signals as

$$s_{it} = \frac{X_{it} - m_i}{\tilde{\sigma}_i} \quad (24)$$

where  $m_i$  and  $\tilde{\sigma}_i$  are mean and equilibrium standard deviation of residuals,<sup>22</sup> and  $X_{it}$  is the level of integrated residual return of asset  $i$  at time  $t$ . This signal is a normalized version of residual state, which suggests a simple trading strategy as follows. If the signal is strongly positive, the asset is assumed to be over-valued. Then we open a short position, expecting that the price will go down. Once it comes back to a lower level, we close the short position. Similarly, we can open a long position if the signal is strongly negative, and so on. The four thresholds we use are given as follows.

$$\text{open short position}(s_{it} > 1.25) \rightarrow \text{close short position} (s_{it} < 0.5) \quad (25)$$

$$\text{open long position}(s_{it} < -1.25) \rightarrow \text{close long position} (s_{it} > -0.5) \quad (26)$$

The scheme for the threshold based rule is illustrated in Figure 2.

### 3.2 Three constraints

For the practical implementation of the strategy and the proper assessment of its performance, we impose the following three conditions.

#### 3.2.1 Market-neutrality

A market-neutral strategy generates returns that are independent of the market environment. Recall we have the factor model for  $R_{it}$ ,

$$R_{it} = \sum_{k=1}^p L_{ik} F_{kt} + U_{it}. \quad (27)$$

---

<sup>21</sup>For selective trading, trades are rejected if the estimation error is too high.

<sup>22</sup> $\tilde{\sigma}_i = \sigma_i / \sqrt{\kappa_i}$ , where  $\sigma_i$  is the standard deviation of residual returns.

Let  $q_{it}$  be the capital amount invested in asset  $i$  at time  $t$ . Then we can write the gain at each time by multiplying  $q_{it}$  to  $R_{it}$ .

$$\sum_{i=1}^N q_{it} R_{it} = \sum_{i=1}^N q_{it} \left[ \sum_{k=1}^p L_{ik} F_{kt} + U_{it} \right] \quad (28)$$

$$= \sum_{k=1}^p \left[ \left\{ \sum_{i=1}^N L_{ik} q_{it} \right\}_{kt} F_{kt} \right] + \sum_{i=1}^N q_{it} U_{it} \quad (29)$$

where we switched the summation in the second equality. The gain can be decomposed into the factor gain and the residual gain. The condition for market-neutrality, or hedging the so-called *beta risk*, implies there is no gain from the market factors. Eliminating the first term on the RHS, we have the following expression for market-neutrality

$$\sum_{i=1}^N L_{ik} q_{it} = 0, \text{ for each } k = 1, \dots, p \text{ and for each } t \quad (30)$$

### 3.2.2 Dollar-neutrality

We also want our strategy to be dollar-neutral, which just means symmetry between long and short holdings,

$$\sum_{i=1}^N q_{it} = 0, \text{ for each } t \quad (31)$$

With this condition, there is no net amount of money investment. This also ensures that we can have a fair comparison or assessment of the strategy.

### 3.2.3 Leverage: total amount of investment

When it comes to comparing different strategies, the total absolute amount matters. In order to normalize this effect, we need to restrict the total investment.

$$\sum_{i=1}^N |q_{it}| = I, \text{ for each } t \quad (32)$$

The leverage constant,  $I$ , is set to achieve a desired level of volatility of returns.

## 3.3 Optimization for asset allocations

The rule displayed in Figure 2 determines only whether to buy or sell an asset. The main objective of this section is to determine *how much*. First of all, it is worthwhile to note that not all assets' allocations need to be changed each time. For example, if the trading signal of an asset does not indicate a change of its position from time  $t$  to  $t + 1$ , then the amount already invested in this asset is preserved. This means that the adjustment should be applied only to stocks that are to take new positions. This actually makes sense and it

avoids unnecessary transaction costs.

Let us start out by defining the following sets of assets:

$$\mathcal{C}_t : \text{assets whose positions are changed from } t-1 \text{ to } t \quad (33)$$

$$\mathcal{U}_t : \text{assets whose positions are unchanged from } t-1 \text{ to } t \quad (34)$$

$$\mathcal{L}_t : \text{assets that are long positions at } t \quad (35)$$

$$\mathcal{S}_t : \text{assets that are short positions at } t \quad (36)$$

$$\mathcal{O}_t : \text{assets that are zero positions at } t \quad (37)$$

Note that these sets are completely determined by the trading signals and the threshold-based rule. Then we can rewrite the three conditions as

$$\text{Positions:} \quad \text{sgn}(Q_{it}) = \begin{cases} +1, & i \in \mathcal{L}_t \\ 0, & i \in \mathcal{O}_t \\ -1, & i \in \mathcal{S}_t \end{cases} \quad (38)$$

$$\text{Market-neutrality:} \quad \sum_{i \in \mathcal{C}_t} L_{ik} q_{it} + \sum_{i \in \mathcal{U}_t} L_{ik} q_{it} = 0, \text{ for each } k \text{ and } t \quad (39)$$

$$\text{Dollar-neutrality:} \quad \sum_{i \in \mathcal{C}_t} q_{it} + \sum_{i \in \mathcal{U}_t} q_{it} = 0, \text{ at each } t \quad (40)$$

$$\text{Total leverage:} \quad \sum_{i \in \mathcal{C}_t} |q_{it}| + \sum_{i \in \mathcal{U}_t} |q_{it}| = I, \text{ at each } t \quad (41)$$

Note that at time  $t$ ,  $\{q_{it}|i \in \mathcal{U}_t\} = \{q_{i,t-1}|i \in \mathcal{U}_t\}$ , since the position of the asset  $i$  is unchanged from  $t-1$  to  $t$ , by definition of  $\mathcal{U}_t$ . Thus, the summations including  $\mathcal{U}_t$  can be treated as constants at time  $t$ .

In order to find  $q_{it}$ , we set up an optimization. Finding a suitable objective function is a crucial issue. In this paper, we want to avoid introducing any other quantity such as expected returns or covariance matrices, since they are hard to estimate or they can lead to other intricate issues. Instead, we take one of the conditions and modify it to the objective function. In particular, we consider the market-neutrality condition:  $\sum_{i=1}^N L_{ik} q_{it} = 0$  for each  $k$ . Motivated from this, we set the objective function as

$$\left| \sum_{i=1}^N L_{i1} q_{it} \right| + \cdots + \left| \sum_{i=1}^N L_{ip} q_{it} \right| \quad (42)$$

which is to be minimized. Other conditions will be the constraints in the optimization problem.<sup>23</sup> Therefore,

---

<sup>23</sup>Other conditions may be used for the objective function. Our choice is made solely for better numerical stability in solving the problem.

to find the asset allocations at each time  $t$ , we solve the following problem.

$$\min_{q_{it}, i \in \mathcal{C}_t} \sum_{k=1}^p \left| \sum_{i \in \mathcal{C}_t} L_{ik} q_{it} \right| \quad (43)$$

$$\text{s.t.} \quad \text{sgn}(q_{it}) = \begin{cases} +1, & i \in \mathcal{L}_t \\ 0, & i \in \mathcal{O}_t \\ -1, & i \in \mathcal{S}_t \end{cases} \quad (44)$$

$$\sum_{i \in \mathcal{C}_t} q_{it} = - \sum_{i \in \mathcal{U}_t} q_{it} \quad (45)$$

$$\sum_{i \in \mathcal{C}_t} |q_{it}| = I - \sum_{i \in \mathcal{U}_t} |q_{it}| \quad (46)$$

Again, the first constraint is for the sign of each investment that is determined by trading signals. The second constraint is for a dollar-neutrality, and the third is from the leverage condition. The summations over  $\mathcal{U}_t$  are written on the right hand sides, to indicate that they are known quantities at time  $t$  and regarded as constants.

We can rewrite the problem in a simpler vector-matrix form. Let us employ notation for two vectors,  $q_t$  and  $\tilde{q}_t$  such that

$$q_t = \{q_{it}, i \in \mathcal{C}_t\} \quad (47)$$

$$\tilde{q}_t = \{q_{it}, i \in \mathcal{U}_t\} \quad (48)$$

That is,  $q_t$  and  $\tilde{q}_t$  represent the dollar investments for stocks that have changed and not changed their positions, respectively, at time  $t$ . Our goal is to find  $q$ , given the fact that  $\tilde{q}$  is determined at  $t$ . Noting that the summation of absolute values can be simply rewritten as 1-norm, the vector-matrix expression is given by

$$\min_{q_t} \|Lq_t\|_1 \quad (49)$$

$$\text{s.t.} \quad Aq_t \preceq 0 \quad (50)$$

$$A_{eq}q_t = b_{eq} \quad (51)$$

$$c(q_t) = d \quad (52)$$



where

$$\underline{A}_{N_c \times N_c} : A_{ij} = \begin{cases} -\delta_{ij}, & i \in \mathcal{L}_t \\ +\delta_{ij}, & i \in \mathcal{S}_t \\ 0, & \text{otherwise} \end{cases}, \quad (53)$$

$$\underline{A}_{(N_c+1) \times N_c}^{eq} = \begin{bmatrix} \tilde{A} \\ 1^T \end{bmatrix}, \quad \text{with } \tilde{A}_{N_c \times N_c} : \{\tilde{A}_{ij} = \delta_{ij}, i \in \mathcal{O}_t\} \quad (54)$$

$$\underline{b}_{(N_c+1) \times 1}^{eq} = \begin{bmatrix} 0 \\ 1^T \tilde{q}_t \end{bmatrix} \quad (55)$$

$$c(q_t) = \|q_t\|_1 \quad (56)$$

$$d = I - \|\tilde{q}_t\|_1 \quad (57)$$

Here  $N_c$  is the number of changed positions at each time, defined as  $N_c = |\mathcal{C}_t|$ .<sup>24</sup> This constrained optimization can be solved numerically. There is a numerical issue regarding the portfolio size. If we consider larger portfolios, the optimization has more degree of freedom to control, so it is likely that the constraints can be satisfied more tightly. However, at the same time, solving a higher dimensional optimization problem can accumulate more numerical error. It turns out from the back-test results that larger portfolios result mostly in better performance than smaller ones.<sup>25</sup>

## 4 Empirical results

The back-testing results are presented in this section. We start out by describing the testing setups.

