Statistical Arbitrage in the U.S. Equities Market

Marco Avellaneda[*][†] and Jeong-Hyun Lee[*]

July 11, 2008

Abstract

We study model-driven statistical arbitrage strategies in U.S. equities. Trading signals are generated in two ways: using Principal Component Analysis and using sector ETFs. In both cases, we consider the residuals, or idiosyncratic components of stock returns, and model them as a mean-reverting process, which leads naturally to "contrarian" trading signals.

The main contribution of the paper is the back-testing and comparison of market-neutral PCA- and ETF- based strategies over the broad universe of U.S. equities. Back-testing shows that, after accounting for transaction costs, PCA-based strategies have an average annual Sharpe ratio of 1.44 over the period 1997 to 2007, with a much stronger performances prior to 2003: during 2003-2007, the average Sharpe ratio of PCA-based strategies was only 0.9. On the other hand, strategies based on ETFs achieved a Sharpe ratio of 1.1 from 1997 to 2007, but experience a similar degradation of performance after 2002. We introduce a method to take into account daily trading volume information in the signals (using "trading time" as opposed to calendar time), and observe significant improvements in performance in the case of ETF-based signals. ETF strategies which use volume information achieve a Sharpe ratio of 1.51 from 2003 to 2007.

The paper also relates the performance of mean-reversion statistical arbitrage strategies with the stock market cycle. In particular, we study in some detail the performance of the strategies during the liquidity crisis of the summer of 2007. We obtain results which are consistent with Khandani and Lo (2007) and validate their "unwinding" theory for the quant fund drawdown of August 2007.

1 Introduction

The term statistical arbitrage encompasses a variety of strategies and investment programs. Their common features are: (i) trading signals are systematic, or

---

[*]Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, N.Y. 10012 USA

[†]Finance Concepts SARL, 49-51 Avenue Victor-Hugo, 75116 Paris, France.

1

rules-based, as opposed to driven by fundamentals, (ii) the trading book is market-neutral, in the sense that it has zero beta with the market, and (iii) the mechanism for generating excess returns is statistical. The idea is to make many bets with positive expected returns, taking advantage of diversification across stocks, to produce a low-volatility investment strategy which is uncorrelated with the market. Holding periods range from a few seconds to days, weeks or even longer.

Pairs-trading is widely assumed to be the "ancestor" of statistical arbitrage. If stocks P and Q are in the same industry or have similar characteristics (*e.g.* Exxon Mobile and Conoco Phillips), one expects the returns of the two stocks to track each other after controlling for beta. Accordingly, if P_t and Q_t denote the corresponding price time series, then we can model the system as

$$ln(P_t/P_{t_0}) = \alpha(t - t_0) + \beta ln(Q_t/Q_{t_0}) \ + \ X_t \tag{1}$$

or, in its differential version,

$$\frac{dP_t}{P_t} \ = \ \alpha dt \ + \ \beta \frac{dQ_t}{Q_t} \ + dX_t, \tag{2}$$

where X_t is a stationary, or mean-reverting, process. This process will be referred to as the *cointegration residual*, or *residual*, for short, in the rest of the paper. In many cases of interest, the drift α is small compared to the fluctuations of X_t and can therefore be neglected. This means that, after controlling for beta, the long-short portfolio oscillates near some statistical equilibrium. The model suggests a contrarian investment strategy in which we go long 1 dollar of stock P and short β dollars of stock Q if X_t is small and, conversely, go short P and long Q if X_t is large. The portfolio is expected to produce a positive return as valuations converge (see Pole (2007) for a comprehensive review on statistical arbitrage and co-integration). The mean-reversion paradigm is typically associated with market over-reaction: assets are temporarily under- or over-priced with respect to one or several reference securities (Lo and MacKinley (1990)).

Another possibility is to consider scenarios in which one of the stocks is expected to out-perform the other over a significant period of time. In this case the co-integration residual should not be stationary. This paper will be principally concerned with mean-reversion, so we don't consider such scenarios.

"Generalized pairs-trading", or trading groups of stocks against other groups of stocks, is a natural extension of pairs-trading. To explain the idea, we consider the sector of biotechnology stocks. We perform a regression/cointegration analysis, following (1) or (2), for each stock in the sector with respect to a benchmark sector index, *e.g.* the Biotechnology HOLDR (BBH). The role of the stock Q would be played by BBH and P would an arbitrary stock in the biotechnology sector. The analysis of the residuals, based of the magnitude of X_t, suggests typically that some stocks are cheap with respect to the sector, others expensive and others fairly priced. A generalized pairs trading book, or statistical arbitrage book, consists of a collection of "pair trades" of stocks relative to the ETF (or, more generally, factors that explain the systematic stock

2

returns). In some cases, an individual stock may be held long against a short position in ETF, and in others we would short the stock and go long the ETF. Due to netting of long and short positions, we expect that the net position in ETFs will represent a small fraction of the total holdings. The trading book will look therefore like a long/short portfolio of single stocks. This paper is concerned with the design and performance-evaluation of such strategies.

The analysis of residuals will be our starting point. Signals will be based on relative-value pricing within a sector or a group of peers, by decomposing stock returns into systematic and idiosyncratic components and statistically modeling the idiosyncratic part. The general decomposition may look like

$$\frac{dP_t}{P_t} \;=\; \alpha dt \;+\; \sum_{j=1}^{n} \beta_j \, F_t^{(j)} \;+\; dX_t, \tag{3}$$

where the terms $F_t^{(j)}, j = 1, ..., n$ represent returns of risk-factors associated with the market under consideration. This leads to the interesting question of how to derive equation (3) in practice. The question also arises in classical portfolio theory, but in a slightly different way: there we ask what constitutes a "good" set of risk-factors from a risk-management point of view. Here, the emphasis is instead on the residual that remains *after* the decomposition is done. The main contribution of our paper will be to study how different sets of risk-factors lead to different residuals and hence to different profit-loss (PNL) for statistical arbitrage strategies.

Previous studies on mean-reversion and contrarian strategies include Lehmann (1990), Lo and MacKinlay (1990) and Poterba and Summers (1988). In a recent paper, Khandani and Lo (2007) discuss the performance of the Lo-MacKinlay contrarian strategies in the context of the liquidity crisis of 2007 (see also references therein). The latter strategies have several common features with the ones developed in this paper. Khandani and Lo (2007) market-neutrality is enforced by ranking stock returns by quantiles and trading "winners-versus-losers", in a dollar-neutral fashion. Here, we use risk-factors to extract trading signals, i.e. to detect over- and under-performers. Our trading frequency is variable whereas Khandani-Lo trade at fixed time-intervals. On the parametric side, Poterba and Summers (1988) study mean-reversion using auto-regressive models in the context of international equity markets. The models of this paper differ from the latter mostly in that we immunize stocks against market factors, i.e. we consider mean-reversion of residuals (relative prices) and not of the prices themselves.

The paper is organized as follows: in Section 2, we study market-neutrality using two different approaches. The first method consists in extracting risk-factors using Principal Component Analysis (Jolliffe (2002)). The second method uses industry-sector ETFs as proxies for risk factors. Following other authors, we show that PCA of the correlation matrix for the broad equity market in the U.S. gives rise to risk-factors that have economic significance because they can be interpreted as long-short portfolios of industry sectors. Furthermore, the stocks that contribute the most to a particular factor are not necessarily the largest capitalization stocks in a given sector. This suggests that, unlike ETFs,

3

PCA-based risk factors are not biased towards large-capitalization stocks. We also observe that the variance explained by a fixed number of PCA eigenvectors varies significantly across time, leading us to conjecture that the number of explanatory factors needed to describe stock returns is variable and that this variability is linked with the investment cycle, or the changes in the risk-premium for investing in the equity market.[1]

In Section 3 and 4, we construct the trading signals. This involves the statistical estimation of the process X_t for each stock at the close of each trading day, using historical data prior to the close. Estimation is always done looking back at the historical record, thus simulating decisions which would take place in real automatic trading. Using daily end-of-day (EOD) data, we perform a full calculation of daily trading signals, going back to 1996 in some cases and to 2002 in others, across the broad universe of stocks with market-capitalization of more than 1 billion USD at the trade date.[2]

The estimation and trading rules are kept simple to avoid data-mining. For each stock in the universe, the parameter estimation is done using a 60-day trailing estimation window, which corresponds roughly to one earnings cycle. The length of the window is fixed once and for all in the simulations and is not changed from one stock to another. We use the same fixed-length estimation window, we choose as entry point for trading any residual that deviates by 1.25 standard deviations from equilibrium, and we exit trades if the residual is less than 0.5 standard deviations from equilibrium, uniformly across all stocks.

In Section 5 we back-test different strategies which use different sets of factors to generate residuals, namely: synthetic ETFs based on capitalization-weighted indices, actual ETFs, a fixed number of factors generated by PCA, a variable number of factors generated by PCA. Due to the mechanism described aboive used to generate trading systems, the simulation is out-of-sample, in the sense that the estimation of the residual process at time t uses information available only before this time. In all cases, we assume a slippage/transaction cost of 0.05% or 5 basis points per trade (a round-trip transaction cost of 10 basis points).

In Section 6, we consider a modification of the strategy in which signals are estimated in "trading time" as opposed to calendar time. In the statistical analysis, using trading time on EOD signals is effectively equivalent to multiplying daily returns by a factor which is inversely proportional to the trading volume for the past day. This modification accentuates (i.e. tends to favor) contrarian price signals taking place on low volume and mitigates (i.e. tends not to favor) contrarian price signals which take place on high volume. It is as if we "believe more" a print that occurs on high volume and less ready to bet against it. Back-testing the statistical arbitrage strategies using trading-time signals leads to improvements in most strategies, suggesting that volume information is valuable in the mean-reversion context, even at the EOD time-scale.

---

[1]See Scherer and Avellaneda (2002) for similar observations for Latin American debt securities in the 1990's.

[2]The condition that the company must have a given capitalization at the trade date (as opposed to at the time this paper was written), avoids survivorship bias.

In Section 7, we discuss the performance of statistical arbitrage in 2007, and particularly around the inception of the liquidity crisis of August 2007. We compare the performances of the mean-reversion strategies with the ones studied in the recent work of Khandani and Lo (2007). Conclusions are presented in Section 8.

2 A quantitative view of risk-factors and market-neutrality

We divide the world schematically into "indexers' and "market-neutral agents". Indexers seek exposure to the entire market or to specific industry sectors. Their goal is generally to be long the market or sector with appropriate weightings in each stock. Market-neutral agents seek returns which are uncorrelated with the market.

Let us denote by $\{R_i\}_{i=1}^N$ the returns of the different stocks in the trading universe over an arbitrary one-day period (from close to close). Let F represent the return of the "market portfolio" over the same period, (*e.g.* the return on a capitalization-weighted index, such as the S&P 500). We can write, for each stock in the universe,

$$R_i \; = \; \beta_i F \; + \; \tilde{R}_i, \tag{4}$$

which is a simple regression model decomposing stock returns into a systematic component $\beta_i F$ and an (uncorrelated) idiosyncratic component $\tilde{R}_i$. Alternatively, we consider multi-factor models of the form

$$R_i \; = \; \sum_{j=1}^m \beta_{ij} F_j \; + \; \tilde{R}_i. \tag{5}$$

Here there are m factors, which can be thought of as the returns of "benchmark" portfolios representing systematic factors. A trading portfolio is said to be market-neutral if the dollar amounts $\{Q_i\}_{i=1}^N$ invested in each of the stocks are such that

$$\overline{\beta}_j \; = \; \sum_{i=1}^N \beta_{ij} Q_i \; = 0, \quad j = 1, 2, ..., m. \tag{6}$$

The coefficients $\overline{\beta}_j$ correspond to the portfolio betas, or projections of the portfolio returns on the different factors. A market-neutral portfolio has vanishing portfolio betas; it is uncorrelated with the market portfolio or factors that drive the market returns. It follows that the portfolio returns satisfy

$$\sum_{i=1}^{N} Q_i R_i = \sum_{i=1}^{N} Q_i \left[\sum_{j=1}^{m} \beta_{ij} F_j \right] + \sum_{i=1}^{N} Q_i \tilde{R}_i$$

$$= \sum_{j=1}^{m} \left[\sum_{i=1}^{N} \beta_{ij} Q_i \right] F_j + \sum_{i=1}^{N} Q_i \tilde{R}_i$$

$$= \sum_{i=1}^{N} Q_i \tilde{R}_i \qquad (7)$$

Thus, a market-neutral portfolio is affected only by idiosyncratic returns. We shall see below that, in G8 economies, stock returns are explained by approximately $m=15$ factors (or between 10 and 20 factors), and that the the systematic component of stock returns explains approximately 50% of the variance (see Plerou $et\ al.$ (2002) and Laloux $et\ al.$ (2000)). The question is how to define "factors".

