Statistical arbitrage in the US equities market

Marco Avellaneda & Jeong-Hyun Lee

Statistical arbitrage in the US equities market

MARCO AVELLANEDA*†‡ and JEONG-HYUN LEE†

†Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012, USA
‡Finance Concepts, 49-51 Avenue Victor-Hugo, Paris 75116, France

We study model-driven statistical arbitrage in US equities. Trading signals are generated in two ways: using Principal Component Analysis (PCA) or regressing stock returns on sector Exchange Traded Funds (ETFs). In both cases, the idiosyncratic returns are modelled as mean-reverting processes, which leads naturally to 'contrarian' strategies. We construct, back-test and compare market-neutral PCA- and ETF-based strategies applied to the broad universe of US equities. After accounting for transaction costs, PCA-based strategies have an average annual Sharpe ratio of 1.44 over the period 1997 to 2007, with stronger performances prior to 2003. During 2003–2007, the average Sharpe ratio of PCA-based strategies was only 0.9. ETF-based strategies had a Sharpe ratio of 1.1 from 1997 to 2007, experiencing a similar degradation since 2002. We also propose signals that account for trading volume, observing significant improvement in performance in the case of ETF-based signals. ETF-strategies with volume information achieved 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 detail the performance of the strategies during the liquidity crisis of the summer of 2007, following Khandani and Lo [Social Science Research Network (SSRN) working paper, 2007].

Keywords: Alternative investments; Cointegration; Correlation modelling; Quantitative trading strategies

1. Introduction

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

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

or, in its differential version,

$$\frac{\mathrm{d}P_t}{P_t} = \alpha \, \mathrm{d}t + \beta \frac{\mathrm{d}Q_t}{Q_t} + \mathrm{d}X_t, \qquad (2)$$

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

*Corresponding author. Email: avellane@CIMS.nyu.edu

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 MacKinlay 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 be 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 Exchange Traded Fund (ETF) (or, more generally, factors that explain the systematic stock 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 is 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 modelling the idiosyncratic part. The general decomposition may look like

$$\frac{\mathrm{d}P_t}{P_t} = \alpha \, \mathrm{d}t + \sum_{j=1}^{n} \beta_j F_t^{(j)} + \mathrm{d}X_t, \qquad (3)$$

where the terms $F_t^{(j)}$, $j = 1, \ldots, n$, 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 Poterba and Summers (1988), Lehmann (1990), and Lo and MacKinlay (1990). 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. In 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 (PCA) (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 US gives rise to risk-factors that have economic significance because they can be interpreted as long–short portfolios of industry sectors. However, the stocks that contribute the most to a particular factor are not necessarily the largest capitalization stocks in a given sector. This suggests that PCA-based risk factors may not be as biased towards large-capitalization stocks as ETFs, as the latter are generally capitalization-weighted. We also observe that the variance explained by a fixed number of PCA eigenvectors varies significantly across time, which leads us to conjecture that the number of explanatory factors needed to describe stock returns (to separate systematic returns from residuals) 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.† This might explain some of the differences that we found in performance between the PCA and ETF methods.

In sections 3 and 4, we construct the trading signals. This involves the statistical estimation of the residual process for each stock at the close of each trading day, using 60 days of historical data prior to that date. Estimation is always done *looking back 60 days* from the trade date, thus simulating decisions which might take place in real trading. The trading signals correspond to significant deviations of the residual process from its estimated mean. Using daily end-of-day (EOD) data, we perform a calculation of daily trading signals, going back to 1996 in the case of PCA strategies and to 2002 in the case of ETF strategies, across the universe of stocks with

---

†See Scherer and Avellaneda (2002) for similar observations for Latin American debt securities in the 1990s.

market-capitalization of more than 1 billion USD at the trade date. The condition that the company must have a given capitalization at the trade date (as opposed to at the time when the paper was written), avoids survivorship bias.