### 4.1 Back-testing setups

#### 4.1.1 Schemes

Figure 3 provides a schematic diagram on how the data is used for back-testings. During the “portfolio selection” period, we use historical data to estimate the mean-reversion speed of each asset’s residual. The estimations are collected with a moving window, and they are sorted by the average mean-reversion speed.<sup>26</sup> At the end of training, the high-ranked stocks are chosen to form the trading portfolio. In the “trading” period, the trading signals are generated. Then we dynamically adjust the asset allocations based on the solution of the optimization problem we discussed in section 3.3. Note that the OU-estimations are carried out in both

---

<sup>24</sup>We found that the matrices and vectors used in this optimization problem are sparse, since the proportion of active signals are usually small.

<sup>25</sup>The effect of the portfolio size is discussed in section 4.7.

<sup>26</sup>or the inverse of the estimated mean-reversion times.

periods: One is for sorting estimated mean-reversion speeds and the other is for generating signals.

#### 4.1.2 Parameters

Trading results cannot of course be independent of the choice of parameters and of the testing environments. Therefore, we consider the following parameter ranges for our tests.<sup>27</sup>

1. Estimation window: 30, 60, 90, 120 days
2. Number of factors: 5, 10, 15, 20 factors are removed to obtain residuals.
3. Trading periods: 5 regimes - 2005-2006 / 2007-2008 / 2009-2010 / 2011-2012 / 2013-2014
4. Portfolio selection: risk-controlled, high-capitalization, low-capitalization, random

The window lengths are chosen for reliable estimations. For example, 5 days are insufficient to derive mean-reversion properties from daily data. On the other hand, a window that is too long can cause over-fitting or overemphasis of old information. We choose the range from 30 to 120 days which is also consistent with the actual reporting cycles of companies. As for the number of factors, our intention is to find the number of principal components that effectively remove strong cross-correlations between assets. From the returns data of 378 stocks<sup>28</sup>, we found that taking out more than 30 principal components makes residuals too noisy, which may cause instability in estimations. In order to have meaningful mean-reverting residuals, we extract 1 to 20 factors. Lastly, the performance of the trading strategy is evaluated and summarized for each two-year period. Although the overall performance for the whole period (ten years in this paper) may be important for long-term profits, persistent performance with no bad periods is preferred. Thus, five different two year periods are considered: pre-crisis (2005-2006), in-crisis (2007-2008), post-crisis (2009-2010), afterward (2011-2012), and recent years (2013-2014).

#### 4.1.3 Performance measure

The performance from trading can be measured by the cumulative profit-and-loss (PnL). We use the following formula.<sup>29</sup>

$$E_{t+\Delta t} = E_t + E_t r \Delta t + \sum_{i=1}^N q_{it} R_{it} \Delta t - \sum_{i=1}^N q_{it} r \Delta t - \sum_{i=1}^N |q_{i(t+\Delta t)} - q_{it}| \epsilon \quad (58)$$

where  $E_t$  is the value of the portfolio at time  $t$ ,  $r$  is the reference rate of cash, and  $\epsilon$  is for frictional costs. Note that the fourth term on the right hand side must be zero if the dollar neutral condition is completely satisfied. However, there are numerical tolerance imposed for the feasibility, we keep that term here.

<sup>27</sup>Other parameters are also considered later.

<sup>28</sup>The number of factors can vary with the asset universe used.

<sup>29</sup>We use the same formula as in [16].

For each period, the annualized Sharpe ratio for the PnL is calculated as the standardized excessive returns:

$$SR_{ann} = \frac{\mu_{ann} - r}{\sigma_{ann}} \quad (59)$$

where  $\mu_{ann}$  and  $\sigma_{ann}$  are the average and standard deviation of annualized return of PnL, respectively, and the risk-free rate  $r$  is a constant.<sup>30</sup> Annualized total returns and volatility of PnL returns are examined as well.

#### 4.1.4 Market data

Our analysis is based on stocks from S&P500. We extracted  $T = 3780$  records (Jan 2, 2000 - Dec 31, 2014) of daily prices for  $N = 378$  stocks, which have persisted during this period. The data of daily stock prices and market capitalization are obtained from Bloomberg Database Services: `PX LAST` recorded the closing price on the last trading day and `CUR MKT CAP` is used to record the prevailing market capitalization at each day.

## 4.2 Sample PnL plots

Before presenting the back-test results, we first present typical PnL plots that are generated from controlled and random portfolios with different transaction costs. Our trading algorithm works quite well regardless of the prevailing market regime.

## 4.3 Mean-reversion control

As illustrated in Section 2.3, stocks are chosen based on the quality scores we calculate before each period so as to form the trading portfolio. We also compare performance with other portfolios, including a randomly chosen portfolio, and high- and low- capitalized firms portfolios. The left plot in Figure 5 shows the performance of the four different portfolios. The portfolio size here is  $N = 75$  and the transaction cost is 5bp ( $\epsilon = 0.0005$  in equation (58)), and no  $R^2$ -screening is applied. Our portfolio selection using fast mean-reverting residuals outperforms other portfolios. In particular, we found no significant drawdown in all five periods when using the controlled portfolio.<sup>31</sup> This shows that our approach is robust to market changes, yielding consistent results during both bearish and bullish markets.

In contrast, the performance with a randomly chosen portfolio is different. The Sharpe ratios of the random portfolio are lower and more unstable. It is also clear that the mean reversion strategy with the risk-controlled portfolio is far superior to low- and high- cap portfolios as well. Moreover, the controlled portfolio gives high-returns and low-volatility, whereas for other portfolios, the high returns are accompanied by high

---

<sup>30</sup>We use  $r = 2\%$  in this paper.

<sup>31</sup>For example, with a window of 60 days and 5 factors, the Sharpe ratios are close to or higher than 2 over all periods, and the average annual return of the strategy is 20%, which are convincing results. See Table 4 for detailed results.

volatilities.

To further assess the impact of our approach, we also test with an *anti*-controlled portfolio, which is formed from the stocks with long mean-reversion times. We present the performance from this portfolio in Table 10 in the Appendix, and as expected, the results are the worst among the all other portfolios considered. Although there are few parameter choices that yield a good performance, the overall fluctuations of PnL are significantly larger than those from other portfolios. We also confirm this with experiments using other portfolio sizes.

To confirm and illustrate the effect of mean-reversion control in the out-of-sample tests, we also investigate the realized mean-reversion time of the residuals. By this, we mean the realized buy-sell time intervals, or the duration between opening and closing of a position:

$$\text{Realized mean-reversion time} := \text{length of days of active positions} \quad (60)$$

This quantity can be easily calculated by tracking the positions while trading. In Figure 6, we summarize the estimated and realized mean-reversion times. For estimated mean-reversion times, we take the inverse of mean-reverting speed scaled by the data observation interval:  $\tau_i = \frac{1}{\kappa_i \Delta t}$  for asset  $i$ , and then we average. As shown in the left plot, estimated the mean-reversion times for the controlled portfolio are about 7 days, while those for other portfolios are 8 or more days. The difference is also evident for realized values. On average, the realized mean-reversion times of the controlled portfolio are 0.5 to 1 day shorter than those of other portfolios. Since there is no strict definition of mean-reversion time, the observed difference in scale is expected. The main point here is that our control of the estimation process works well and the realized mean-reversion times for the controlled portfolio are clearly shorter than those from other portfolios. It also supports the observed effectiveness of the training based on estimated mean-reversion times.

#### 4.4 Selective trading via $R^2$ -value screening

The reliability of a trading signal relies on the estimation quality of OU-parameters. Rather than keeping and using all signals, we reject some of them if the goodness of fit of the corresponding OU estimations is significantly lower than an acceptable level. The least squares estimation uses two consecutive integrated residuals,  $X_{k+1}$  and  $X_k$ , where  $X_k$  is defined as the sum of residual returns from time 1 to  $k$ . With longer windows,  $X_{k+1}$  and  $X_k$  are close, and the  $R^2$ -value from the OU-estimations tends to be larger. We set different thresholds for different window lengths so that the total number of trades is decreased by 50% from this screening.

The effect of selective trading via  $R^2$ -value screening on Sharpe ratios is clear as seen from the right plot of Figure 5. We find that in most periods and parameter ranges, the  $R^2$ -screening boosts up performance. Comparing results from before and after applying the  $R^2$ -screening, we can see that the Sharpe ratios (+0.22, +0.47, +0.74, and +0.63 on average) and the average returns (+3.55%, +4.5%, +5.55%, and +3.93% on

average) are increased for controlled, random, low-cap, and high-cap portfolios, respectively.<sup>32</sup> An important consequence of these observations is that higher goodness-of-fit in parameter estimation is essential for increasing the reliability of trading signals, especially for a well diversified statistical arbitrage strategy.

## 4.5 Capitalization

As we have seen in Figure 5, in addition to the random portfolio, we also compare other portfolios that are formed by according to capitalization. The assets are sorted by market capitalization, and low- and high-capitalized assets are selected to form portfolios for trading. We note first that none of them outperform the controlled portfolio in all periods. Second, we find that the high-cap portfolio performs very well during 2007-2008 and 2009-2010, compared to low-cap portfolio. The most likely explanation for this is that small firms rely less on market information and so are less sensitive to market changes, as compared to large firms. Consequently, during volatile periods market impact is greater on large firms, and this is reflected by principal component analysis. Thus the residuals of large-cap portfolios are well-decoupled from the market. The third observation is that the volatility resulting from high-cap portfolios is lower than for other portfolios. The volatility level on average is about  $\sim 3.87\%$ , which is close to the level of the controlled portfolio ( $3.68\%$ ). Lastly, when the transaction costs are high, the low-cap portfolio performs much better than high-cap portfolio. See Table 8 and Table 9 in the Appendix.

## 4.6 Number of factors, window lengths, and transaction costs

Most studies in statistical arbitrage do not provide sufficient guidance or discussion about the role of important parameters, such as the number of factors, window lengths, and transaction costs. In this subsection we consider some issues associated with them.

First, the number of factors to be used varies depending on the portfolio size and the window length, since a larger portfolio or a longer window generally require more factors to be removed to get reliable residuals. This is because when using statistical factor models, with the factors constructed from observed data, the explaining power of a single factor can decrease when the data size increases. Second, setting a window length must take into account the trade-off between the benefit from accessing more data and the loss from using old information. Lastly, transaction costs also matter. For example, frequent trading may increase chances of gaining, but at the same time it will bring losses if the transaction costs are high.