2.1 The PCA approach: can you hear the shape of the market?

A first approach for extracting factors from data is to use Principal Components Analysis (Jolliffe (2002)). This approach uses historical share-price data on a cross-section of N stocks going back, say, M days in history. For simplicity of exposition, the cross-section is assumed to be identical to the investment universe, although this need not be the case in practice.[3] Let us represent the stocks return data, on any given date t_0, going back $M+1$ days as a matrix

$$R_{ik} = \frac{S_{i(t_0-(k-1)\Delta t)} - S_{i(t_0-k\Delta t)}}{S_{i(t_0-k\Delta t)}}, \ k = 1, ..., M, \ i = 1, ..., N,$$

where S_{it} is the price of stock i at time t adjusted for dividends and $\Delta t = 1/252$. Since some stocks are more volatile than others, it is convenient to work with standardized returns,

$$Y_{ik} = \frac{R_{ik} - \overline{R}_i}{\overline{\sigma}_i}$$

where

$$\overline{R}_i = \frac{1}{M} \sum_{k=1}^{M} R_{ik}$$

and

$$\overline{\sigma}_i^2 = \frac{1}{M-1} \sum_{k=1}^{M} (R_{ik} - \overline{R}_i)^2$$

---

[3]For instance, the analysis can be restricted to the members of the S&P500 index in the US, the Eurostoxx 350 in Europe, etc.

The empirical correlation matrix of the data is defined by

$$\rho_{ij} = \frac{1}{M-1}\sum_{k=1}^{M} Y_{ik}Y_{jk}, \tag{8}$$

which is symmetric and non-negative definite. Notice that, for any index i, we have

$$\rho_{ii} = \frac{1}{M-1}\sum_{k=1}^{M}(Y_{ik})^2 = \frac{1}{M-1}\frac{\sum_{k=1}^{M}(R_{ik}-\overline{R}_i)^2}{\sigma_i^2} = 1.$$

The dimensions of ρ are typically 500 by 500, or 1000 by 1000, but the data is small relative to the number of parameters that need to be estimated. In fact, if we consider daily returns, we are faced with the problem that very long estimation windows $M \gg N$ don't make sense because they take into account the distant past which is economically irrelevant. On the other hand, if we just consider the behavior of the market over the past year, for example, then we are faced with the fact that there are considerably more entries in the correlation matrix than data points.

The commonly used solution to extract meaningful information from the data is Principal Components Analysis.[4] We consider the eigenvectors and eigenvalues of the empirical correlation matrix and rank the eigenvalues in decreasing order:

$$N \geq \lambda_1 > \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_N \geq 0.$$

We denote the corresponding eigenvectors by

$$v^{(j)} = \left(v_1^{(j)}, ..., v_N^{(j)}\right) , \quad j = 1, ..., N.$$

A cursory analysis of the eigenvalues shows that the spectrum contains a few large eigenvalues which are detached from the rest of the spectrum (see Figure 1). We can also look at the density of states

$$D(x,y) = \frac{\{\#\text{of eigenvalues between } x \text{ and } y\}}{N}$$

(see Figure 2). For intervals (x,y) near zero, the function $D(x,y)$ corresponds to the "bulk spectrum" or "noise spectrum" of the correlation matrix. The eigenvalues at the top of the spectrum which are isolated from the bulk spectrum are obviously significant. The problem that is immediately evident by looking at Figures 1 and 2 is that there are less "detached" eigenvalues than industry sectors. Therefore, we expect that the boundary between "significant" and "noise" eigenvalues to be somewhat blurred and to correspond to be at the

---

[4]We refer the reader to Laloux *et al.* (2000), and Plerou *et al.* (2002) who studied the correlation matrix of the top 500 stocks in the US in this context.
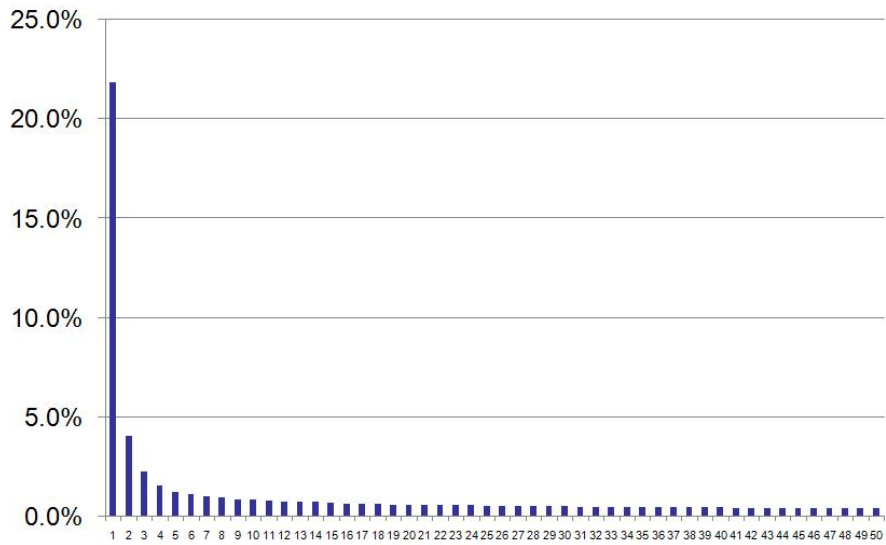
7

Figure 1: Eigenvalues of the correlation matrix of market returns computed on May 1 2007 estimated using a 1-year window (measured as percentage of explained variance)
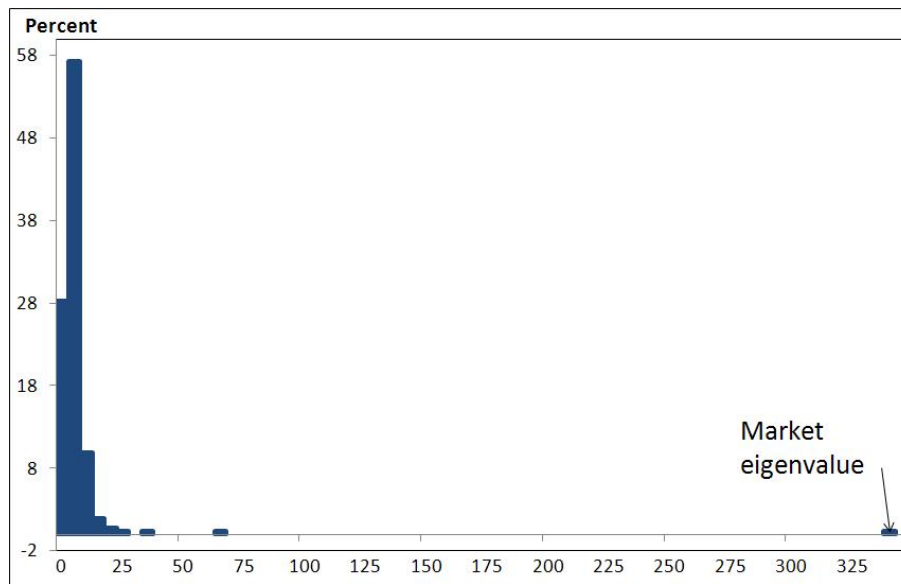

Figure 2: The density of states for May 1-2007 estimated using a year window

8

edge of the "bulk spectrum". This leads to two possibilities: (a) we take into account a fixed number of eigenvalues to extract the factors (assuming a number close to the number of industry sectors) or (b) we take a variable number of eigenvectors, depending on the estimation date, in such a way that a sum of the retained eigenvalues exceeds a given percentage of the trace of the correlation matrix. The latter condition is equivalent to saying that the truncation explains a given percentage of the total variance of the system.

Let $\lambda_1, ..., \lambda_m$, $m < N$ be the significant eigenvalues in the above sense. For each index j, we consider a the corresponding "eigenportfolio", which is such that the respective amounts invested in each of the stocks is defined as

$$Q_i^{(j)} = \frac{v_i^{(j)}}{\overline{\sigma}_i}.$$

The eigenportfolio returns are therefore

$$F_{jk} = \sum_{i=1}^{N} \frac{v_i^{(j)}}{\overline{\sigma}_i} R_{ik} \quad j = 1, 2, ..., m. \tag{9}$$

It is easy for the reader to check that the eigenportfolio returns are uncorrelated in the sense that the empirical correlation of F_j and $F_{j'}$ vanishes for $j \neq j'$. The factors in the PCA approach are the eigenportofolio returns.
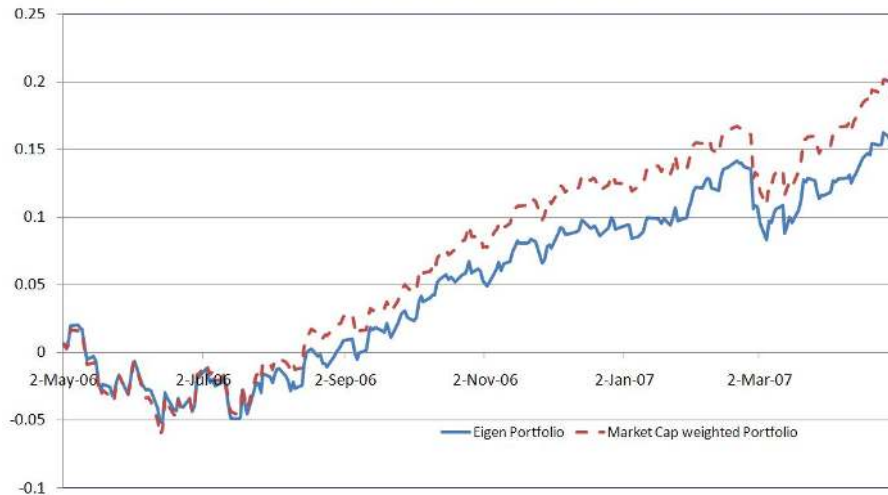

Figure 3: Comparative evolution of the principal eigenportfolio and the capitalization-weighted portfolio from May 2006 to April 2007. Both portfolios exhibit similar behavior.

Each stock return in the investment universe can be decomposed into its projection on the m factors and a residual, as in equation (4). Thus, the PCA

approach delivers a natural set of risk-factors that can be used to decompose our returns. It is not difficult to verify that this approach corresponds to modeling the correlation matrix of stock returns as a sum of a rank-m matrix corresponding to the significant spectrum and a diagonal matrix of full rank,

$$\overline{\rho}_{ij} = \sum_{k=0}^{m} \lambda_k \, v_i^{(k)} v_j^{(k)} + \epsilon_{ii}^2 \delta_{ij},$$

where δ_{ij} is the Kronecker delta and ϵ_{ii}^2 is given by

$$\epsilon_{ii}^2 = 1 - \sum_{k=0}^{m} \lambda_k \, v_i^{(k)} v_i^{(k)}$$

so that $\overline{\rho}_{ii} = 1$. This means that we keep only the significant eigenvalues/eigenvectors of the correlation matrix and add a diagonal "noise" matrix for the purposes of conserving the total variance of the system.

2.2 Interpretation of the eigenvectors/eigenportfolios

As pointed out by several authors (see for instance, Laloux et $al.$(2000)), the dominant eigenvector is associated with the "market portfolio", in the sense that all the coefficients $v_i^{(1)}$, $i = 1, 2.., N$ are positive. Thus, the eigenportfolio has positive weights $Q_i^{(1)} = \frac{v_i^{(1)}}{\overline{\sigma}_i}$. We notice that these weights are inversely proportional to the stock's volatility. This weighting is consistent with the capitalization-weighting, since larger capitalization companies tend to have smaller volatilities. The two portfolios are not identical but are good proxies for each other,[5] as shown in Figure 3.

To interpret the other eigenvectors, we observe that (i) the remaining eigenvectors must have components that are negative, in order to be orthogonal to $v^{(i)}$; (ii) given that there is no natural order in the stock universe, the "shape analysis" that is used to interpret the PCA of interest-rate curves (Litterman and Scheinkman (1991) or equity volatility surfaces (Cont and Da Fonseca (2002)) does not apply here. The analysis that we use here is inspired by Scherer and Avellaneda (2002), who analyzed the correlation of sovereign bond yields across different Latin American issuers (see also Plerou $et.$ $al.$(2002) who made similar observations). We rank the coefficients of the eigenvectors in decreasing order:

$$v_{n_1}^{(2)} \geq \, v_{n_2}^{(2)} \geq \, ... \, \geq \, v_{n_N}^{(2)},$$

the sequence n_i representing a re-labeling of the companies. In this new ordering, we notice that the "neighbors" of a particular company tend to be in the

---

[5]The positivity of the coefficients of the first eigenvector of the correlation matrix in the case when all assets have non-negative correlation follows from Krein's Theorem. In practice, the presence of commodity stocks and mining companies implies that there are always a few negatively correlated stock pairs. In particular, this explains why there are a few negative weights in the principal eigenportfolio in Figure 4.

same industry group. This property, which we call *coherence*, holds true for $v^{(2)}$ and for other high-ranking eigenvectors. As we descend in the spectrum towards the noise eigenvectors, the property that nearby coefficients correspond to firms in the same industry is less true and coherence will not hold for eigenvectors of the noise spectrum (almost by definition!). The eigenportfolios can therefore be interpreted as "pairs-trading" or, more generally, long-short positions, at the level of industries or sectors.
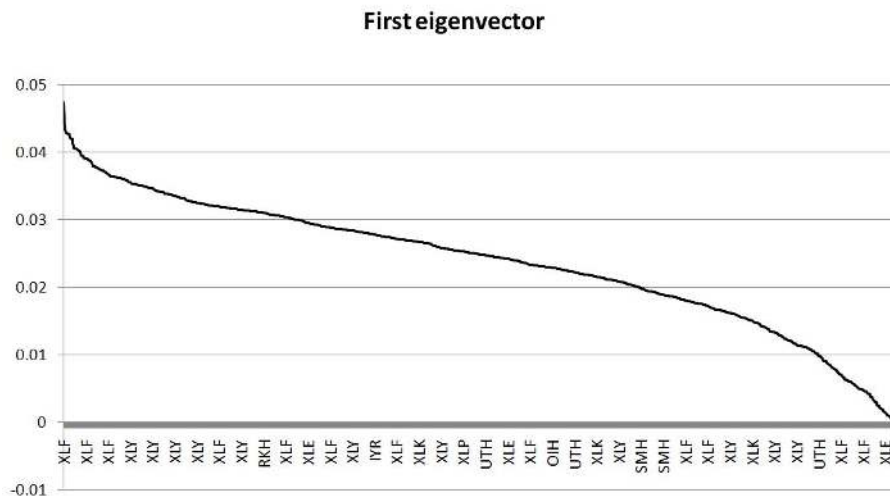
First eigenvector

Figure 4: First eigenvector sorted by coefficient size. The x-axis shows the ETF corresponding to the industry sector of each stock.

2.3 The ETF approach: using the industries

Another method consists in using the returns of sector ETFs as factors. In this approach, we select a sufficiently diverse set of ETFs and perform multiple regression analysis of stock returns on these factors. Unlike the case of eigen-portfolios, ETF returns are not uncorrelated, so there can be redundancies: strongly correlated ETFs may lead to large factor loadings with opposing signs for stocks that belong to or are strongly correlated to different ETFs. To remedy this, we can perform a robust version of multiple regression analysis to obtain the coefficients β_{ij}. For example, the *matching pursuit* algorithm (Davis, Mallat & Avellaneda (1997)) which favors sparse representations is preferable to a full multiple regression. Another class of regression methods known as *ridge regression* achieves the similar goal of sparse representations (see, for instance Jolliffe (2002)). Finally, a simple approach, which we use in our back-testing strategies, associates to each stock a sector ETF (following the partition of the market in

11

Figure 5: Second eigenvector sorted by coefficient size. Labels as in Figure 4.

Figure 6: Third eigenvector sorted by coefficient size. Labels as in Figure 4.

Top 10 Stocks	Bottom 10 Stocks
Energy, oil and gas	Real estate, financials, airlines
Suncor Energy Inc.	American Airlines
Quicksilver Res.	United Airlines
XTO Energy	Marshall & Isley
Unit Corp.	Fifth Third Bancorp
Range Resources	BBT Corp.
Apache Corp.	Continental Airlines
Schlumberger	M & T Bank
Denbury Resources Inc.	Colgate-Palmolive Company
Marathon Oil Corp.	Target Corporation
Cabot Oil & Gas Corporation	Alaska Air Group, Inc.

Table 1: The top 10 stocks and bottom 10 stocks in second eigenvector.

Top 10 Stocks	Bottom 10 Stocks
Utility	Semiconductor
Energy Corp.	Arkansas Best Corp.
FPL Group, Inc.	National Semiconductor Corp.
DTE Energy Company	Lam Research Corp.
Pinnacle West Capital Corp.	Cymer, Inc.
The Southern Company	Intersil Corp.
Consolidated Edison, Inc.	KLA-Tencor Corp.
Allegheny Energy, Inc.	Fairchild Semiconductor International
Progress Energy, Inc.	Broadcom Corp.
PG&E Corporation	Cellcom Israel Ltd.
FirstEnergy Corp.	Leggett & Platt, Inc.

Table 2: The top 10 stocks and bottom 10 stocks in third eigenvector.

Figure 7) and performs a regression of the stock returns on the corresponding ETF returns.

Let $I_1, I_2, ..., I_m$ represent a class of ETFs that span the main sectors in the economy, and let R_{I_j} denote the corresponding returns. The ETF decomposition takes the form

$$R_i \ = \ \sum_{j=1}^{m} \beta_{ij} R_{I_j} \ + \tilde{R}_i.$$

The tradeoff between the ETF method and the PCA method is that in the former we need to have some prior knowledge of the economy to know what is a "good" set of ETFs to explain returns. The advantage is that the interpretation of the factor loadings is more intuitive than for PCA. Nevertheless, based on the notion of coherence alluded to in the previous section, it could be argued that the ETF and PCA methods convey similar information. There is a caveat, however: ETF holdings give more weight to large capitalization companies, whereas PCA has no *a priori* capitalization bias. As we shall see, these nuances are reflected in the performance of statistical arbitrage strategies based on different risk-factors.

Figure 7 shows a sample of industry sectors number of stocks of companies with capitalization of more than 1 billion USD at the beginning of January 2007, classified by sectors. The table gives an idea of the dimensions of the trading universe and the distribution of stocks corresponding to each industry sector. We also include, for each industry, the ETF that can be used as a risk-factor for the stocks in the sector for the simplified model (11).

3   A relative-value model for equity pricing

We propose a quantitative approach to stock pricing based on relative performance within industry sectors or PCA factors. In the last section, we present a modification of the signals which take into account the trading volume in the stocks as well, within a similar framework. This model is purely based on price data, although in principle it could be extended to include fundamental factors, such changes in analysts' recommendations, earnings momentum, and other quantifiable factors.

We shall use continuous-time notation and denote stock prices by $S_i(t), ..., S_N(t)$, where t is time measured in years from some arbitrary starting date. Based on the multi-factor models introduced in the previous section, we assume that stock returns satisfy the system of stochastic differential equations

$$\frac{dS_i(t)}{S_i(t)} = \alpha_i \, dt + \sum_{j=1}^{N} \beta_{ij} \frac{dI_j(t)}{I_j(t)} + dX_i(t), \tag{10}$$

where the term

Sector	ETF	Num of Stocks	Market Cap (unit: 1M/usd)		
			Average	Max	Min
Internet	HHH	22	10,350	104,500	1,047
Real Estate	IYR	87	4,789	47,030	1,059
Transportation	IYT	46	4,575	49,910	1,089
Oil Exploration	OIH	42	7,059	71,660	1,010
Regional Banks	RKH	69	23,080	271,500	1,037
Retail	RTH	60	13,290	198,200	1,022
Semiconductors	SMH	55	7,303	117,300	1,033
Utility	UTH	75	7,320	41,890	1,049
Energy	XLE	75	17,800	432,200	1,035
Financial	XLF	210	9,960	187,600	1,000
Industrial	XLI	141	10,770	391,400	1,034
Technology	XLK	158	12,750	293,500	1,008
Consumer Staples	XLP	61	17,730	204,500	1,016
Healthcare	XLV	109	14,390	192,500	1,025
Consumer discretionary	XLY	207	8,204	104,500	1,007
Total		1417	11,291	432,200	1,000

January, 2007

Figure 7: Trading universe on January 1, 2007: breakdown by sectors.

15

$$\sum_{j=1}^{N} \beta_{ij} \frac{dI_j(t)}{I_j(t)}$$

represents the systematic component of returns (driven by the returns of the eigenportfolios or ETFs). To fix ideas, we place ourselves in the ETF framework. In this context, $I_j(t)$ represents the mid-market price of the jth ETF used to span the market. The coefficients β_{ij} are the corresponding loadings.

In practice, only ETFs that are in the same industry as the stock in question will have significant loadings, so we could also work with the simplified model

$$
\begin{aligned}
\beta_{ij} &= \frac{Cov(R_i, R_{I_j})}{Var(R_{I_j})} \quad \text{if stock } \#i \text{ is in industry } \#j \\
&= 0 \qquad\qquad\quad \text{otherwise}
\end{aligned}
\tag{11}
$$

where each stock is regressed to a single ETF representing its "peers".

The idiosyncratic component of the return is given by

$$\alpha_i dt + dX_i(t).$$

Here, the α_i represents the drift of the idiosyncratic component, *i.e.* $\alpha_i dt$ is the excess rate of return of the stock in relation to market or industry sector over the relevant period. The term $dX_i(t)$ is assumed to be the increment of a stationary stochastic process which models price fluctuations corresponding to over-reactions or other idiosyncratic fluctuations in the stock price which are not reflected the industry sector.

Our model assumes (i) a drift which measures systematic deviations from the sector and (ii) a price fluctuation that is mean-reverting to the overall industry level. Although this is very simplistic, the model can be tested on cross-sectional data. Using statistical testing, we can accept or reject the model for each stock in a given list and then construct a trading strategy for those stocks that appear to follow the model and yet for which significant deviations from equilibrium are observed.

Based on these considerations, we introduce a parametric model for $X_i(t)$ which can be estimated easily, namely, the Ornstein-Uhlembeck process:

$$dX_i(t) = \kappa_i \left(m_i - X_i(t) \right) dt + \sigma_i dW_i(t), \quad \kappa_i > 0. \tag{12}$$

This process is stationary and auto-regressive with lag 1 (AR-1 model). In particular, the increment $dX_i(t)$ has unconditional mean zero and conditional mean equal to

$$E\left\{ dX_i(t) | X_i(s), s \leq t \right\} = \kappa_i \left(m_i - X_i(t) \right) dt .$$

The conditional mean, or forecast of expected daily returns, is positive or negative according to the sign of $m_i - X_i(t)$.

The parameters of the stochastic differential equation, α_i, κ_i, m_i and σ_i, are specific to each stock. They are assumed to vary slowly in relation to the Brownian motion increments $dW_i(t)$, in the time-window of interest. We estimate the statistics for the residual process on a window of length 60 days, assuming that the parameters are constant over the window. This hypothesis is tested for each stock in the universe, by goodness-of-fit of the model and, in particular, by analyzing the speed of mean-reversion.

If we assume momentarily that the parameters of the model are constant, we can write

$$X_i(t_0 + \Delta t) = e^{-\kappa_i \Delta t} X_i(t_0) + \left(1 - e^{-\kappa_i \Delta t}\right) m_i + \sigma_i \int_{t_0}^{t_0 + \Delta t} e^{-\kappa_i(t_0 + \Delta t - s)} dW_i(s) .$$

(13)

Letting Δt tend to infinity, we see that equilibrium probability distribution for the process $X_i(t)$ is normal with

$$E\left\{X_i(t)\right\} = m_i \quad \text{and} \quad Var\left\{X_i(t)\right\} = \frac{\sigma_i^2}{2\kappa_i}.$$

(14)

According to Equation (10), investment in a market-neutral long-short portfolio in which the agent is long \$1 in the stock and short β_{ij} dollars in the jth ETF has an expected 1-day return

$$\alpha_i dt + \kappa_i \left(m_i - X_i(t)\right) dt .$$

The second term corresponds to the model's prediction for the return based on the position of the stationary process $X_i(t)$: it forecasts a negative return if $X_i(t)$ is sufficiently high and a positive return if $X_i(t)$ is sufficiently low.

The parameter κ_i is called the speed of mean-reversion and

$$\tau_i = \frac{1}{\kappa_i}$$

represents the characteristic time-scale for mean reversion. If $\kappa \gg 1$ the stock reverts quickly to its mean and the effect of the drift is negligible. In our strategies, and to be consistent with the estimation procedure that uses constant parameters, we are interested in stocks with fast mean-reversion, i.e. such that

$$\tau_i \ll T_1.$$

4 Signal generation

Based on this simple model, we defined several trading signals. We considered an estimation window of 60 business days i.e. $T_1 = 60/252$. This estimation window incorporates at least one earnings cycle for the company. Therefore, we
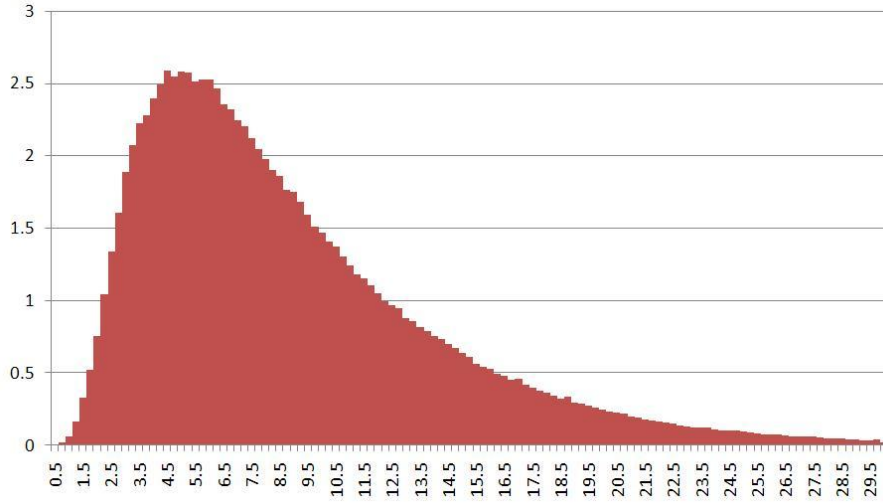
17

Figure 8: Empirical distribution of the characteristic time to mean-reversion τ_i (in business days) for the year 2007, for the stock universe under consideration. The descriptive statistics are given below.

	Days
Maximum	30
75%	11
Median	7.5
25%	4.9
Minimum	0.5
Fast days	36 %

Table 3: Descriptive statistics on the mean-reversion time τ.