We have found that the relations among these quantities are rather complicated, and a specific parameter set that leads to best performance is not clearly identified. However, considering an extensive range of parameters (see Figure 7), we observe the following.

1. Performance with short windows is more sensitive to the number of factors. If the window is short, not

---

<sup>32</sup>The volatility increases by limited amount.

many factors can be used. Thus, subtracting too many factors quickly degrades performance, compared to using longer windows.

2. The results with shorter window (with a smaller number of factors) are best when transaction costs are small. However, performance drops significantly as the transaction costs increase (to 15bp and above). In particular, if the transaction costs are high, then using long windows provides better and more stable performance. This is not surprising, since when it comes to trading environments with high transaction costs, frequent trading activities are not desirable.

The performance of a strategy considered (denoted by  $\Omega$ ) depends on three elements: the number of factors ( $p$ ), the length of estimation window ( $T$ ), and transaction cost ( $\epsilon$ ):

$$\Omega = \Omega(p, T, \epsilon) \quad (61)$$

The best choices, or the optimal values, in the sense of performance for  $p$  and  $T$  are entangled.

$$p^* = p^*(T, \epsilon) \quad (62)$$

$$T^* = T^*(p, \epsilon) \quad (63)$$

Thus, the performance of a strategy must be discussed always along with  $p$  and  $T$ .<sup>33</sup> The relations are visualized in Figure 7 as a surface of Sharpe ratios with respect to  $p$  and  $T$ . As seen from these plots, when the transaction costs increase the performances with shorter windows become unstable, while performance with longer windows is less so. As for the number of factors, any number larger than 10 yields unacceptable results in this context.

## 4.7 Portfolio sizes

In many cases we consider, as the portfolio size increases the strategy becomes more stable, as can be seen in Figure 8. This is consistent with the well-known idea of diversification. Since the total leverage is fixed regardless of the number of stocks participated, the optimization process has more range and give more stable results. However, this is not true in general. For example, in the case of a controlled portfolio we find that using  $N = 75$  stocks results in a higher Sharpe ratio on average than using  $N = 100$  stocks.

This also illustrates the role of the proposed optimization problem (Eq.43) given the total leverage. Figure 9 presents the performance without doing the optimization for asset allocation. That is, we allocate an identical amount to each asset, given the long/short signal of each. As expected, the performance is very unstable, since the allocation in this case may not satisfy the market-neutral conditions.

---

<sup>33</sup>Here we assume that the relations can change with the total number of stocks ( $N$ ) and data frequency ( $\Delta t$ ).

## 4.8 Recent years results

We found that our investment algorithm becomes less effective in recent years, 2013-2014. We need to examine this period more carefully. Since the essence of our approach is based on mean-reversion time control, we can check whether there is any unexpected change in realized mean-reversion times. Actually, Figure 6 already has a clue for this. We see that the buy-sell interval for the controlled portfolio is relatively longer both in 2005-2006 and 2013-2014. We note that the selected residuals in the controlled portfolio are expected to have shorter mean-reversion times. However, if they do not behave as predicted, the reliability of whole approach may drop. It appears that in 2005-2006, the relatively longer realized mean-reversion time did not affect performance, but in 2013-2014, we have seem to affect it. Unfortunately, we do not found a clear explanation for why this happens, but only a guess that the residuals in the recent market period have abnormally higher temporal correlations, which results in slower mean-reversions than expected.

## 4.9 Time-varying number of factors by variance explanations

So far, we have been using the fixed number of factors. This assumption can be relaxed by introducing a dynamic number of factors. Estimating the number of factors in high-dimensional factor models is beyond the scope of this study.<sup>34</sup>

Results with a dynamic number of factors with moving windows is presented in Figure 10. We can see that the numbers range from 1 to 25, which shows that there is time-varying correlation among assets. It can be also found that the market becomes very condensed during late-2008, mid-2010, and late-2011. We note that high volatility of markets is directly linked to the strong correlation of assets. The largest eigenvalue dynamics depicted in Figure 11 also illustrates this. Given a variance level, the corresponding number of factors is smaller when using shorter windows (60 days in Figure 10). This is because each factor brings higher proportion of variance, compared to using longer windows (120 days in Figure 10).

With a dynamic number factors we repeat the back-tests as before. Table 2 and Table 3 are the results for a controlled portfolio with transaction costs of 5bp and 20bp, respectively. The result with 20% is the best when the transaction costs are low. However, the performance from the variance explanation level of 30% to 40% are more stable with high-cost settings. We point out that there is no unique level of variance explanation that yields an all-the-time best performance. This would imply that using a dynamic number of factors solely based on variance explanation might not be ideal.

---

<sup>34</sup>However, a commonly used method is to determine the number of principal components that explain a certain level of total variance of returns. Instead of fixing the number of factors, we now fix the percentage of explained variance, such as 20%, 30%, 40%, 50%, and 60%, and the number of factors changes according to the variance levels.

## 5 Conclusions

Mean-reversion is the critical property for many statistical arbitrage strategies, and this paper studies its risk in detail. We have proposed a new approach to control the risk of mean-reversion time. Following the well-known Ornstein-Uhlenbeck modeling for residual processes, stocks are ranked by the estimated mean-reversion speeds. Then high-ranked stocks are selected to form the trading portfolio. This simple portfolio formation procedure combined with selective trading via goodness-of-fit in trading signals estimation effectively makes the strategy more reliable. Moreover, the optimization problem for trading volume plays a crucial role in obtaining market and dollar neutrality. All these contribute to robust performance regardless of market regime. The results obtained from back-testings have demonstrated the effectiveness of the approach.

Statistical arbitrage algorithms generally involve many parameters. There are, however, no “best” parameters to use. The exploration of the associated parameter regimes provides some insight into the relation between these parameters and trading performance.

## 6 Acknowledgement

Joongyeub Yeo gratefully acknowledges the support of ILJU Foundation Scholarship.

## References

- [1] Gatev E, Goetzmann WN, Rouwenhorst KG. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*. 2006;19 (3):797–827.
- [2] Do B, Faff R. Does simple pairs trading still work? *Financial Analysts Journal*. 2010;66 (4):83–95.
- [3] Do B, Faff R. Are pairs trading profits robust to trading costs? *Journal of Financial Research*. 2012;35 (2):261–287.
- [4] Fama EF, French K. Permanent and temporary components of stock prices. *Journal of Political Economics*. 1988;96(2):246–273.
- [5] Poterba JM, Summers LH. Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*. 1988;22(1):27–59.
- [6] Jorion P, Sweeney R. Mean reversion in real exchange rates: evidence and implications for forecasting. *Journal of International Money and Finance*. 1996;15(4):535–550.
- [7] Chaudhuri K, Wu Y. Random walk versus breaking trend in stock prices: evidence from emerging markets. *Journal of Banking and Finance*. 2003;27:575–592.
- [8] Carcano G, Falbo P, Stefani S. Speculative trading in mean reverting markets. *European Journal of Operation Research*. 2005;163:132–144.



- [9] Elliott RJ, van der Hoek J, Malcolm WP. Pairs trading. *Quantitative Finance*. 2005;5(3):271-276.
- [10] Do B, Faff R, Hamza K. A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 Financial Management Association European Conference*. 2006;.
- [11] Vidyamurthy G. *Pairs Trading: Quantitative Methods and Analysis*. Hoboken, N.J.: John Wiley & Sons; 2004.
- [12] Stock JH, Watson MW. Testing for common trends. *Journal of the American Statistical Association*. 1988;83 (404):1097.
- [13] Ross SA. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*. 1976;13 (3):341–360.
- [14] Bertram WK. Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications*. 2010;389 (11):2234–2243.
- [15] Jurek JW, Yang H. Dynamic portfolio selection in arbitrage. Working Paper, Harvard University. 2007;.
- [16] Avellaneda M, Lee JH. Statistical arbitrage in the US equities market. *Quantitative Finance*. 2010;10(7):761–782.
- [17] Lo AW. Maximum likelihood estimation of generalized It processes with discretely sampled data. *Econometric Theory*. *Econometric Theory*. 1988;4:231–247.
- [18] Bergstrom AR. *Continuous Time Econometric Modelling*. Oxford University Press; 1990.
- [19] Pedersen A. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observation. *Scandinavian Journal of Statistics*. 1995;22:55–71.
- [20] Ait-Sahalia Y. Maximum Likelihood Estimation of Discretely Sampled diffusion: A Closed-form Approximation Approach. *Econometrica*. 2002;70:223–262.
- [21] Yu J. Bias in the estimation of mean reversion parameter in continuous time models. *Journal of Econometrics*. 2012;169:114–122.
- [22] Phillips PCB, Yu J. Jackknifing bond option prices. *Review of Financial Studies*. 2005;18:707–742.

## A Supplemental tables

Table	Portfolio	$N$	$\epsilon$	$R^2$ -screening
Table 2	Controlled	Dynamic	5bp	No
Table 3	Controlled	Dynamic	20bp	No
Table 4	Controlled	75	5bp	No
Table 5	Controlled	75	5bp	Yes
Table 6	Random	75	5bp	No
Table 7	Random	75	5bp	Yes
Table 8	Low-Cap	75	5bp	No
Table 9	High-Cap	75	5bp	No
Table 10	anti-Controlled	75	5bp	No
Table 11	Controlled	100	5bp	No
Table 12	Controlled	50	5bp	No
Table 13	Controlled	25	5bp	No
Table 14	Controlled	75	10bp	No
Table 15	Controlled	75	No Cost	No

Table 1: Summary of tables in appendix. Tables with other parameters are available on request.