ETF	Abs(Alpha)	Beta	Kappa	Reversion days	EquiVol	Abs(m)
HHH	0.20%	0.69	38	7	4%	3.3%
IYR	0.11%	0.90	39	6	2%	1.8%
IYT	0.18%	0.97	41	6	4%	3.0%
RKH	0.10%	0.98	39	6	2%	1.7%
RTH	0.17%	1.02	39	6	3%	2.7%
SMH	0.19%	1.01	40	6	4%	3.2%
UTH	0.09%	0.81	42	6	2%	1.4%
XLF	0.11%	0.83	42	6	2%	1.8%
XLI	0.15%	1.15	42	6	3%	2.4%
XLK	0.17%	1.03	42	6	3%	2.7%
XLP	0.12%	1.01	42	6	2%	2.0%
XLV	0.14%	1.05	38	7	3%	2.5%
XLY	0.16%	1.03	39	6	3%	2.5%
Total	0.15%	0.96	40	6	3%	2.4%

Figure 9: Statistical averages for the estimated OU parameters corresponding to all stocks over 2007.

expect that it reflects so some extent fluctuations in the price which take place along the cycle. We selected stocks with mean-reversion times less than $1/2$ period ($\kappa > 252/30 = 8.4$). Typical descriptive statistics for signal estimation are presented in Figure 9. For the details of the estimation of the O-U process and more statistical details on signal generation see the Appendix.

4.1   Pure mean-reversion

We focus only on the process $X_i(t)$, neglecting the drift α_i. We know that the equilibrium variance is

$$\sigma_{eq,i} \;=\; \frac{\sigma_i}{\sqrt{2\kappa_i}} \;=\; \sigma_i \sqrt{\frac{\tau_i}{2}}$$

Accordingly, we define the dimensionless variable

$$s_i = \frac{X_i(t) - m_i}{\sigma_{eq,i}} \; . \tag{15}$$

We call this variable the *s-score*.[6] See Figure 11 for a graph showing the evolution of the s-score for residuals of JPM against the Financial SPDR, XLF. The s-score measures the distance to equilibrium of the cointegrated residual in units standard deviations, i.e. how far away a given stock is from the theoretical equilibrium value associated with our model.

Our basic trading signal based on mean-reversion is

$$
\begin{aligned}
\text{buy to open if } \; s_i \;\; &< \;\; -\overline{s}_{bo} \\
\text{sell to open if } \; s_i \;\; &> \;\; +\overline{s}_{so} \\[2mm]
\text{close short position if } \; s_i \;\; &< \;\; +\overline{s}_{bc} \\
\text{close long position } \;\; s_i \;\; &> \;\; -\overline{s}_{sc}
\end{aligned}
$$

$$\tag{16}$$

where the cutoff values are determined empirically. Entering a trade, *e.g.* buy to open, means buying one dollar of the corresponding stock and selling β_i dollars of its sector ETF or, in the case of using multiple factors, β_{i1} dollars of ETF #1, β_{i2} dollars of ETF #2, ..., β_{im} dollars of ETF #m. Similarly, closing a long position means selling stock and buying ETFs.

Since we expressed all quantities in dimensionless variables, we expect the cutoffs $\overline{s}_{bo}, \overline{s}_{bo}, \overline{s}_{bc}, \overline{s}_{sc}$ to be valid across the different stocks. We selected the cutoffs empirically, based on simulating strategies from 2000 to 2004 in the case of ETF factors. Based on this analysis, we found that a good choice of cutoffs is

---

[6]See the Appendix for practical details on estimating the s-score.

$$\bar{s}_{bo} = \bar{s}_{so} = 1.25$$
$$\bar{s}_{bc} = 0.75 \quad \text{and} \quad \bar{s}_{sc} = 0.50$$

Thus, we enter trades when the s-score exceeds 1.25 in absolute value. Closing short trades sooner (at 0.75) gives slightly better results than 0.50. For closing long trades, we choose 0.50. (see Figure 10)
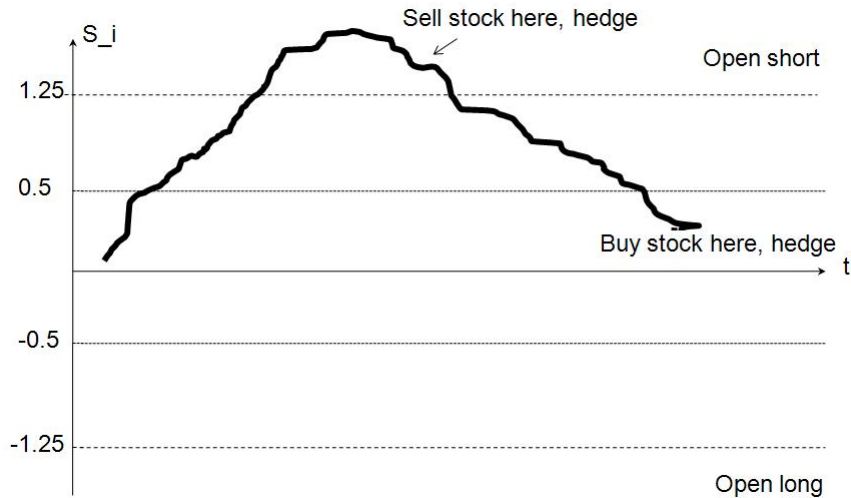

Figure 10: Schematic evolution of the s-score and the associated signal, or trading rule.

The rationale for opening trades only when the s-score s_i is far from equilibrium is to trade only when we think that we detected an anomalous excursion of the co-integration residual. We then need to consider when we close trades. Closing trades when the s-score is near zero also makes sense, since we expect most stocks to be near equilibrium most of the time. Thus, our trading rule detects stocks with large "excursions" and trades assuming these excursions will revert to the mean in a period of the order of the mean-reversion time τ_i.

4.2 Mean-reversion with drift

In the previous signal, the presence of the drift was ignored (implicitly it was assumed that the effect of the drift was irrelevant in comparison with mean-reversion). We incorporate the drift by considering the conditional expectation of the residual return over a period of time dt, namely,
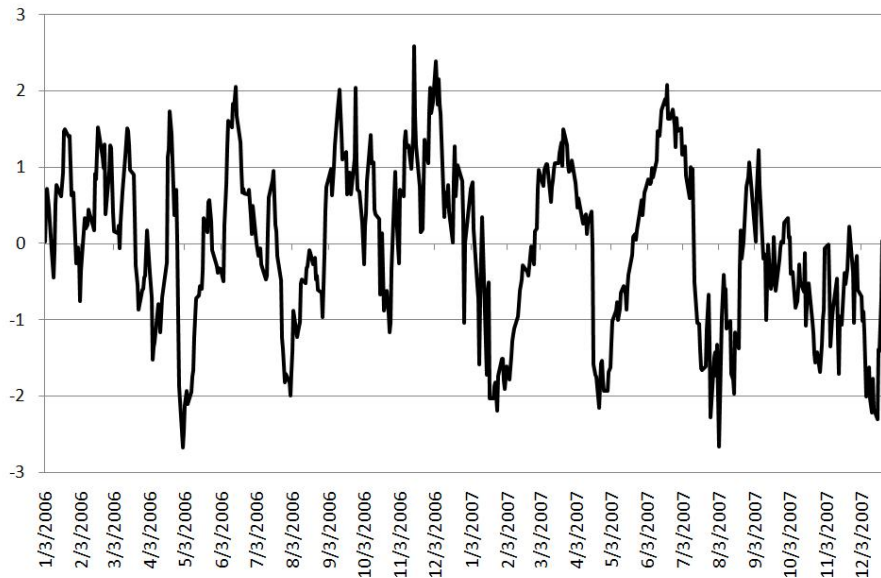
Figure 11: Evolution of the s-score of JPM (vs. XLF) from January 2006 to December 2007.

$$
\begin{aligned}
\alpha_i \, dt \; + \; \kappa_i(m_i - X_i(t)) \, dt \;\; &= \;\; \kappa_i \, \left(\frac{\alpha_i}{\kappa_i} + m_i - X_i(t) \right) dt \\
&= \;\; \kappa_i \, \left(\frac{\alpha_i}{\kappa_i} - \sigma_{eq,i} \, s_i \right) dt .
\end{aligned}
$$

This suggests that the dimensionless decision variable is the "modified s-score" (see Figure 12)

$$
s_{mod,i} \;\; = \;\; s_i - \frac{\alpha_i}{\kappa_i \, \sigma_{eq,i}} = s_i - \frac{\alpha_i \tau_i}{\sigma_{eq,i}} . \tag{17}
$$

To make contact with the analysis of the pure mean-reversion strategy, consider for example the case of shorting stock. In the previous framework, we short stock if the s-score is large enough. The modified s-score is larger if α_i is negative, and smaller if α_i is positive. Therefore, it will be harder to generate a short signal if we think that the residual has an upward drift and easier to short if we think that the residual has a downward drift. If the s-score is zero, the signal reduces to buying when the drift is high enough and selling when the drift is low. Since the drift can be interpreted as the slope of a 60-day moving average, we have therefore a "built-in" momentum strategy in this second signal. A calibration exercise using the training period 2000-2004 showed that

22
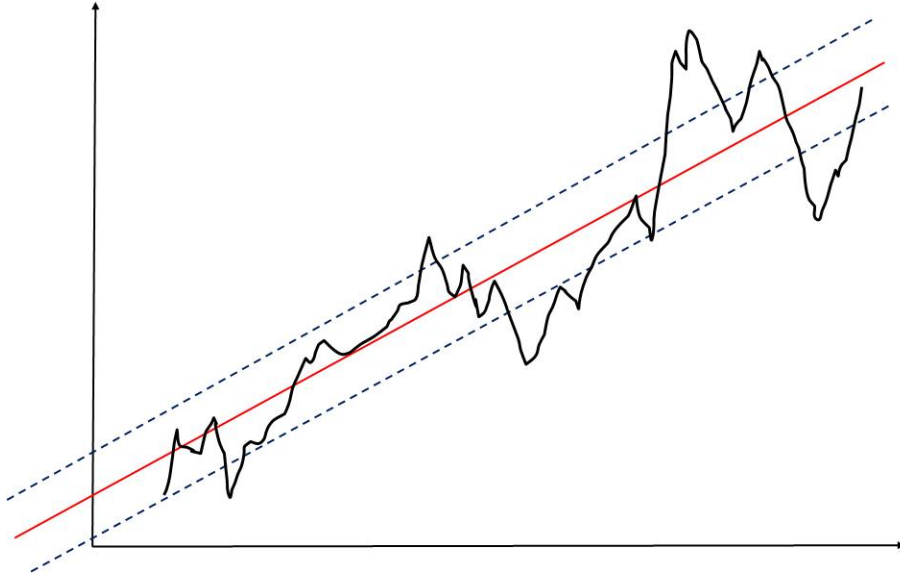
Figure 12: Including the drift in signal generation

the cutoffs defined in the previous strategy are also acceptable for this one. We notice, however, that the drift parameter has values of the order of 15 basis points and the average expected reversion time is 7 days, whereas the equilibrium volatility of residuals is on the order of 300 bps. The expected average shift for the modified s-score is of the order of $0.15 \times 7/300 \approx 0.3$. In practice, the effect of incorporating a drift in these time-scales is minor.[7]

5   Back-testing results

The back-testing experiments consisted in running the signals through historical data, with the estimation of parameters (betas, residuals), signal evaluations and portfolio re-balancing performed daily. We assumed that all trades are done at the closing price of that day. As mentioned previously, we assumed a round-trip transaction cost per trade of 10 basis points, to incorporate an estimate of price slippage and other costs as a single friction coefficient.

Let E_t represent the portfolio equity at time t. The basic PNL equation for the strategy has the following form:

---

[7]Back-testing shows that this is indeed the case.We shall not present back-testing results with the modified s-scores for the sake of brevity.

$$E_{t+\Delta t} \;=\; E_t + E_t\, r\,\Delta t + \sum_{i=1}^{N} Q_{it}\, R_{it} - \left(\sum_{i=1}^{N} Q_{it}\right) r\,\Delta t$$

$$+ \;\; \sum_{i=1}^{N} Q_{it}\, D_{it}/S_{it} \;-\; \sum_{i=1}^{N} |Q_{i\,(t+\Delta t)} - Q_{it}|\,\epsilon\,,$$

$$Q_{it} \;=\; E_t\,\Lambda_t,$$

where Q_{it} is the dollar investment in stock i at time t, R_{it} is the stock return from corresponding to the period $(t, t+\Delta t)$, r represents the interest rate (assuming, for simplicity, no spread between long and short rates), $\Delta t = 1/252$, D_{it} is the dividend payable to holders of stock i over the period $(t, t+\Delta t)$(when t=ex-dividend date), S_{it} is the price of stock i at time t, and $\epsilon = 0.0005$ is the slippage term alluded to above. The last line in the equation states that the money invested in stock i is proportional to the total equity in the portfolio. The proportionality factor, Λ_t, is stock-independent and chosen so that the portfolio has a desired level of leverage on average. For example, if we have 100 stocks long and 100 short and we wish to have a "2+2" leverage, then $\Lambda_t = 2/100$. In practice this number is adjusted only for new positions, so as not to incur transaction costs for stock which are already held in the portfolio. [8] In other words, Λ_t controls the maximum fraction of the equity that can be invested in any stock, and we take this bound to be equal for all stocks. In practice, especially when dealing with ETFs as risk-factors, we modulated the leverage coefficient on a sector-by-sector basis.[9]

Given the discrete nature of the signals, the investment strategy that we propose is "bang-bang": there is no continuous trading. Instead, the full amount is invested on the stock once the signal is active (buy-to-open, short-to-open) and the position is unwound when the s-score indicates a closing signal. This all-or-nothing strategy, which might seem inefficient at first glance, turns out to outperform making continuous portfolio adjustments.

5.1 Synthetic ETFs as factors

The first set of experiments were done using 15 *synthetic capitalization-weighted industry-sector indices* as risk-factors (see Figure 7). The reason for using synthetic ETFs was to be able to back-test strategies going back to 1996, when most ETFs did not exist. A series of daily returns for a synthetic index is calculated for each sector and recorded for the 60 days preceding the estimation date. We performed a regression of stock returns on the corresponding sector index and extracted the corresponding residual series. To ensure market-neutrality,

---

[8]Hence, strictly speaking, the leverage factor is weakly dependent on the available signals.

[9]Other refinements that can be made have to do with using different leverage according to the company's market capitalization or choosing a sector-dependent leverage that is inversely proportional to the average volatility of the sector.

we added to the portfolio an S&P 500 index futures hedge which was adjusted daily and kept the overall portfolio beta-neutral.

Since we expect that, in aggregate, stocks are correctly priced, we experimented with adjusting the means of the OU processes so that the total mean would be zero. In other words, we introduced the adjusted means for the residuals

$$\overline{m}_i \ = \ m_i - \frac{1}{N}\sum_{j=1}^{N} m_j, i = 1, 2, ..., N. \tag{18}$$

This modification has the effect of removing "model bias" and is consistent with market-neutrality. We obtained consistently better results in back-testing than when using the estimated m_i and adopted it for all other strategies as well. The results of back-testing with synthetic ETFs are shown in Figure 13 and Table 14.
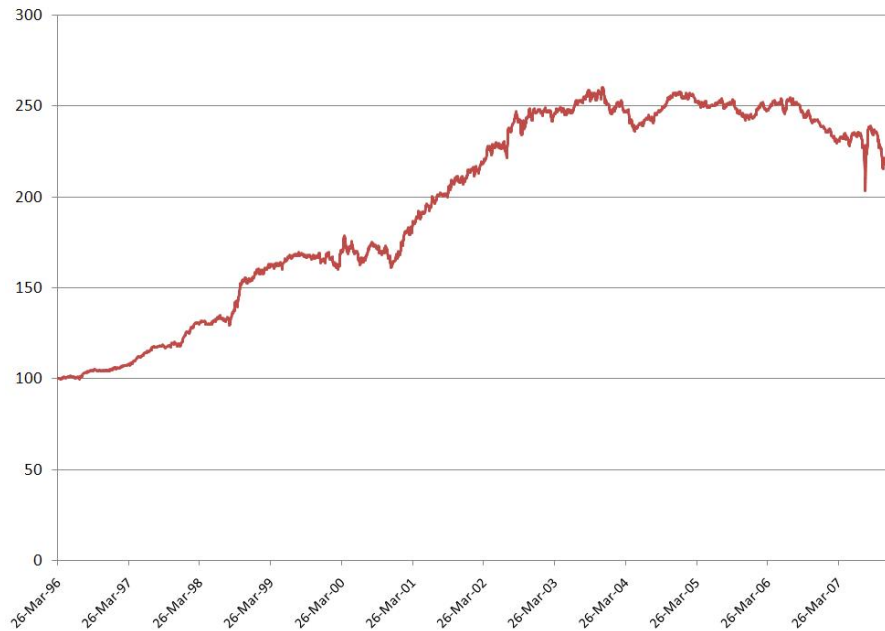
Figure 13: Historical PNL for the strategy using synthetic ETFs as factors from 1996-2007

5.2 Actual ETFs

Back-testing with actual ETFs was possible only from 2002 onward, due to the fact that many ETFs did not exist previously. We simulated strategies going

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
1996	1.7	1.7	(1.2)	1.0	0.8	(0.6)	0.6	1.4	0.6	2.3	0.5	1.5	(0.5)	1.1	0.4	1.7
1997	0.1	1.5	(0.0)	2.5	1.2	1.1	2.2	1.1	(1.0)	2.3	0.6	1.1	0.4	1.5	2.9	3.6
1998	0.9	(0.5)	(0.5)	0.8	2.5	1.8	2.4	2.0	1.1	2.1	0.8	3.0	0.1	(0.1)	2.8	3.4
1999	(1.0)	(1.3)	1.5	(1.3)	(0.7)	0.3	1.2	(1.2)	1.4	1.9	1.1	1.9	(1.1)	0.1	0.6	0.8
2000	(0.4)	1.0	1.2	(0.6)	2.1	0.1	(0.7)	0.7	1.0	0.2	(0.8)	0.9	0.1	(0.5)	(1.1)	0.3
2001	(0.9)	2.8	0.7	0.6	2.7	1.5	(0.9)	0.6	1.6	0.1	1.9	1.9	0.6	1.4	3.3	2.9
2002	1.9	1.5	(0.1)	1.0	2.1	0.7	(0.5)	(1.1)	(1.3)	1.6	0.8	2.0	1.3	0.0	1.8	2.0
2003	0.5	0.0	(0.4)	(0.4)	2.6	1.3	(1.3)	(0.9)	0.1	(0.4)	(0.8)	2.5	(0.6)	(1.0)	(1.1)	0.1
2004	0.7	0.1	1.2	0.3	1.3	(0.4)	0.1	(1.1)	0.6	0.1	1.1	1.2	(0.0)	(0.8)	(0.0)	0.8
2005	0.1	(2.1)	(0.3)	(0.8)	(0.1)	0.2	0.5	(2.1)	0.0	(0.8)	(0.1)	1.0	(1.1)	(0.6)	(0.5)	(1.3)
2006	(0.7)	(1.8)	(0.1)	(0.3)	1.6	(0.4)	(0.2)	0.3	(0.7)	(1.1)	0.9	0.7	(0.9)	(1.0)	1.1	(0.5)
2007	2.1	(2.1)	0.6	(1.4)	(1.1)	(0.9)	0.1	(1.1)	(0.8)	(1.0)	1.0	(0.0)	0.0	(0.6)	1.1	(0.5)
Since inception	0.4	0.1	0.2	0.1	1.2	0.4	0.3	(0.1)	0.2	0.6	0.6	1.5	(0.2)	(0.0)	0.9	1.1

Figure 14: Sharpe ratios for the strategy using synthetic ETFs as factors : 1996-2007. The Sharpe Ratio is defined as $\mu - r/\sigma$, where μ, r, σ are the annualized return, interest rate and standard deviation of the PNL.

back to 2002, using regression on a single ETF to generate residuals. The results
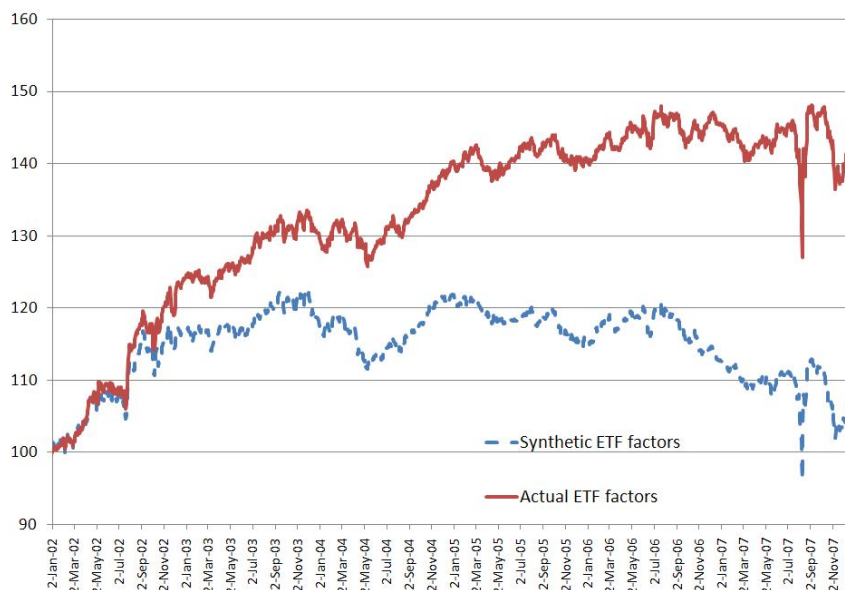are displayed on Figure 15 and Table 16.

Figure 15: Historical PNL for the strategy using actual ETFs as factors, com-
pared with the one using synthetic ETFs : 2002-2007. Notice the strong out-
performance by the strategy which uses actual ETFs.

We observe that using actual ETFs improved performance considerably. An
argument that might explain this improvement is that ETFs are traded, whereas
the synthetic ETFs are not, therefore providing better price information.

5.3   PCA with 15 eigenportfolios

The back-testing results for signals generated with 15 PCA factors are shown
in Figures 17 and 18 and Table 19. We observe that the 15-PCA strategy
out-performs the actual ETF strategy after 2002.

5.4   Using a variable number of PCA factors

We also tested strategies based on a variable number of factors, chosen so as
to explain a given level of variance. In this approach, we retain a number of
eigen-portfolios (factors) such that the sum of the corresponding eigenvectors is
equal to a set percentage. The number of eigenvalues (or eigenvectors) which
are needed to explain 55% of the total variance of the correlation matrix varies

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
2002	1.9	2.1	1.4	0.6	2.4	1.5	(0.7)	(0.2)	(0.2)	1.8	0.7	1.5	1.8	(0.1)	2.4	2.7
2003	(0.2)	0.8	(0.3)	(0.5)	1.4	1.1	(1.0)	(0.1)	0.5	0.6	(0.6)	2.6	0.3	(0.4)	(0.4)	0.8
2004	0.9	1.6	(0.7)	0.4	0.5	0.1	0.2	(0.4)	0.6	0.6	1.4	1.9	0.5	(0.6)	0.3	1.6
2005	0.3	(1.5)	0.8	(0.6)	0.3	0.5	0.5	(1.1)	(0.1)	0.9	0.6	1.3	(0.7)	0.2	0.0	0.1
2006	(0.2)	(1.3)	0.0	(0.2)	0.9	(0.1)	0.5	1.7	(0.5)	(0.6)	1.7	1.7	(0.0)	(0.4)	2.0	0.7
2007	(0.4)	(0.3)	0.0	(1.3)	(1.2)	(0.7)	0.9	(0.7)	(1.0)	(0.6)	1.1	0.6	0.4	(0.5)	1.3	(0.2)
Since inception	0.4	0.2	0.2	(0.3)	0.7	0.4	0.1	(0.1)	(0.1)	0.5	0.8	1.6	0.4	(0.3)	0.9	0.9

Figure 16: Sharpe ratios for actual 15 ETFs as factors : 2002-2007. We observe, for the purpose of comparison, that the average Sharpe ratio from 2003 to 2007 was 0.6.

28

Figure 17: PNL corresponding 15 PCA factors, compared with synthetic ETFs from 1997-2007

across time. This variability is displayed in Figure 20 and Figure 21. We also looked at other cutoffs and report similar results in Figure 24. The periods over which the number of eigenvectors needed to explain a given level of variance is small, appear to be those when the risk-premium for equities is relatively high. For instance, the latter parts of 2002 and 2007, which correspond respectively the aftermath of the Internet bubble and the bursting of the subprime bubble, are periods for which the variance is concentrated on a few top eigenvectors/eigenvalues. In contrast, 2004-2006 is a period where the variance is distributed across a much larger set of modes.

Back-testing the strategy with 55% explained variance shows that it is comparable but slighly inferior to taking a fixed number of eigenvectors (see Figure 22 and Table 23). In the same vein, we studied the performance of other strategies with a variable number of PCA eigenportfolios explaining different levels of variance. In Table 27 and Figure 25, we display the performances of strategies using 45%, 55% and 65% compared with the PCA strategies with 1 eigen portfolio and with 15 eigenportfolios. The conclusion is that 55% PCA is the best performing among the three strategies and is comparable, but slightly inferior, to the 15 PCA strategy. We also observed that taking a high cutoff such as 75% of explained variance leads to steady losses, probably due to the fact that transaction costs dominate the small residual noise that remains in the system after 'defactoring' (see Figure 26). Similarly, on the opposite side of the spectrum,

Figure 18: Comparison of strategies with 15 PCA factors and the using actual ETFs in the period 2002-2007. 15-PCA outperforms significantly the ETF strategy.