## Tables

Table 2: Dynamic number of factors: Controlled portfolio with  $N = 100$ , 5bp cost and no  $R^2$ -screening

Windows	Var. Exp.	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	20%	2.32	3.10	2.47	3.17	2.52	2.72	40.03%	5.22%
	30%	2.18	2.99	2.53	2.50	2.07	2.45	34.74%	5.22%
	40%	2.69	3.19	2.18	2.71	1.55	2.46	32.42%	5.00%
	50%	1.52	2.80	2.53	2.79	3.08	2.54	27.94%	4.46%
	60%	0.92	1.65	3.12	2.04	1.64	1.87	16.72%	4.12%
60	20%	2.19	2.35	1.45	2.31	0.49	1.76	25.79%	6.09%
	30%	2.80	2.17	1.45	2.34	0.79	1.91	27.75%	6.06%
	40%	1.59	2.25	1.54	1.81	2.11	1.86	23.93%	5.58%
	50%	1.60	3.26	1.33	1.77	1.42	1.88	22.49%	5.16%
	60%	1.83	2.30	0.94	1.59	0.48	1.43	14.95%	4.81%
90	20%	1.39	1.55	1.83	1.93	0.02	1.35	17.08%	5.78%
	30%	1.02	1.64	1.83	1.97	0.08	1.31	16.50%	5.71%
	40%	1.78	1.96	1.71	1.80	1.12	1.67	19.49%	5.20%
	50%	1.05	2.47	1.42	3.01	0.55	1.70	17.76%	4.71%
	60%	0.00	0.00	0.00	0.00	0.00	0.00	0.00%	—%
120	20%	1.15	0.69	2.30	1.33	1.98	1.49	17.44%	5.98%
	30%	1.45	0.65	2.30	1.33	2.03	1.55	17.60%	5.85%
	40%	1.70	1.90	2.09	1.51	1.43	1.72	21.32%	5.38%
	50%	1.07	0.72	1.28	1.08	1.80	1.19	11.58%	5.18%
	60%	1.06	1.58	1.56	1.25	0.15	1.12	11.56%	4.63%
Mean		1.65	2.06	1.89	2.01	1.33	1.79	21.95%	5.27%

Dynamic number of factors are applied, based on variance explanation by principal components. Transaction costs are 5bp. All values are annualized.

Table 3: Dynamic number of factors: Controlled portfolio with  $N = 100$ , 20bp cost and no  $R^2$ -screening

Windows	Var. Exp.	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	20%	(0.17)	1.65	0.88	0.49	(0.38)	0.50	7.70%	5.83%
	30%	(0.43)	1.51	0.92	(0.22)	(0.71)	0.21	5.33%	5.89%
	40%	(0.25)	1.60	0.53	(0.19)	(1.48)	0.04	4.07%	5.64%
	50%	(1.88)	1.14	0.42	(0.28)	(0.81)	(0.28)	1.96%	5.09%
	60%	(3.19)	(0.12)	0.50	(1.34)	(2.69)	(1.37)	(2.66)%	4.87%
60	20%	0.83	1.58	0.64	0.99	(0.91)	0.63	9.47%	6.54%
	30%	1.33	1.42	0.64	1.03	(0.67)	0.75	10.56%	6.50%
	40%	(0.04)	1.40	0.71	0.36	0.16	0.52	7.84%	6.00%
	50%	(0.39)	2.19	0.55	0.21	(0.87)	0.34	6.68%	5.55%
	60%	(0.67)	1.32	(0.09)	(0.15)	(2.10)	(0.34)	2.17%	5.22%
90	20%	0.43	0.96	1.12	0.92	(1.05)	0.48	7.12%	6.08%
	30%	0.01	1.04	1.12	0.95	(0.99)	0.42	6.70%	6.02%
	40%	0.42	1.26	1.00	0.70	(0.30)	0.62	8.28%	5.49%
	50%	(0.39)	1.61	0.64	1.77	(0.94)	0.54	6.97%	4.97%
	60%	(0.16)	0.28	0.32	0.14	(0.70)	(0.02)	2.52%	4.99%
120	20%	0.39	0.28	1.75	0.55	0.97	0.79	9.04%	6.25%
	30%	0.65	0.24	1.75	0.55	0.99	0.84	9.23%	6.11%
	40%	0.74	1.33	1.57	0.75	0.26	0.93	11.61%	5.62%
	50%	(0.08)	0.13	0.78	0.21	0.55	0.32	4.52%	5.41%
	60%	(0.26)	0.83	0.93	0.23	(1.35)	0.08	3.97%	4.87%
Mean		(0.16)	1.08	0.83	0.38	(0.65)	0.30	6.16%	5.65%

Dynamic number of factors are applied, based on variance explanation by principal components. Transaction costs are 20bp here. All values are annualized.

Table 4: Controlled portfolio (N=75,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	2.52	1.92	2.92	2.37	2.62	2.47	20.88%	3.84%
	10	0.57	1.68	0.36	0.81	1.16	0.92	6.40%	3.30%
	15	0.64	0.82	0.66	0.69	(0.27)	0.51	4.16%	2.67%
	20	(1.33)	(0.86)	(1.04)	(0.18)	(1.37)	(0.95)	(0.15)%	2.23%
60	5	1.80	2.58	1.63	2.19	1.93	2.02	20.08%	4.55%
	10	1.74	1.11	2.17	1.16	0.83	1.40	11.36%	3.99%
	15	1.64	1.60	2.11	0.63	0.61	1.32	9.38%	3.33%
	20	0.56	1.43	0.93	0.27	0.03	0.64	5.12%	3.04%
90	5	1.34	2.03	2.36	2.22	0.65	1.72	16.97%	4.47%
	10	2.42	1.15	1.11	1.63	2.03	1.67	13.04%	4.08%
	15	0.53	1.81	2.23	1.52	1.00	1.42	11.43%	3.72%
	20	1.03	1.61	2.13	1.52	1.35	1.53	10.15%	3.22%
120	5	1.23	2.55	1.82	1.85	1.25	1.74	18.15%	4.74%
	10	2.17	1.08	2.61	1.39	1.64	1.78	14.82%	4.07%
	15	1.33	1.90	2.10	1.45	0.52	1.46	12.35%	3.94%
	20	0.52	1.93	1.18	0.78	0.98	1.08	8.25%	3.72%
Mean		1.17	1.52	1.58	1.27	0.94	1.29	11.40%	3.68%

The out-of-sample performance of mean-reversion controlled portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is not applied here. Sharpe ratios are summarized for each window length, number of extracted factors, and period. In addition, mean Sharpe ratios, mean returns, and volatility of return of cumulative PnL are presented. All values are annualized.

Table 5: Controlled portfolio (N=75,  $\epsilon=5\text{bp}$ ,  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	2.96	2.57	2.24	2.59	1.73	2.42	27.94%	4.75%
	10	0.77	1.96	1.21	1.98	1.42	1.47	11.95%	4.08%
	15	0.75	1.53	1.34	1.66	0.58	1.17	7.75%	3.15%
	20	0.73	(0.72)	0.25	0.01	0.01	0.06	2.24%	2.51%
60	5	1.93	2.00	2.79	1.93	2.73	2.28	27.27%	4.99%
	10	1.55	0.59	2.75	1.28	1.78	1.59	14.56%	4.64%
	15	2.51	1.70	1.21	1.03	1.15	1.52	12.85%	4.23%
	20	1.77	1.54	1.29	0.97	(0.15)	1.08	8.34%	3.62%
90	5	1.40	1.17	1.82	2.54	0.98	1.58	17.24%	5.28%
	10	1.37	2.81	1.22	1.93	1.24	1.71	17.89%	4.87%
	15	1.54	1.70	2.22	1.53	0.93	1.59	15.51%	4.51%
	20	0.60	1.37	1.92	1.51	2.47	1.57	11.88%	3.86%
120	5	1.00	2.66	1.92	1.82	0.95	1.67	20.60%	5.41%
	10	2.34	2.79	2.14	1.11	0.29	1.73	18.79%	4.82%
	15	1.94	2.10	1.08	1.35	0.68	1.43	13.01%	4.47%
	20	2.14	1.70	1.18	0.66	0.68	1.27	11.35%	4.47%
Mean		1.58	1.72	1.66	1.49	1.09	1.51	14.95%	4.35%

The out-of-sample performance of mean-reversion controlled portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is applied here. Compared to Table 4, Sharpe ratios and returns are significantly boosted up. All values are annualized.

Table 6: Random portfolio (N=75,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	0.82	1.42	0.69	1.61	1.00	1.11	9.88%	4.54%
	10	1.31	(1.09)	0.07	0.34	(1.02)	(0.08)	1.38%	3.81%
	15	(0.00)	(0.75)	0.29	0.43	(0.26)	(0.06)	1.73%	2.99%
	20	(1.12)	(0.37)	(1.45)	(1.14)	(1.78)	(1.17)	(0.67)%	2.41%
60	5	0.97	2.20	1.16	0.73	0.48	1.11	12.25%	5.11%
	10	(0.31)	0.90	1.19	0.23	1.16	0.63	6.23%	4.43%
	15	(0.27)	0.32	1.17	(0.81)	(0.26)	0.03	2.60%	4.02%
	20	(0.34)	0.09	0.26	(1.33)	(0.20)	(0.30)	1.06%	3.62%
90	5	(0.07)	1.38	1.16	0.78	0.70	0.79	8.64%	5.20%
	10	(0.05)	1.06	1.19	(0.06)	0.29	0.49	6.41%	5.09%
	15	0.13	1.69	1.55	0.69	(0.10)	0.79	8.12%	4.29%
	20	(0.00)	1.12	(0.01)	(0.32)	1.01	0.36	3.82%	4.25%
120	5	0.12	0.79	(0.22)	1.33	1.12	0.63	6.27%	5.65%
	10	0.82	0.85	1.25	0.60	1.54	1.01	10.66%	5.26%
	15	1.05	1.80	1.33	0.94	0.96	1.22	10.91%	4.37%
	20	0.03	0.66	1.30	0.14	0.58	0.54	5.87%	4.45%
Mean		0.19	0.75	0.68	0.26	0.33	0.44	5.95%	4.34%

The out-of-sample performance of randomly chosen portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is not applied here. The performance from random portfolio has much fluctuations and unstable. As a result, the Sharpe ratios are significantly lower than tha from controlled portfolio. All values are annualized.

Table 7: Random portfolio (N=75,  $\epsilon=5\text{bp}$ ,  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	1.60	2.52	1.10	2.08	2.16	1.89	19.30%	4.82%
	10	1.29	0.02	1.00	1.40	1.22	0.98	6.99%	3.69%
	15	1.32	1.19	0.79	2.32	1.48	1.42	8.11%	2.95%
	20	(0.09)	0.41	(0.52)	(0.31)	(0.15)	(0.13)	1.81%	2.34%
60	5	0.94	2.41	1.95	2.19	0.18	1.53	19.94%	5.73%
	10	0.83	1.74	1.23	2.18	2.02	1.60	14.48%	4.56%
	15	0.42	0.96	2.14	0.80	1.34	1.13	9.26%	4.01%
	20	0.66	1.53	1.75	(0.01)	1.02	0.99	7.82%	3.55%
90	5	1.25	0.94	1.47	1.89	1.42	1.40	14.90%	5.35%
	10	0.71	1.68	1.92	1.46	1.19	1.39	14.65%	4.94%
	15	1.47	1.80	1.39	1.16	1.05	1.38	12.85%	4.53%
	20	0.95	1.56	(0.07)	1.45	(0.01)	0.78	6.49%	4.82%
120	5	0.37	1.38	0.58	1.07	1.43	0.96	10.49%	5.89%
	10	1.45	0.97	1.06	1.85	0.82	1.23	14.44%	6.17%
	15	1.87	0.86	1.65	1.11	0.49	1.19	12.45%	5.05%
	20	0.84	0.33	1.81	1.62	1.01	1.12	10.10%	4.54%
Mean		0.99	1.27	1.20	1.39	1.04	1.18	11.50%	4.56%

The out-of-sample performance of randomly chosen portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is applied here. The performance from random portfolio has been improved from R2-screening. All values are annualized.



Table 8: Low-cap portfolio (N=75,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	0.62	0.95	0.03	0.22	1.69	0.70	7.01%	5.83%
	10	0.32	(0.43)	(0.90)	(0.34)	(0.95)	(0.46)	(0.29)%	4.69%
	15	0.55	(0.66)	(0.49)	(0.28)	(1.71)	(0.52)	0.16%	3.68%
	20	(1.35)	(0.81)	0.29	(1.37)	(1.00)	(0.85)	(0.33)%	2.92%
60	5	1.49	1.44	1.20	0.65	2.03	1.36	18.44%	6.62%
	10	1.00	0.23	0.42	1.11	1.46	0.84	8.06%	5.41%
	15	0.52	(0.12)	1.28	1.22	(0.22)	0.54	5.54%	4.51%
	20	(0.13)	(0.47)	0.34	1.76	0.21	0.34	3.61%	4.13%
90	5	2.24	1.68	0.78	1.81	1.04	1.51	19.22%	6.11%
	10	1.84	1.55	0.62	1.12	1.06	1.24	13.14%	5.55%
	15	1.05	0.74	1.49	1.04	1.47	1.16	11.06%	4.91%
	20	1.39	(0.01)	0.87	1.00	0.42	0.73	6.90%	4.66%
120	5	1.22	0.85	0.91	1.46	1.94	1.27	15.42%	6.29%
	10	0.24	1.04	1.70	0.88	0.61	0.89	11.72%	6.13%
	15	1.37	0.59	1.01	1.21	0.88	1.01	10.18%	5.25%
	20	1.15	(0.22)	0.61	1.16	1.24	0.79	7.25%	5.12%
Mean		0.84	0.40	0.63	0.79	0.64	0.66	8.57%	5.11%

The out-of-sample performance of low-cap portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is not applied here. On average, it performs well during pre-crisis and post-crisis. All values are annualized.

Table 9: High-cap portfolio (N=75,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	(0.06)	1.37	2.14	(0.09)	1.10	0.89	7.98%	3.97%
	10	(0.71)	0.08	0.91	(0.51)	(0.56)	(0.16)	1.89%	3.41%
	15	(0.72)	(0.43)	0.38	(0.83)	(1.39)	(0.60)	0.56%	2.76%
	20	(1.94)	(0.92)	(2.05)	(1.53)	(2.06)	(1.70)	(1.63)%	2.25%
60	5	1.56	2.05	1.42	1.04	2.34	1.68	15.07%	4.36%
	10	1.18	0.70	2.04	0.40	0.38	0.94	7.70%	3.91%
	15	0.29	0.97	1.80	(0.45)	0.36	0.59	5.27%	3.39%
	20	0.49	0.28	0.32	(0.66)	(0.11)	0.06	2.58%	3.01%
90	5	0.89	1.73	1.80	(0.23)	0.95	1.03	11.69%	5.04%
	10	0.13	1.39	1.84	0.24	(0.03)	0.71	8.97%	4.65%
	15	(0.52)	1.10	0.83	(0.52)	0.83	0.34	4.46%	3.91%
	20	0.33	0.78	0.88	0.71	(1.32)	0.28	3.98%	3.63%
120	5	0.19	0.27	0.25	0.75	1.38	0.56	5.38%	5.30%
	10	0.34	1.12	1.18	0.45	1.41	0.90	8.51%	4.60%
	15	0.51	0.17	0.79	0.67	0.49	0.53	4.94%	4.09%
	20	0.99	1.54	1.09	0.69	0.15	0.89	7.36%	3.69%
Mean		0.18	0.76	0.98	0.01	0.25	0.44	5.92%	3.87%

The out-of-sample performance of high-cap portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. R2-screening is not applied here. On average, it performs well during crisis. All values are annualized.

Table 10: *Anti*-controlled portfolio (N=75,  $\epsilon$ =5bp, no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	1.40	1.53	0.67	0.17	2.25	1.20	11.22%	4.90%
	10	0.41	0.29	(0.12)	(0.10)	(1.60)	(0.23)	1.42%	3.84%
	15	(0.23)	(0.85)	(1.51)	(0.47)	(1.02)	(0.82)	(0.89)%	3.26%
	20	(1.01)	(1.70)	(0.78)	(1.47)	(1.73)	(1.34)	(1.32)%	2.64%
60	5	1.01	0.61	(0.42)	1.65	0.40	0.65	6.15%	5.63%
	10	(0.10)	1.32	0.23	0.96	0.88	0.66	6.19%	4.65%
	15	(0.43)	0.94	0.71	0.03	(0.07)	0.24	3.92%	4.13%
	20	(0.77)	(1.12)	(0.04)	(1.34)	0.02	(0.65)	(0.61)%	3.76%
90	5	0.87	0.45	1.82	(0.00)	0.85	0.80	9.27%	5.67%
	10	0.56	0.95	0.54	(0.28)	0.74	0.50	6.04%	5.34%
	15	0.36	0.27	0.87	(0.17)	1.69	0.61	5.86%	4.86%
	20	(0.58)	(0.35)	0.00	(0.08)	0.39	(0.12)	1.39%	3.96%
120	5	0.69	0.48	(0.40)	1.62	2.63	1.00	7.83%	6.09%
	10	0.44	0.98	0.99	0.91	2.52	1.17	10.44%	5.00%
	15	(0.56)	0.06	1.11	1.71	1.29	0.72	6.42%	4.68%
	20	(0.47)	0.64	0.75	0.03	0.18	0.23	3.79%	4.50%
Mean		0.10	0.28	0.28	0.20	0.59	0.29	4.82%	4.56%

The out-of-sample performance of anti-controlled portfolio. That is, stocks with slower mean-reverting residuals are selected to form a portfolio. The number of selected stocks is 75, and the transaction costs are 5bp per volume. All values are annualized.

Table 11: Controlled portfolio (N=100,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	1.58	2.06	2.24	2.54	1.50	1.99	15.73%	3.77%
	10	0.02	2.07	0.21	0.40	0.70	0.68	5.30%	3.17%
	15	(0.39)	(0.14)	0.69	(0.16)	(1.10)	(0.22)	1.72%	2.59%
	20	(1.43)	(0.54)	(1.37)	(1.13)	(1.88)	(1.27)	(0.65)%	2.22%
60	5	1.89	1.85	1.28	2.17	1.65	1.77	16.24%	4.51%
	10	1.07	1.76	2.14	0.75	0.17	1.18	9.69%	3.78%
	15	1.26	1.05	1.86	0.22	(0.40)	0.80	6.38%	3.32%
	20	0.93	1.54	0.47	0.28	(0.14)	0.62	4.75%	3.02%
90	5	1.65	1.66	1.49	1.66	0.74	1.44	13.75%	4.59%
	10	1.91	1.61	0.67	1.18	1.32	1.34	10.48%	4.09%
	15	0.50	2.56	1.87	1.27	0.85	1.41	10.27%	3.49%
	20	1.04	2.56	1.12	1.75	0.82	1.46	9.16%	3.09%
120	5	1.55	2.30	1.87	1.37	1.75	1.77	18.02%	4.69%
	10	2.21	2.26	2.64	1.81	1.48	2.08	18.39%	3.96%
	15	1.77	2.04	2.62	1.29	0.89	1.72	13.60%	3.65%
	20	0.96	0.85	0.96	0.65	0.87	0.86	6.72%	3.71%
Mean		1.03	1.59	1.30	1.00	0.58	1.10	9.97%	3.60%

The out-of-sample performance of controlled portfolio. The number of selected stocks is 100, and the transaction costs are 5bp per volume. R2-screening is not applied here. On average, the Sharpe ratios and returns are slightly lower than that of portfolio size of 75. The returns the All values are annualized.

Table 12: Controlled portfolio (N=50,  $\epsilon=5\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	2.55	1.46	1.95	1.02	1.69	1.73	19.22%	5.16%
	10	0.04	0.80	0.06	0.70	(0.51)	0.22	3.34%	4.43%
	15	(0.02)	0.38	(0.33)	0.44	(0.67)	(0.04)	1.98%	3.70%
	20	(1.38)	(0.58)	(1.17)	(0.22)	(2.26)	(1.12)	(1.29)%	3.12%
60	5	0.25	0.61	2.14	1.43	0.68	1.02	12.81%	6.14%
	10	1.06	1.17	1.74	1.19	0.25	1.08	11.07%	4.93%
	15	1.06	(0.03)	1.61	0.55	0.16	0.67	6.53%	4.43%
	20	1.25	1.57	(0.32)	0.55	(0.88)	0.43	3.91%	4.51%
90	5	1.12	1.62	0.83	1.53	0.85	1.19	14.73%	5.95%
	10	1.83	1.08	1.68	0.27	1.04	1.18	13.28%	5.49%
	15	0.78	1.54	1.22	0.59	0.56	0.94	9.60%	4.94%
	20	0.58	1.88	1.62	0.84	0.09	1.00	9.23%	4.16%
120	5	2.11	0.73	0.96	0.97	1.57	1.27	14.96%	6.11%
	10	1.73	1.79	2.30	0.73	0.88	1.49	18.87%	5.52%
	15	1.58	1.30	2.14	0.72	0.36	1.22	13.79%	5.12%
	20	1.27	0.60	1.92	0.34	0.73	0.97	9.84%	4.75%
Mean		0.99	0.99	1.15	0.73	0.28	0.83	10.12%	4.90%

The out-of-sample performance of controlled portfolio. The number of selected stocks is 50, and the transaction costs are 5bp per volume. R2-screening is not applied here. On average, the Sharpe ratios and returns are more unstable and much lower than that of portfolio size of 75. The returns and All values are annualized.