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
1997	1.1	0.8	1.1	(0.2)	1.6	(0.4)	0.7	0.4	0.5	1.2	(0.2)	2.0	(0.2)	1.2	(0.0)	1.4
1998	0.9	(0.4)	(0.3)	0.5	0.2	0.6	1.0	(0.6)	(0.3)	0.4	1.1	3.7	(0.9)	0.6	1.0	1.4
1999	(1.2)	(0.5)	0.0	(0.1)	(0.7)	0.5	1.9	(3.4)	(0.8)	1.2	0.5	2.1	0.5	(0.6)	0.7	0.2
2000	(0.1)	1.0	1.3	(0.5)	2.0	0.5	1.0	1.1	(0.2)	0.7	0.6	(0.3)	1.1	(0.3)	1.1	2.2
2001	1.7	2.1	(0.5)	(0.8)	2.6	0.2	(0.4)	(0.3)	(0.8)	0.7	2.8	2.4	(0.3)	(0.2)	3.1	2.6
2002	1.2	1.4	0.2	1.9	2.4	0.9	(0.3)	0.4	1.0	1.5	2.0	1.9	1.6	(0.2)	2.5	3.4
2003	(0.6)	0.8	1.3	(0.3)	1.7	1.0	(1.4)	(1.0)	0.5	0.5	0.1	1.7	(0.8)	1.2	1.7	0.9
2004	0.8	1.8	2.0	0.9	0.6	0.3	(0.4)	0.6	0.3	0.5	1.4	1.7	(0.8)	0.7	2.0	2.2
2005	(0.0)	(0.8)	0.0	1.5	1.0	0.6	1.2	0.2	1.3	1.0	1.9	(0.3)	0.0	(0.1)	(1.3)	1.2
2006	(0.0)	1.2	(1.1)	0.4	1.2	0.6	(0.6)	1.2	(0.3)	(0.2)	2.0	1.7	(0.9)	(1.0)	0.4	1.0
2007	0.3	(0.7)	(0.1)	1.4	(1.8)	(1.3)	(0.9)	0.7	0.1	(1.5)	1.7	1.3	1.2	0.6	(1.6)	(0.7)
Since inception	0.37	0.61	0.35	0.42	0.98	0.31	0.16	(0.07)	0.13	0.54	1.26	1.62	0.04	0.18	0.86	1.44

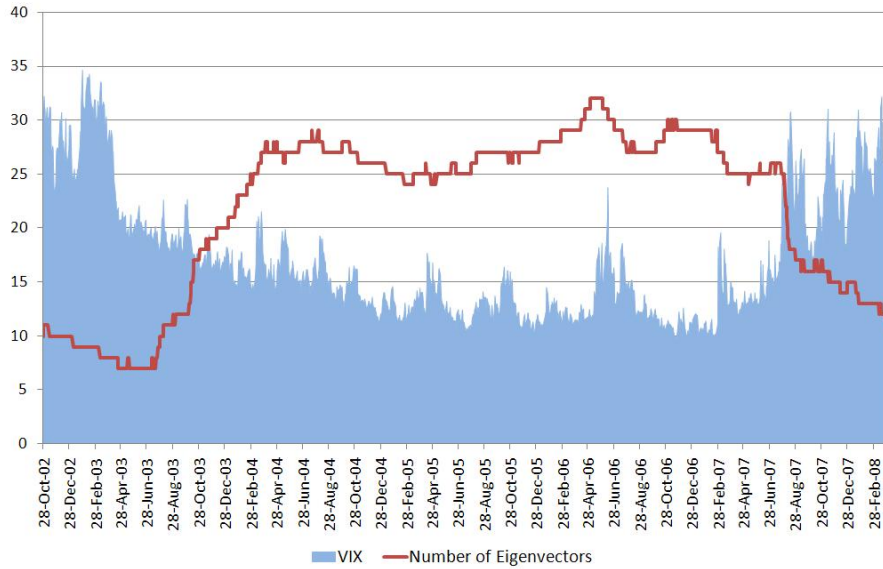Figure 19: Sharpe ratios for 15 PCA factors : 2002-2007

Figure 20: Number of significant eigenvectors needed to explain the variance of the correlation matrix at the 55% level, from 2002 to February 2008. The estimation window for the correlation matrix is 252 days. The boundary of the shaded region represents the VIX CBOT Volatility Index (measured in percentage points).
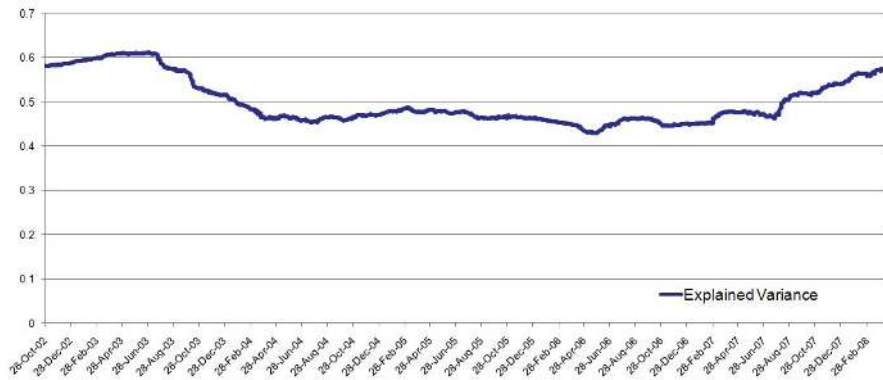

Figure 21: Percentage of variance explained by the top 15 eigenvectors: 2002-February 2008. Notice the increase in the Summer of 2007.
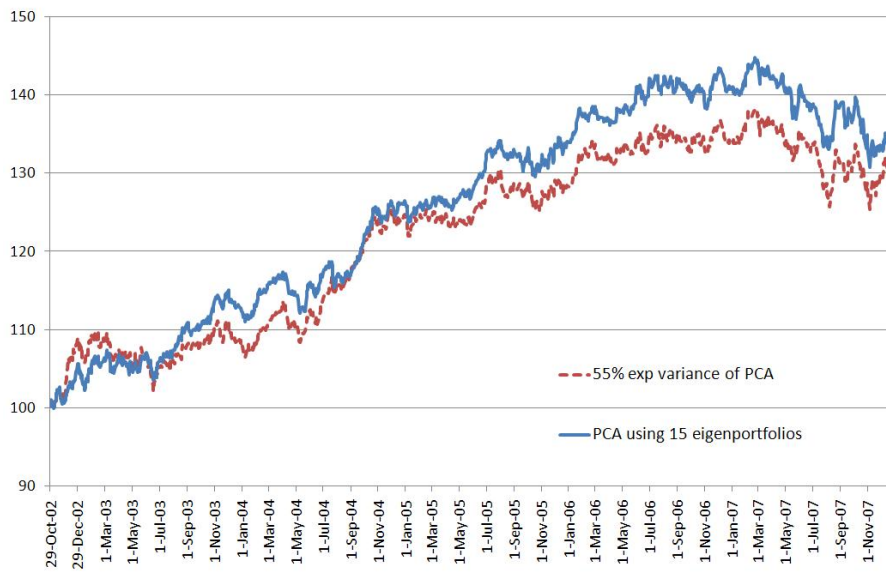
Figure 22: Comparison of the PNLs for the fixed explained variance (55%) of PCA and the 15 PCA strategy: 2002-2007. The performance of the 15 PCA strategy is slightly superior.

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
2003	(1.3)	0.4	(0.8)	(0.4)	1.4	1.4	(1.4)	(1.3)	(0.2)	0.3	(0.6)	2.5	(1.0)	(0.2)	1.1	(0.1)
2004	1.2	0.9	0.6	0.4	0.6	(0.3)	0.5	0.4	1.1	1.2	2.1	2.4	(0.3)	0.7	2.4	2.6
2005	(0.3)	(1.4)	1.1	1.6	(0.6)	(0.3)	1.3	0.2	1.1	1.1	1.6	0.3	(0.6)	0.5	(0.8)	0.6
2006	0.0	0.8	(0.9)	0.2	0.9	0.4	(0.3)	1.2	(0.4)	(0.8)	1.3	0.5	(0.5)	(0.1)	1.2	0.9
2007	1.8	(0.9)	(0.3)	0.6	(2.0)	(0.3)	0.5	1.0	0.3	(1.7)	2.6	0.8	0.2	0.3	(1.0)	(0.4)
Since inception	0.3	(0.0)	(0.1)	0.5	0.1	0.2	0.1	0.3	0.4	(0.0)	1.4	1.3	(0.4)	0.2	0.6	0.7

Figure 23: Sharpe ratios for the fixed explained variance (55%) of PCA : 2002-2007

using just one eigen-portfolio, as in the Capital Asset Pricing Model, gives rise low levels of mean-reversion, higher residual volatility and poor Sharpe ratios. (See Figure (25)).
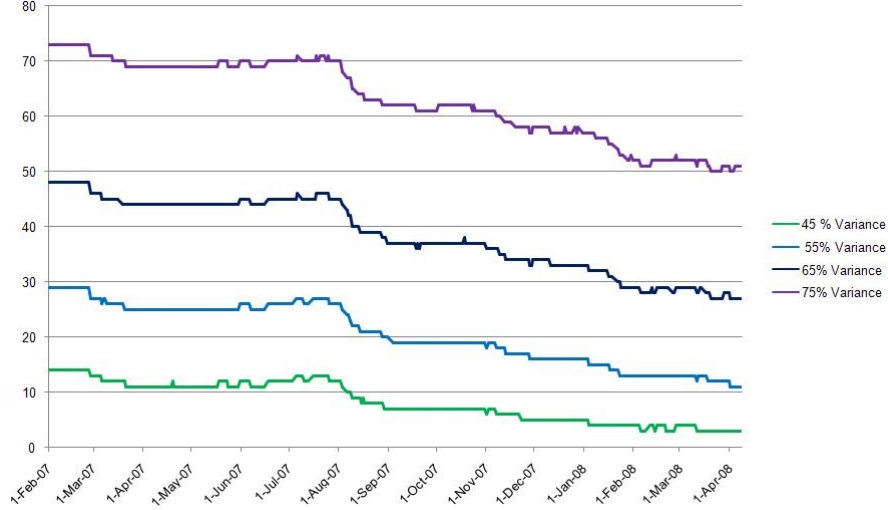

Figure 24: Time-evolution of number of PCA factors for different levels of explained variance: 2002-2007

6   Taking trading volume into account

In this section, we add volume information to the mean-reversion signals. Let V_t represent the cumulative share volume transacted until time t starting from an arbitrary reference time t_0 (say, the date at which the stock was first issued). This is an increasing function which can be viewed as a sum of daily trading volumes and approximated as an integral:

$$V_t = \sum \delta V_k \approx \int_{t_0}^{t} \dot{V}_s \, ds.$$

Historical prices can be viewed on a uniform "time grid" or on a uniform "volume grid" (i.e. the price evolution each time one share is traded). If we denote the latter prices by P_V, we have

$$
\begin{aligned}
S_{t+\Delta t} - S_t &= P_{V(t+\Delta t)} - P_{V(t)} \\
&= \frac{P_{V(t+\Delta t)} - P_{V(t)}}{V(t+\Delta t) - V(t)} \left(V(t + \Delta t) - V(t) \right). \quad (19)
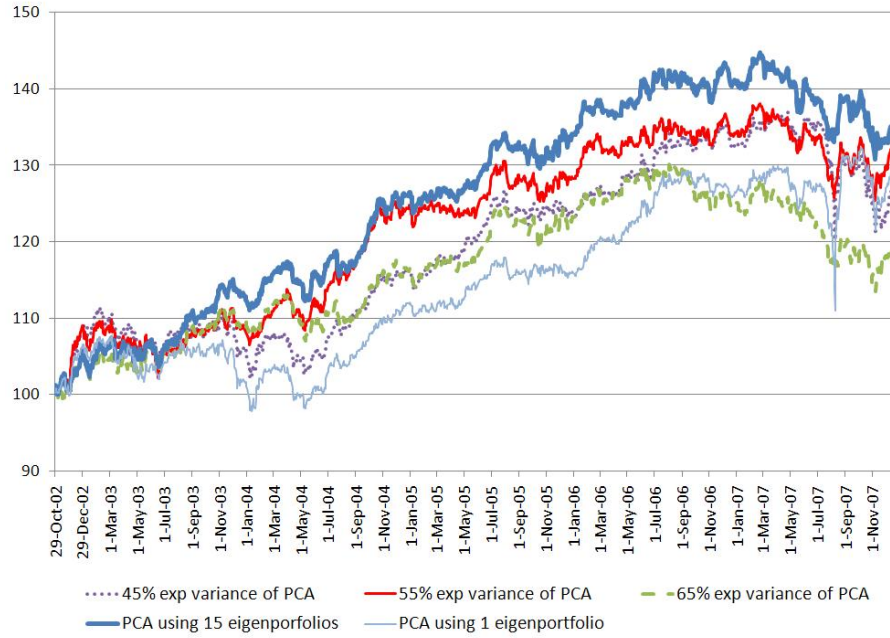\end{aligned}
$$

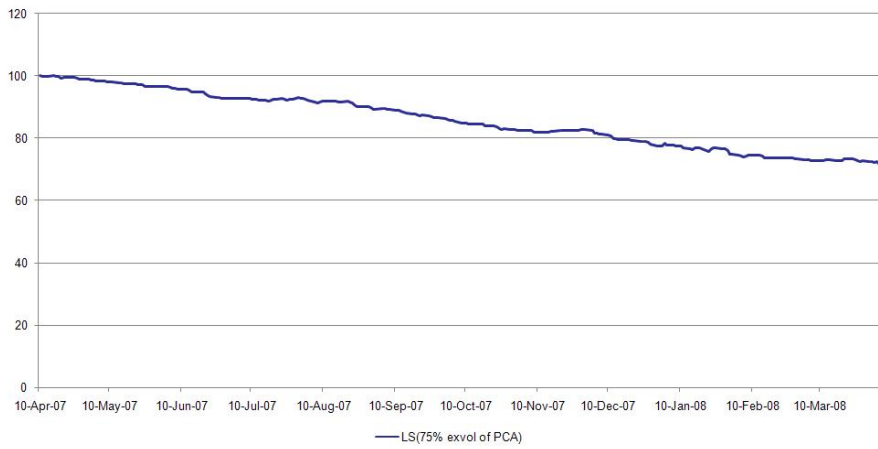Figure 25: PNL for different variance truncation level:2002-2007

Figure 26: Truncation at 75 % of explained variance: 2007- Apr 2008

	1 Eigenportfolios	15 Eigenportfolios	45 % Exp Variance	55% Exp Variance	65% Exp Variance
2003	-0.7	0.9	-0.5	-0.1	0.4
2004	1.7	2.2	1.7	2.6	1.3
2005	0.8	1.2	1.3	0.6	1.0
2006	1.8	1.0	1.3	0.9	0.3
2007	0.0	-0.7	-0.7	-0.4	-0.9
Since Inception	0.7	0.9	0.6	0.7	0.4

Figure 27: Sharpe ratios for variable PCA strategies: 2002- 2007

Thus, the average price change per unit share over the period of interest is

$$\frac{P_{V(t+\Delta t)} - P_{V(t)}}{V(t + \Delta t) - V(t)} = \frac{S_{t+\Delta t} - S_t}{V(t + \Delta t) - V(t)}.$$

This suggests that, instead of the classical daily stock returns, we use the modified returns

$$\overline{R}_t = \frac{S_{t+\Delta t} - S_t}{S_t} \frac{\langle \delta V \rangle}{V(t + \Delta t) - V(t)} = R_t \left(\frac{\langle \delta V \rangle}{V(t + \Delta t) - V(t)} \right) \qquad (20)$$

where $\langle \delta V \rangle$ indicates the average, or typical, daily trading volume calculated over a given trailing window. Measuring mean-reversion in trading time is equivalent to using calendar time and 'weighting' the stock returns as in (20).