Table 13: Controlled portfolio (N=25,  $\epsilon=5$ bp, no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	0.93	0.49	1.39	0.81	2.08	1.14	15.62%	7.03%
	10	(0.12)	1.11	0.78	0.39	0.39	0.51	6.08%	5.50%
	15	(0.13)	(0.24)	1.46	(0.08)	(0.60)	0.08	2.97%	4.68%
	20	(0.97)	(0.16)	(0.18)	(0.11)	(2.02)	(0.69)	(0.33)%	4.28%
60	5	0.68	1.23	1.29	0.82	0.69	0.94	14.26%	7.28%
	10	1.17	1.17	0.63	(0.57)	0.22	0.52	6.87%	5.89%
	15	1.72	0.82	2.31	(0.25)	0.33	0.99	10.44%	5.14%
	20	1.19	2.00	(0.75)	(0.22)	(0.93)	0.26	3.15%	5.27%
90	5	0.16	1.67	1.00	0.50	(0.09)	0.65	10.17%	7.40%
	10	2.25	0.25	1.25	0.96	1.05	1.15	14.09%	6.53%
	15	1.28	1.68	0.85	0.96	0.61	1.08	13.57%	6.30%
	20	0.52	1.24	1.67	0.03	(0.53)	0.59	7.47%	5.47%
120	5	1.62	0.55	0.78	0.55	1.36	0.97	12.38%	6.98%
	10	0.84	0.52	1.73	0.34	0.53	0.79	12.03%	7.01%
	15	1.07	(0.41)	1.40	0.77	(0.46)	0.47	7.33%	6.47%
	20	0.45	0.60	1.47	0.50	0.86	0.78	9.53%	5.83%
Mean		0.79	0.78	1.07	0.34	0.22	0.64	9.10%	6.07%

The out-of-sample performance of controlled portfolio. The number of selected stocks is 25, and the transaction costs are 5bp per volume. R2-screening is not applied here. On average, the Sharpe ratios and returns are more unstable and significantly lower than that of portfolio size of 75. The returns the All values are annualized.

Table 14: Controlled portfolio (N=75,  $\epsilon=10\text{bp}$ , no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	1.09	1.40	0.90	0.59	0.35	0.86	8.52%	4.60%
	10	(1.80)	0.40	(0.81)	(0.28)	(1.73)	(0.84)	(1.06)%	3.90%
	15	(1.77)	(1.28)	(1.19)	(1.59)	(3.26)	(1.82)	(2.98)%	3.24%
	20	(3.85)	(2.59)	(2.85)	(3.05)	(4.39)	(3.35)	(4.88)%	2.68%
60	5	0.35	1.10	1.58	1.17	0.50	0.94	10.75%	5.33%
	10	0.47	1.00	1.21	0.59	(0.62)	0.53	5.67%	4.34%
	15	0.75	(0.54)	1.11	(0.05)	(1.22)	0.01	2.48%	3.85%
	20	0.18	0.77	(0.09)	(0.87)	(1.37)	(0.27)	1.41%	3.48%
90	5	1.00	1.46	1.34	1.49	0.16	1.09	11.71%	5.17%
	10	0.99	1.14	0.57	0.68	0.69	0.81	7.83%	4.87%
	15	0.31	1.55	1.20	0.27	0.38	0.74	6.79%	4.08%
	20	0.35	1.86	1.65	0.60	(0.40)	0.81	6.88%	3.55%
120	5	1.78	2.02	1.30	1.24	1.23	1.51	15.77%	4.95%
	10	1.36	2.27	1.96	1.22	0.53	1.47	14.61%	4.49%
	15	1.89	1.46	1.63	0.41	0.02	1.08	9.85%	4.21%
	20	0.45	0.75	1.18	0.01	0.00	0.48	5.59%	4.32%
Mean		0.22	0.80	0.67	0.15	(0.57)	0.25	6.18%	4.19%

The out-of-sample performance of mean-reversion controlled portfolio. The number of selected stocks is 75, and the transaction costs are 10bp per volume. R2-screening is not applied here. When the transaction costs become higher, the Sharpe ratios dropped down and become unstable. All values are annualized.

Table 15: Controlled portfolio (N=75,  $\epsilon=0$ bp, no  $R^2$ -screening)

Windows	Factors	Periods					All		
		2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	Mean	Return	Volatility
30	5	3.19	2.74	2.38	2.94	2.58	2.76	28.40%	4.24%
	10	0.68	2.44	0.99	2.78	1.26	1.63	11.24%	3.55%
	15	1.27	1.26	1.17	1.79	1.14	1.32	7.86%	2.93%
	20	0.17	0.68	0.39	1.61	1.33	0.84	4.54%	2.37%
60	5	1.44	1.79	2.19	2.19	1.72	1.87	21.60%	5.10%
	10	1.60	1.95	2.21	1.91	0.82	1.70	14.84%	4.13%
	15	2.17	0.80	2.24	1.46	0.45	1.42	10.74%	3.65%
	20	1.80	2.29	0.99	1.07	0.87	1.41	9.30%	3.30%
90	5	1.72	1.94	1.87	2.23	0.96	1.74	19.30%	5.02%
	10	1.83	1.83	1.09	1.46	1.71	1.58	14.81%	4.71%
	15	1.29	2.41	1.88	1.13	1.76	1.69	13.71%	3.94%
	20	1.47	2.91	2.50	1.86	1.00	1.95	14.31%	3.42%
120	5	2.34	2.47	1.77	1.80	1.97	2.07	22.82%	4.83%
	10	2.00	2.84	2.41	1.94	1.42	2.12	21.77%	4.38%
	15	2.57	2.20	2.10	1.15	0.95	1.80	15.99%	4.11%
	20	1.26	1.46	1.61	0.89	0.89	1.22	10.78%	4.20%
Mean		1.67	2.00	1.74	1.76	1.30	1.70	15.13%	3.99%

The out-of-sample performance of mean-reversion controlled portfolio. The number of selected stocks is 75, and there is no transaction cost.  $R^2$ -screening is not applied here. When the transaction costs become higher, the Sharpe ratios dropped down and become unstable. All values are annualized.



## Figures

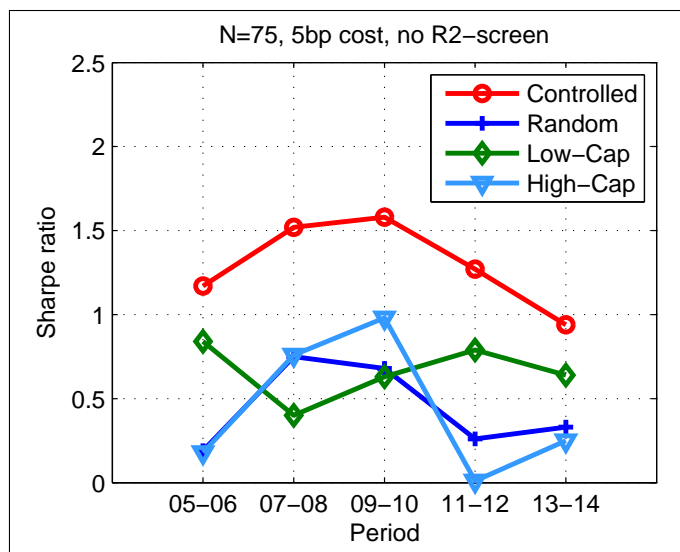


Figure 1: The effect of mean-reversion controlled portfolio on Sharpe ratios. The size of portfolio is 75, the transaction cost is 5bp per volume, and  $R^2$ -screening is not applied here. It is clear to see that the portfolio selection using the estimated mean-reversion speed (red line) performs better than any other portfolios.

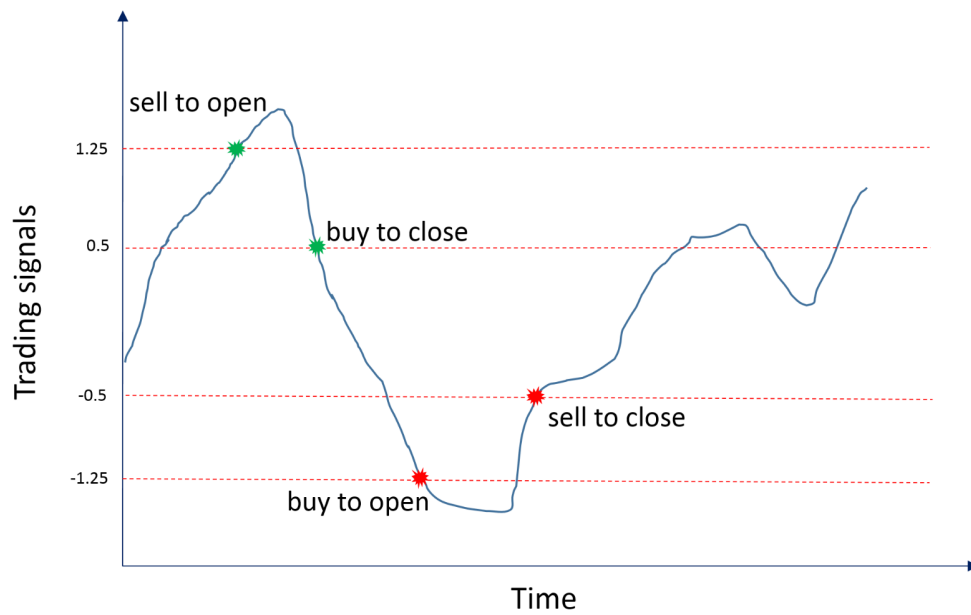


Figure 2: A schematic diagram for a long/short strategy. When the trading signal hits 1.25, it is presumed to be over-valued, so we sell the stock to open the short position (green dots), expecting it to go down soon. If it hits 0.5, we buy the stock to close the short position, which leads us to obtain profits. Similar strategy can be applied to the open/close of long position (red dots).

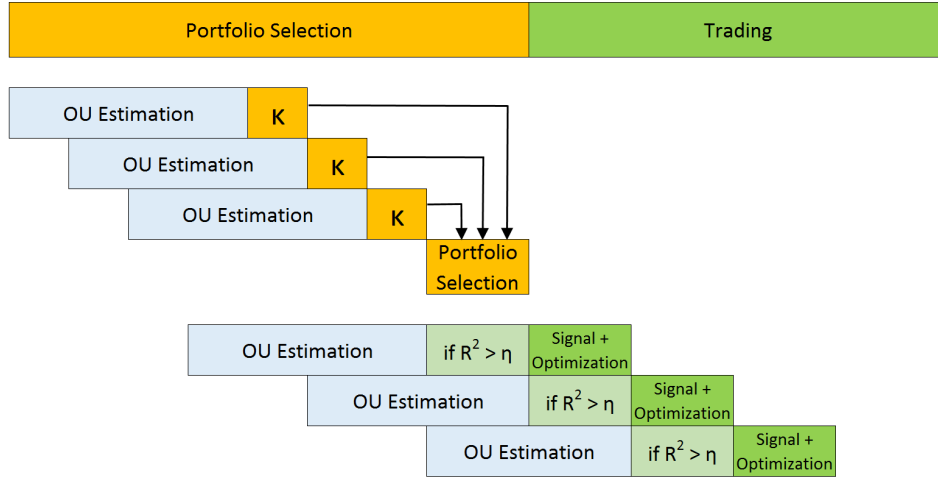


Figure 3: Diagram for trading scheme in a single testing cycle. The mean-reversion speeds (or the inverse of the mean-reversion times) are estimated with a moving estimation window, and the quality scores are recorded until the end of the training period. Based on the quality scores, high-ranked stocks are selected to form the trading portfolio. When the trading signal is judged to be statistically reliable, the optimization problem for asset allocations is solved for trading. Note that the estimation windows for training and for trading are set to be the same.

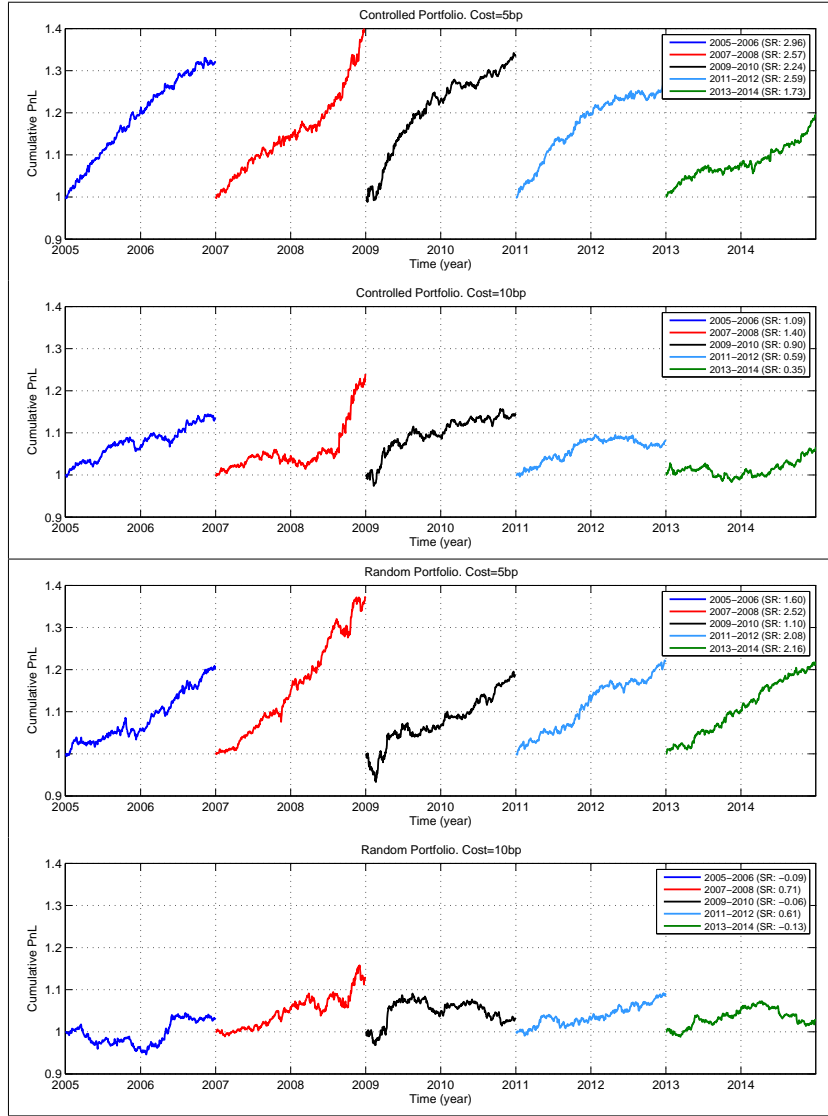


Figure 4: Sample results of cumulative PnL for each period. The results are generated using two portfolios (controlled and random) with two transaction costs (of 5bp and 10bp per trading volume). Here the portfolio size is  $N = 75$ , the number of removed factors are 5 and the window length is 30 business days.  $R^2$ -screening is applied. Note that the performance of controlled portfolio with  $R^2$ -screening is very stable and yields high Sharpe ratios. It is also less affected by increased transaction costs, compared to random portfolio. The results with various parameters are provided in the appendix.

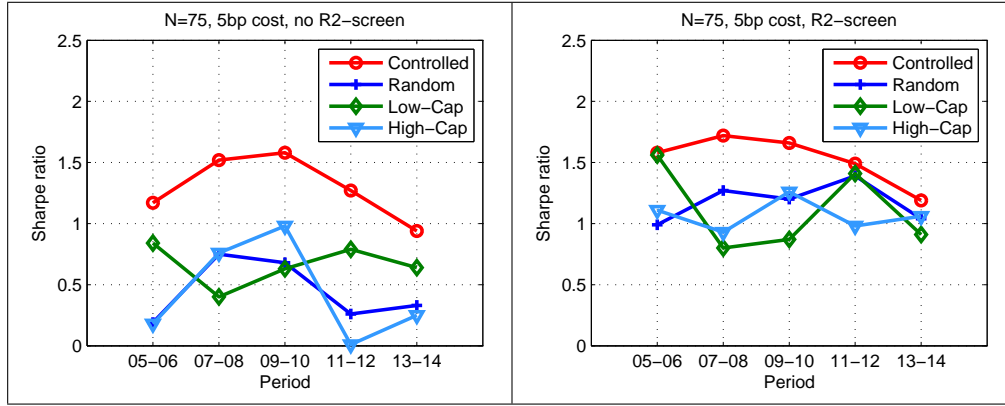


Figure 5: The out-of-sample performance from four different portfolios. The number of selected stocks is 75, and the transaction costs are 5bp per volume. (Left) Without  $R^2$ -screening. Compared with the performance from other portfolios with same condition, the Sharpe ratios are mostly improved when using the (mean-reversion) controlled portfolio. (Right) With  $R^2$ -screening. The overall level of performance enhanced by selective trading based on  $R^2$ -screening. All values are annualized.

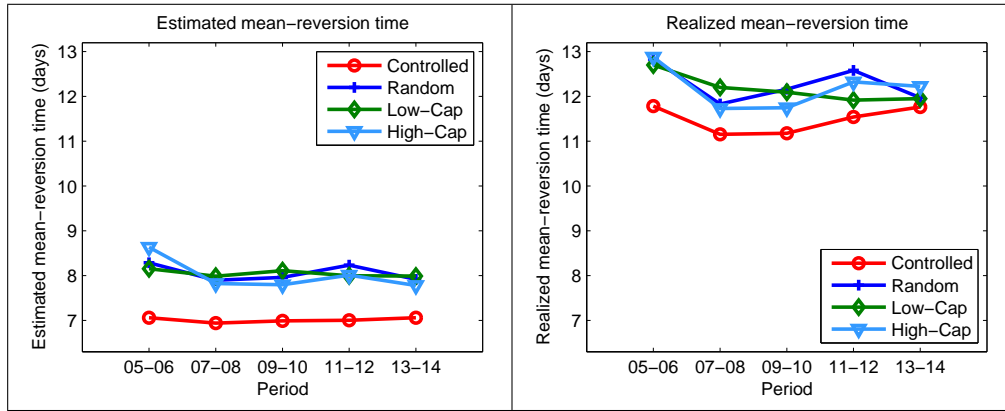


Figure 6: Mean-reversion times (in days): estimated (Left) and realized (Right) values are presented. The mean-reversion controlled portfolio exhibits significantly shorter trading intervals compared to other three portfolios. Note the values are averaged.

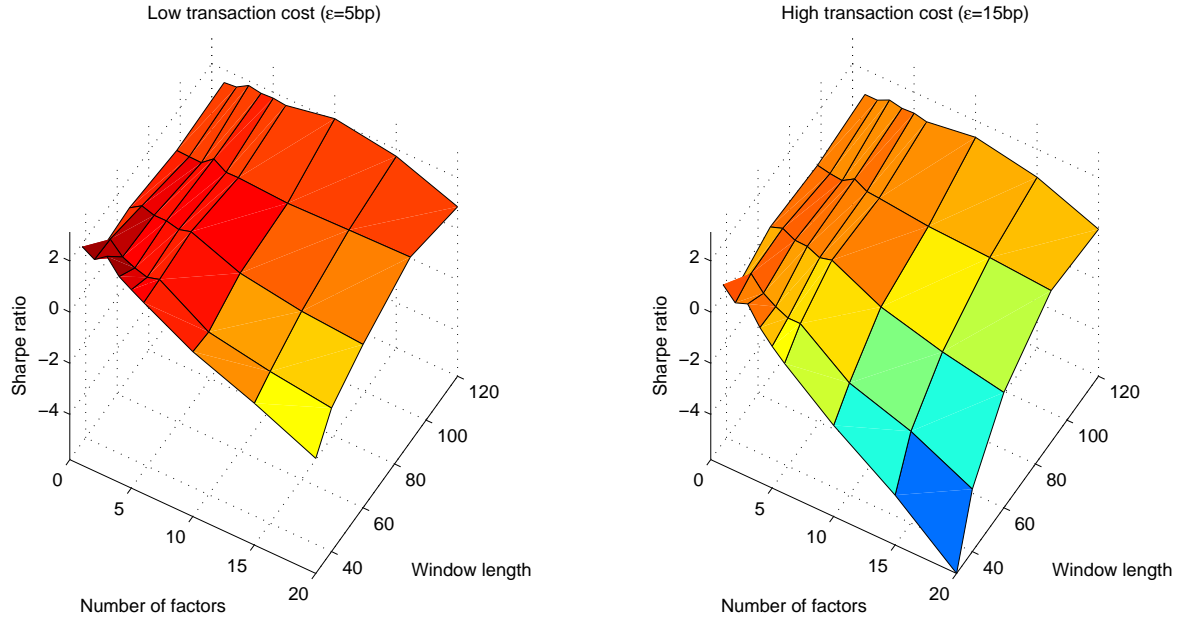


Figure 7: The performance surface as a function of number of factors ( $p$ ) and window length ( $T$ ), for two different transaction costs settings:  $\epsilon=5\text{bp}$  (left) and  $\epsilon=15\text{bp}$  (right). The performance is measured by averaged Sharpe ratios. The interruption from increasing costs is bigger for shorter windows than longer windows.