The modified returns $\overline{R}_t$ are equal to the classical returns if the daily trading volume is typical. Notice that if the trading volume is low, the the factor on the right-hand side of the last equation is larger than unity and $\overline{R}_t > R_t$. Similarly, if volume is high then $\overline{R}_t < R_t$. The concrete effect of the trading-time modification is that mean-reversion strategies are sensitive to how much trading was done immediately before the signal was triggered. If the stock rallies on high volume, an open-to-short signal using classical returns may be triggered. However, if the volume is sufficiently large, then the modified return is much smaller so the residual will not necessarily indicate a shorting signal. Similarly, buying stocks that drop on high volume is discouraged by the trading-time approach.

We back-tested the previous strategies using the trading time approach and found that this technique increases the PNL and the Sharpe ratios unequivocally for stategies with ETF-generated signals (see Figure 28 and Table 29). For PCA-based strategies, we found that the trading time framework does not lead to a significant improvement. Finally, we find that the ETF strategy using trading time is comparable in performance to the 15-PCA/55% PCA strategies (with or

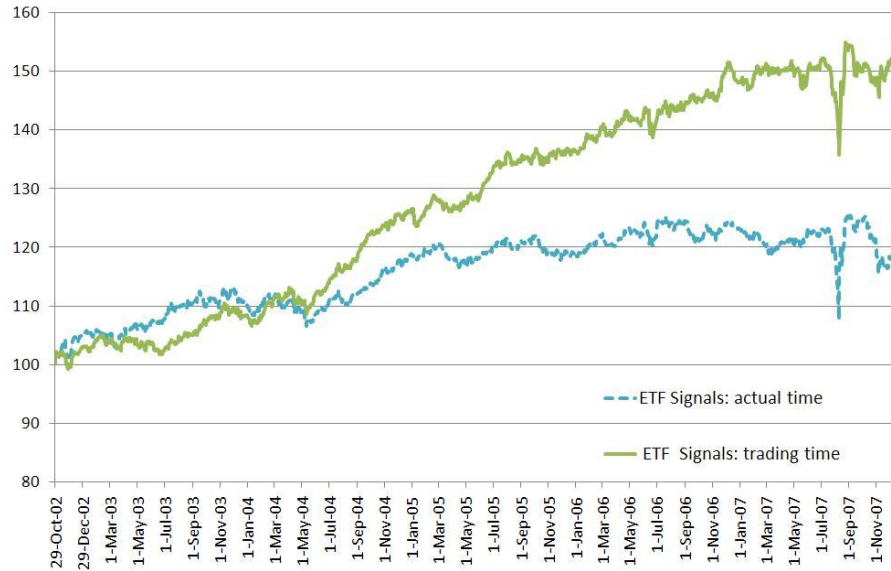without trading time adjustments) (see Figure 30 and Table 31 and also Figure 32).

Figure 28: Comparision of signals in trading time vs. actual time using actual ETFs as factors : 2002-2007

7   A closer look at 2007

It has been widely reported in the media that 2007 was very challenging for quantitative hedge funds; see Khandani and Lo (2007), Barr (2007), Associated Press (2007), Rusli (2007). This was particularly true of statistical arbitrage strategies, who experienced a large drawdown and subsequent partial recovery in the second week of August 2007. Unfortunately for many managers, the size of the drawdown was such that many had to de-leverage their portfolios and did not recover to pre-August levels. Our backtesting results are consistent with the real-world events of 2007 and show a strong drawdown in August 2007 (see below). This drawdown was first reproduced in back-testing by Khandani and Lo(2007) using contrarian strategies.

We analyzed the performance for our stategies in 2007 using ETFs with and without trading time adjustment as well as the 15-PCA strategy (see Figure 33). First, we found that performance was flat or slightly negative in the first part of the year. In early August, we found that mean-reversion strategies experienced a large, sudden drawdown followed by a recovery in about 10 days. In certain cases, our strategies tracked almost identically the Khandani-Lo (2007)

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
2003	0.3	1.4	(1.4)	0.1	(0.1)	0.9	(1.4)	0.1	0.0	(0.8)	1.1	2.1	(0.1)	0.0	1.0	0.9
2004	0.9	2.4	0.2	0.9	1.7	1.8	0.1	0.2	1.8	1.4	1.0	1.6	0.4	0.5	0.4	3.1
2005	0.1	0.1	1.3	0.6	(0.2)	(0.2)	0.8	0.4	1.2	1.5	0.7	2.6	(1.4)	2.1	(0.2)	1.6
2006	2.4	1.6	(1.5)	0.3	1.1	(0.6)	(0.5)	1.3	0.7	0.4	1.6	1.1	(0.5)	(0.0)	1.9	1.5
2007	1.7	(2.4)	0.4	0.8	(1.0)	(0.4)	0.4	0.7	0.7	(1.2)	1.9	2.0	0.3	1.8	0.7	0.4
Since 2003	1.08	0.61	(0.21)	0.52	0.30	0.29	(0.12)	0.54	0.88	0.28	1.26	1.91	(0.28)	0.87	0.76	1.51

Figure 29: Sharpe ratios for signals in trading time using actual ETFs as factors : 2002-2007
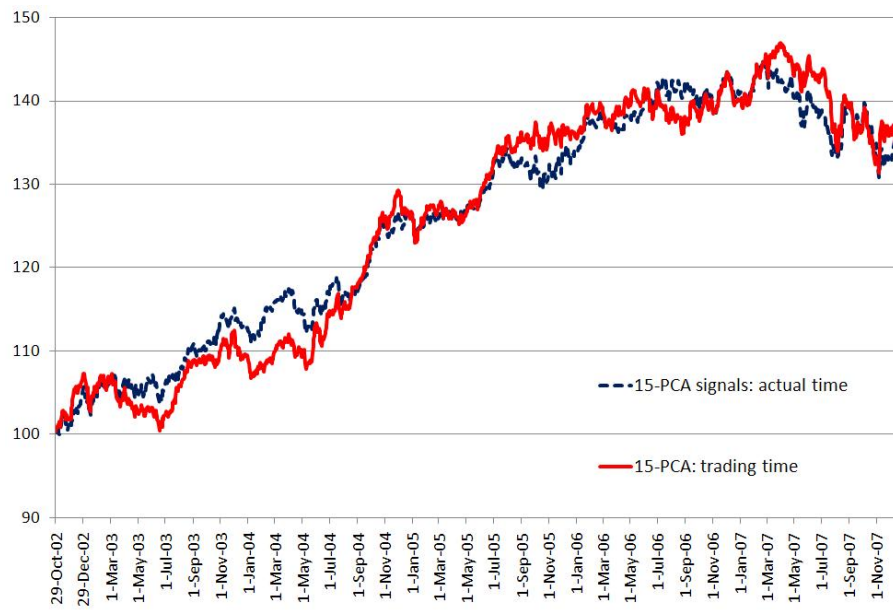

Figure 30: Comparision of signals in trading time vs. actual time using 15 PCAs as factors : 2002-2007

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfolio
2003	0.9	(0.1)	(0.0)	0.6	0.4	0.6	(1.7)	(1.3)	(0.0)	0.9	0.3	1.7	(0.5)	(0.4)	1.5	0.2
2004	1.5	1.8	0.6	0.9	0.9	0.8	(0.5)	0.2	1.7	1.9	1.1	1.4	0.7	0.1	1.4	2.4
2005	(1.1)	(0.8)	0.8	1.0	0.1	1.9	0.4	0.6	1.4	0.8	1.4	1.5	(1.7)	1.7	(0.6)	1.2
2006	0.3	1.2	(0.8)	1.1	0.9	(0.9)	(0.1)	0.8	(0.5)	0.2	0.4	0.0	(0.1)	0.3	0.3	0.6
2007	(0.2)	(0.7)	(0.6)	1.0	(1.1)	(1.9)	0.2	0.8	0.9	(1.7)	2.6	0.9	0.4	0.9	(1.4)	(0.5)
	0.29	0.27	(0.00)	0.93	0.22	0.13	(0.35)	0.21	0.69	0.42	1.16	1.10	(0.23)	0.50	0.24	0.80

Figure 31: Sharpe ratios for signals in trading time using 15 PCAs as factors : 2002-2007
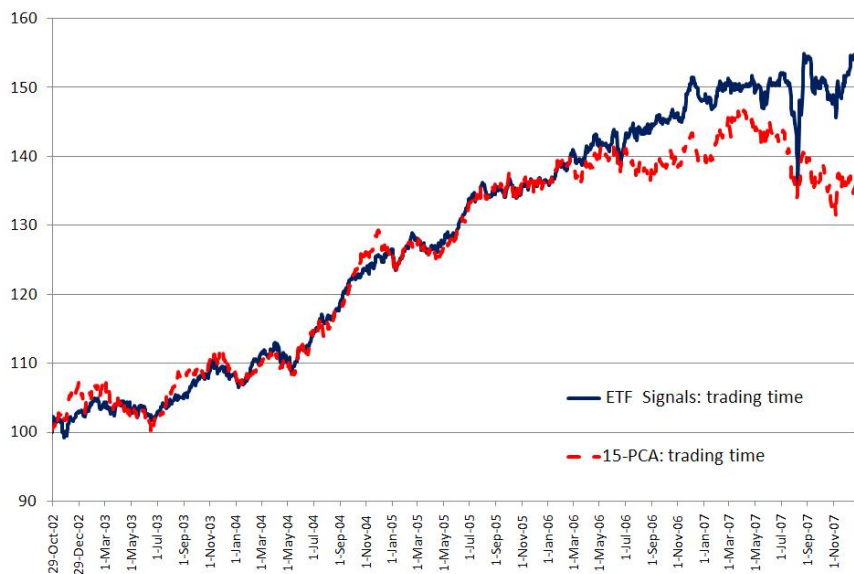
Figure 32: Comparison of ETF and PCA strategies using "trading time".

simulation after adjusting for leverage (KL used 4+4 leverage and we used 2+2 in this paper). The PCA-based strategies showed more resilience during the liquidity event, with a drawdown of 5% as opposed to 10% for the ETF-based strategies (see Figure 33).

Khandani and Lo suggest that the events of 2007 could be due to a liquidity shock caused by funds unwinding their positions. As we have seen, these strategies result in levered portfolios with hundreds of long and short positions in stocks. While each position is small and has probably small impact, the aggregate effect of exiting simultaneously hundreds of positions may have produced the spike shown in Figure 34. A closer look at the PL for different sectors shows, for example, that the Technology and Consumer Discretionary sectors were strongly affected by the shock – and more so than Financials and Real Estate; see Figure 36. This apparently paradoxical result – whereby sectors that are uncorrelated with Financials experience large volatility – is consistent with the unwinding theory of Khandani and Lo. A further breakdown of the performance of the different sectors in August 2007 is given in Figure 35.