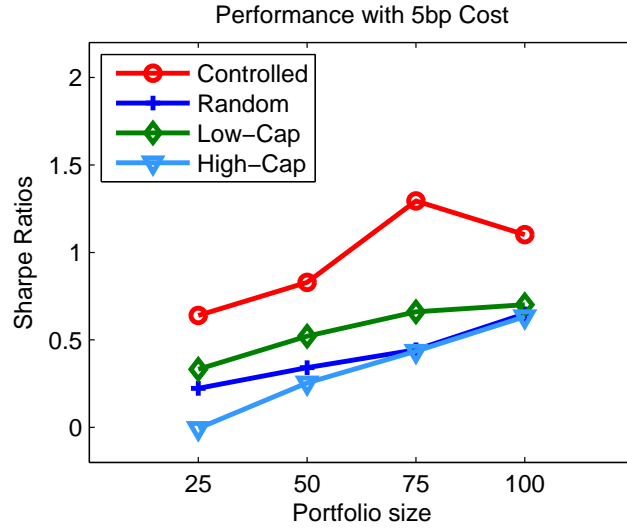


Figure 8: Sharpe ratios with respect to portfolio sizes: X axis represents portfolio sizes (25, 50, 75, and 100 stocks) and Y axis represents Sharpe ratios. The results are presented with transaction costs of 5bp. Note that the values on the Y axis are averaged ones over all number of factors, estimation windows, and periods. In general, the performance is better with larger portfolio sizes. However, for controlled portfolio, using  $N = 75$  is better than using  $N = 100$ .



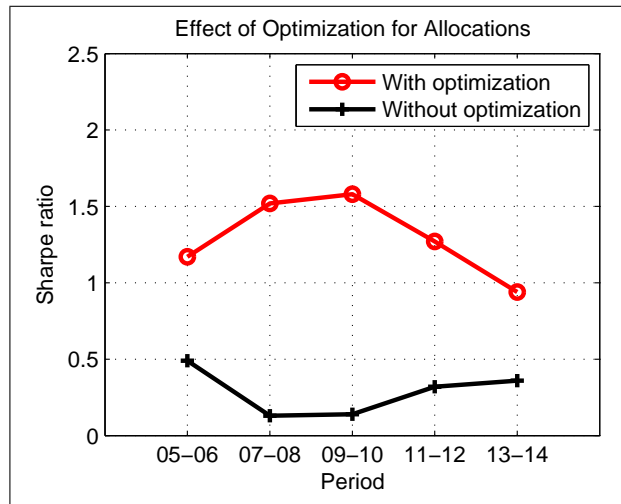


Figure 9: Performance comparison between with and without solving optimization problem for asset allocations. In void of optimized asset allocations, the strategy is unstable and cannot achieve the market-neutrality.

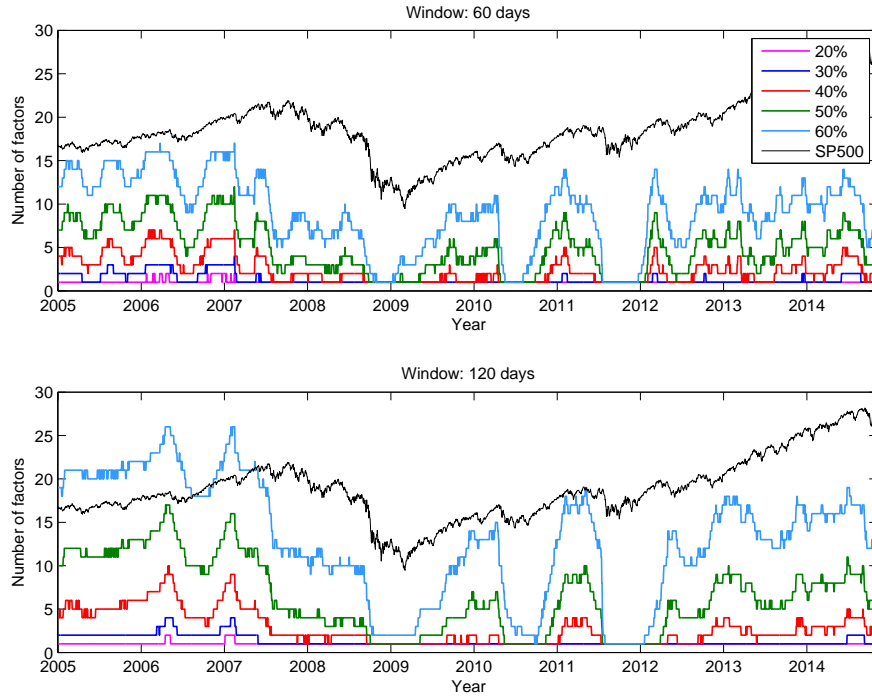


Figure 10: Dynamics of number of factors that explain different levels of variance. The numbers differ with estimation window: 60 days (top) and 120 days (bottom). The total number of stocks is 378. The scaled S&P500 index is compared. The correlation is larger in volatile markets.

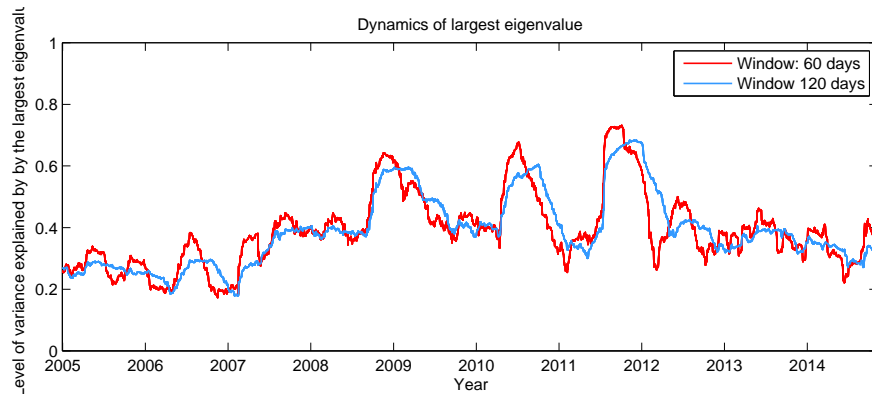


Figure 11: The largest eigenvalue of correlation matrix of normalized original returns with moving window. The largest eigenvalue increase significantly when market becomes volatile. The shorter windows response more sensitively to the market.

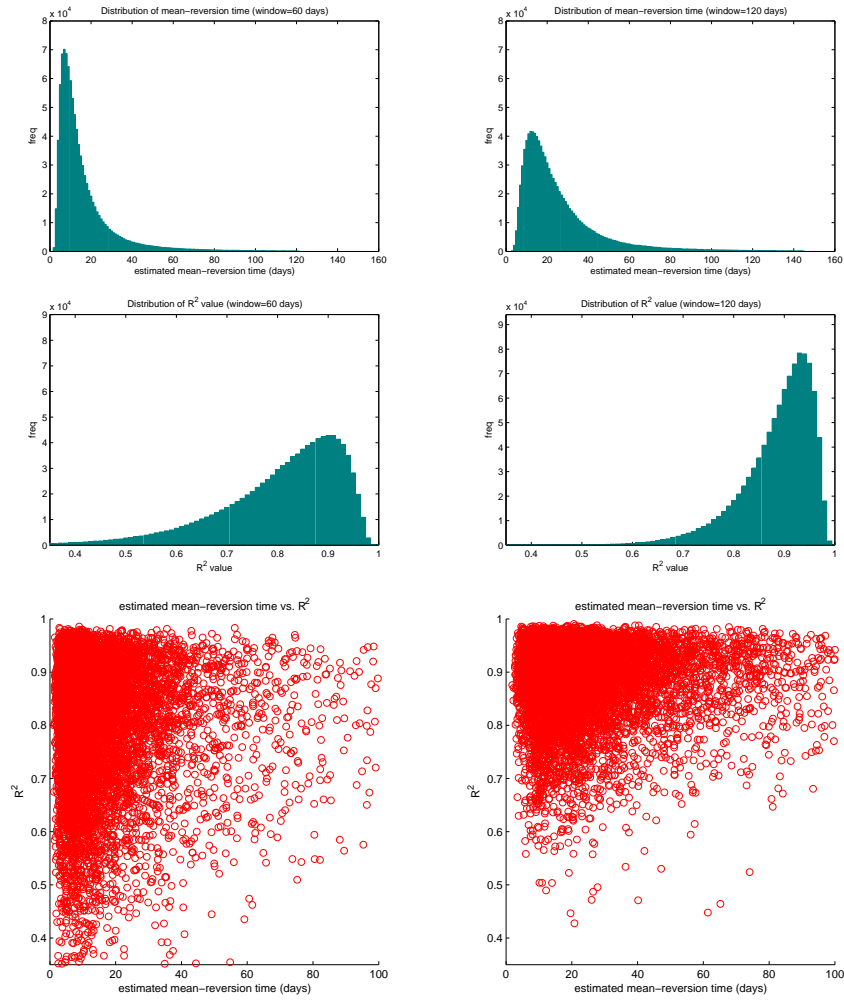


Figure 12: An exemplary statistics of mean reversion time and  $R^2$  values. (Top) Distribution of estimated mean-reversion time. (Middle) Distribution of  $R^2$  values from OU estimations. (Bottom) Scatter plots between the two. Here the number of factors is 20, and estimation window length is 60 (Left) and 120 days (Right). Note that the estimation mean-reversion times and  $R^2$  values are larger for longer estimation window, and vice versa.