8 Conclusions

We presented a systematic approach to statistical arbitrage and for constructing market-neutral portfolio strategies based on mean-reversion. The approach is
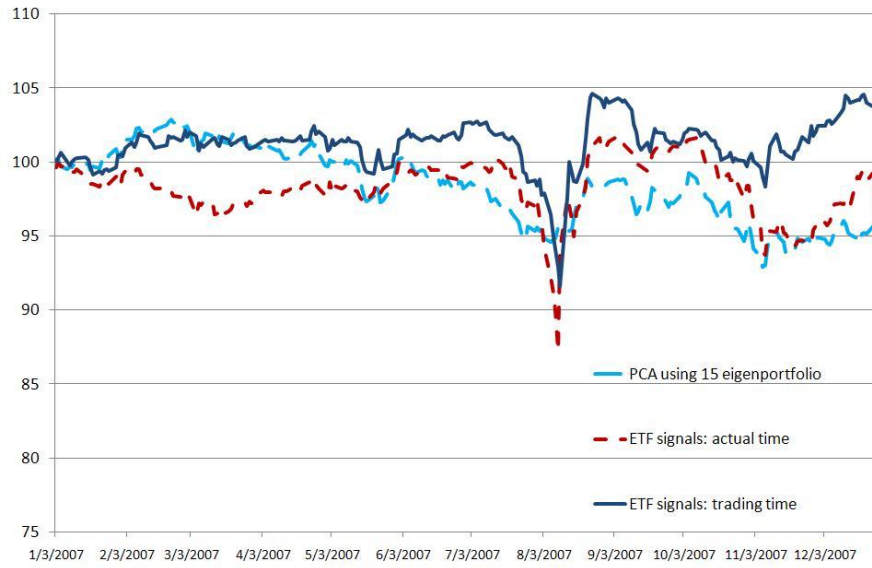
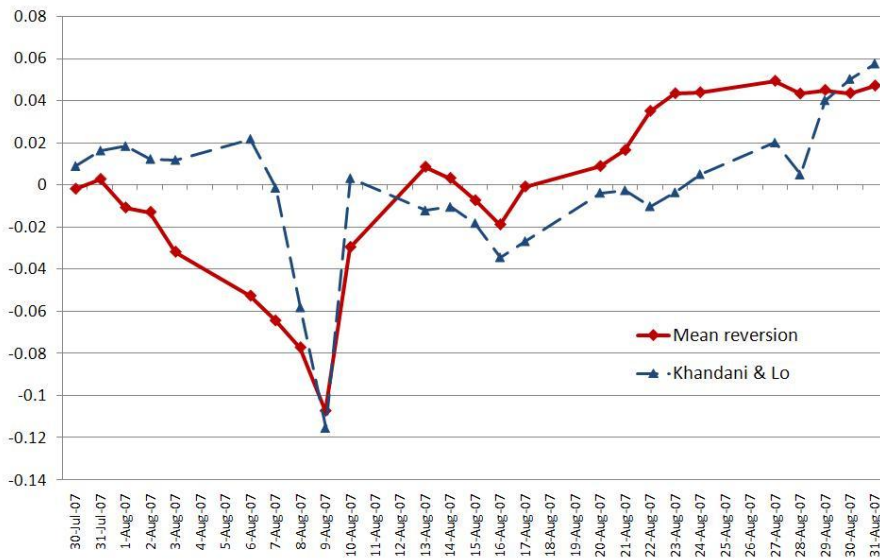Figure 33: Zoom on 2007 of strategies of ETF factor with trading time, ETF factor with actual time and 15 PCA.

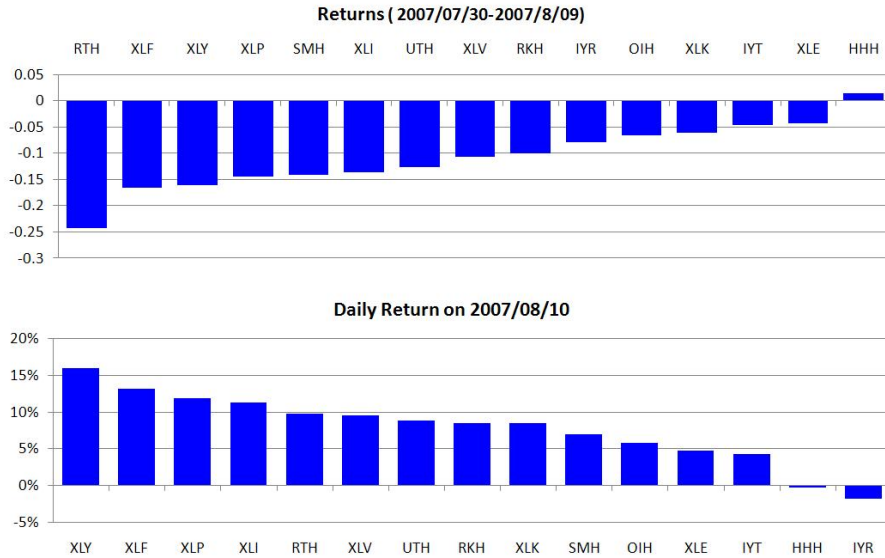Figure 34: Comparison with Khandani & Lo:August 2007, 2+2 leverage on both strategies
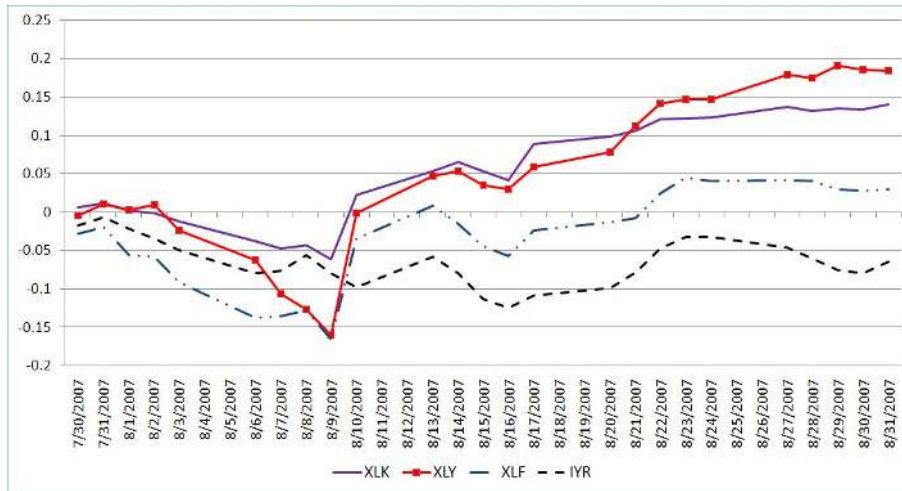
Figure 35: Sector view in Aug 2007

Figure 36: Technology & Consumer vs. Financials & Real Estate : Aug 2007

43

based on decomposing stock returns into systematic and idyosincrating components. This is done using different definitions of risk-factors: ETFs as proxies for industry factors or a PCA-based approach where we extract factors, or eigen-portfolios from the eigenvectors of the empirical correlation matrix of returns.

It is interesting to compare the ETF and PCA methods. In the ETF method, we essentially use 15 ETFs to representing the 'market' fluctuations. It is not difficult to verify that, on average, the systematic component of returns in equity markets, explains between 40% and 60% of the variance of stock returns. This suggests, on the PCA side, that the number of factors needed to explain stock returns should be equal to the number of eigenvalues needed to explain approximately 50% of the variance of the empirical correlation matrix. In practice, we found that this number to vary across time, somehwere between 10 and 30. More precisely, we find that the number varies inversely to the value of the VIX Option volatilty index, suggesting more factors are needed to explain stock returns when volatility is low, and less in times of crisis, or large cross-sectional volatility.

On the performance side, we found that the best results across the entire period were obtained using 15 ETFs or the 15-PCA strategy, or a variable number of PCA factors explaining approximately 55% of the total variance. Trading-time estimation of signals, which is equivalent to weighting returns inversely to the traded volume, seems to benefit particularly the ETF strategy and make it competitive with PCA.

We also note that the performance of mean-reversion strategies appear to benefit from market conditions in which the number of explanatory factors is relatively small. That is, mean-reversion statistical arbitrage works better when we can explain 50% of the variance with a relatively small number of eigenvalues/eigenvectors. The reason for this is that if the "true" number of factors is very large (> 25) then using 15 factors will not be enough to 'defactor the returns', so residuals 'contain' market information that the model is not able to detect. If, on the other hand, we use a large number of factors, the corresponding residuals have small variance, and thus the opportunity of making money, especially in the presence of transaction costs, is diminished.

Finally, we have reproduced the results of Khandani and Lo (2007) and thus place our strategies in the same broad universality class as the contrarian strategies of their paper. Interestingly enough, an analysis of PNL at the sector level shows that the spike of August 2007 was more pronounced in sectors such as Technology and Consumer Discrectionary than in Financials and Real Estate, lending plausibility to the "unwinding theory" of Khandani and Lo.

9   Appendix: estimation of the residual process

We describe our approach for the estimating co-integration residuals as Orstein-Uhlembeck processes and for the calculation of s-scores. We do not claim that this is the most sophisticated or efficient method for estimating the price processes, but simply one that can be readily used (and almost certainly im-

proved) by industry researchers.

For simplicity, we describe the estimation of the OU parameters for the case of ETF regressions, the case of PCA being similar. The first step is to estimate the regression

$$R_n^S = \beta_0 + \beta R_n^I + \epsilon_n, \quad n = 1, 2, ..., 60.$$

relating stock returns to the corresponding ETF returns. Here we assume that returns are chronologically ordered, and R_{60}^S is the last observed return, based on the variation of the closing prices from yesterday to today. Recalling the model (10), we set

$$\alpha = \beta_0 / \Delta t = \beta_0 * 252.$$

Next, we define auxiliary process

$$X_k = \sum_{j=1}^{k} \epsilon_j \quad k = 1, 2, ..., 60,$$

which can viewed as a discrete version of $X(t)$, the OU process that we are estimating. Notice that the regression "forces" the residuals to have mean zero, so we have

$$X_{60} = 0.$$

The vanishing of X_{60} is an artifact of the regression, due to the fact that the betas and the residuals are estimated using the same sample.[10]

The estimation of the OU parameters is done by solving the 1-lag regression model

$$X_{n+1} = a + bX_n + \zeta_{n+1}, \quad n = 1, ..., 59.$$

According to (13), we have

$$
\begin{aligned}
a &= m \left(1 - e^{-\kappa \Delta t} \right) \\
b &= e^{-\kappa \Delta t} \\
\text{Variance}(\zeta) &= \sigma^2 \frac{1 - e^{-2\kappa \Delta t}}{2\kappa}
\end{aligned}
$$

---

[10]This does not have to be the case. For instance, we can use 90 days to estimate the regression and 60 days to estimate the process.

whence

$$
\begin{aligned}
\kappa &= -log(b) * 252 \\
m &= \frac{a}{1-b} \\
\sigma &= \sqrt{\frac{\text{Variance}(\zeta) \cdot 2\kappa}{1-b^2}} \\
\sigma_{eq} &= \sqrt{\frac{\text{Variance}(\zeta)}{1-b^2}}
\end{aligned}
\tag{21}
$$

Fast mean-reversion (compared to the 60-day estimation window) requires that $\kappa > 252/30$, which corresponds to mean-reversion times of the order of 1.5 months at most. In this case, $0 < b < 0.9672$ and the above formulas make sense. If b is too close to 1, the mean-reversion time is too long and the model is rejected for the stock under consideration.

Notice that the s-score, which is defined theoretically as

$$
s = \frac{X(t) - m}{\sigma_{eq}}
$$

becomes, since $X(t) = X_{60} = 0$,

$$
s = \frac{-m}{\sigma_{eq}} = \frac{-a \cdot \sqrt{1-b^2}}{(1-b) \cdot \sqrt{\text{Variance}(\zeta)}}.
$$

The last caveat is that we found that centered means work better, so we set

$$
\overline{m} = \frac{a}{1-b} - \left\langle \frac{a}{1-b} \right\rangle
$$

where brackets denote averaging over different stocks. The s-score is therefore,

$$
s = \frac{-\overline{m}}{\sigma_{eq}} = \frac{-a \cdot \sqrt{1-b^2}}{(1-b) \cdot \sqrt{\text{Variance}(\zeta)}} + \left\langle \frac{a}{1-b} \right\rangle \cdot \sqrt{\frac{1-b^2}{\text{Variance}(\zeta)}}
\tag{22}
$$

References

Associated Press, Quant funds endure August turmoil. *The Motley Fool*, December 6, 2007.

Barr, A., Quant quake shakes hedge-fund giants Goldman, Renaissance, AQR see losses, but also sense opportunity, *Marketwatch*, August 13, 2007.

Cont, R., Da Fonseca, J., Dynamics of implied volatility surfaces. *Quantitative Finance*, 2002, Vol 2, No 1, 45-60.

Davis, G., Mallat, S. and Avellaneda, M., Adaptive greedy approximations. *Constructive Approximations*, 1997, Vol. 13, No. 1, 57-98.

Jolliffe, I. T., *Principal Components Analysis*, Springer Series in Statistics, Springer-Verlag, Heidelberg, 2002.

Khandani, A. E. and Lo, A. W., What happened to the quants in August 2007? *SSRN*, 2007.

Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J. P., Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance, 2000*, Vol. 3, No. 3, 391-397.

Lehmann, B., Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, 1990, Vol. 105, No.1, 1-28.

Litterman, R. and Scheinkman, J. A., Common factors affecting bond returns. *Journal of Fixed Income*, June 1991, 54-61.

Lo, A. W. and MacKinlay, A. C., When are contrarian profits due to stock market overreaction? *The Review of Financial Studies*, 1990, Vol. 3, No. 2, 175-205.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. N., Guhr, T. and Stanley, H. E., Random matrix approach to cross correlations in financial data. *Phys. Rev.*, 2002, E 65, 066126.

Pole, A., *Statistical arbitrage: Algorithmic trading insights and techniques*, Wiley Finance, 2007.

Poterba, J. M. and Summers, L. H., Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, 1988, Vol. 22, 27-59.

Potters, M., Bouchaud, J. P. and Laloux, L., Financial application of random matrix theory: old laces and new pieces. *Acta Physica Polonica B*, 2005, Vol. 36, No. 9, 2767.

Rusli, E. M., Goldman Sachs Alpha to Fail?, *Forbes.com*, August 9, 2007.

Scherer, K. P. and Avellaneda, M., All for one and one for all? A principal component analysis of Latin American brady bond debt from 1994 to 2000. *International Journal of Theoretical and Applied Finance*, 2002, Vol. 5, No. 1, 79-106.