Estimation and trading rules are kept simple to avoid data-mining. For each stock, the estimation of the residual process is done using a 60-day trailing window because this corresponds roughly to one earnings cycle. The length of the window is not changed from one stock to another. We select 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 several strategies which use different sets of factors to generate residuals, namely: (i) synthetic ETFs based on capitalization-weighted indices,† (ii) actual ETFs, (iii) a fixed number of factors generated by PCA, (iv) a variable number of factors generated by PCA. Due to the mechanism described above used to generate trading signals, the simulation is always out-of-sample, in the sense that the estimation of the residual process at time t uses information available only for the 60 days prior to this time. In all trades, 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 favour) contrarian price signals taking place on low volume and mitigates (i.e. tends not to favour) contrarian price signals which take place on high volume. It is as if we 'believe more' a print that occurs on high volume and are 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 context of mean-reversion strategies, even at the EOD sampling frequency and not only for intra-day trading.

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

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

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

The coefficients $\bar{\beta}_j$ 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

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

$$= \sum_{j=1}^m \left[\sum_{i=1}^N \beta_{ij} Q_i \right] F_j + \sum_{i=1}^N Q_i \tilde{R}_i = \sum_{i=1}^N Q_i \tilde{R}_i. \tag{7}$$

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 systematic component of stock returns explains approximately 50% of the variance (see Laloux *et al.* 2000 and Plerou *et al.* 2002). The question is how to define 'factors'.

2.1. The PCA approach

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 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.† Let us represent the stocks return data, on any given date t_0, going back $M + 1$ days

---

†Synthetic ETFs are capitalization-weighted sector indexes formed with the stocks of each industry that are present in the trading universe at the time the signal in calculated. We used synthetic ETFs because most sector ETFs were launched only after 2002.

as a matrix

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

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

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

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

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

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 behaviour 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. In this paper, we always use an estimation window for the correlation matrix of one-year (252 trading days) prior to the trading date.

The commonly used solution to extract meaningful information from the data is to model the correlation matrix.‡ We consider the eigenvectors and eigenvalues of the empirical correlation matrix and rank the eigenvalues in decreasing order:

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

We denote the corresponding eigenvectors by

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

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 fewer 'detached' eigenvalues than industry sectors. Therefore, we expect that the boundary between 'significant' and 'noise' eigenvalues to be somewhat blurred and to correspond to being at the 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 $\lambda_1,\ldots,\lambda_m$, $m < N$ 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 are defined as

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

The eigenportfolio returns are therefore

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

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 eigenportfolio returns.

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 modelling 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,

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

†For instance, the analysis can be restricted to the members of the S&P500 index in the US, the Eurostoxx 350 in Europe, etc.
‡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.
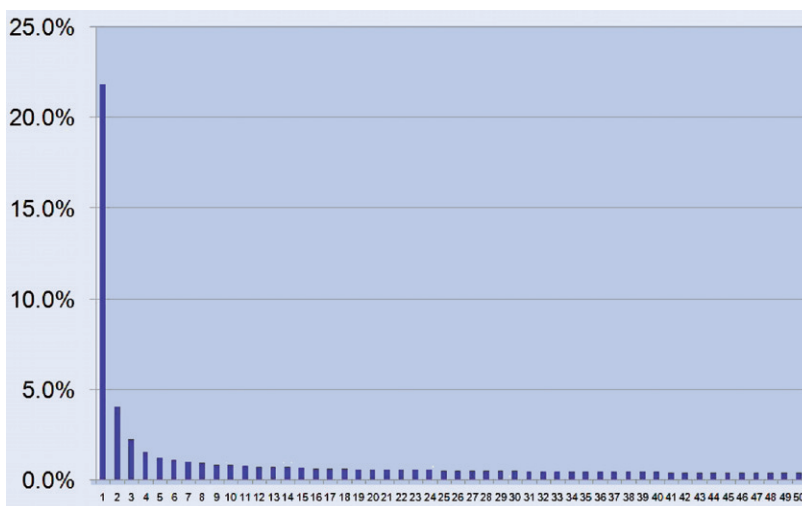
Figure 1. Top 50 eigenvalues of the correlation matrix of market returns computed on 1 May 2007 estimated using a one-year window and a universe of 1417 stocks (see table 3). (Eigenvalues are measured as percentage of explained variance.)
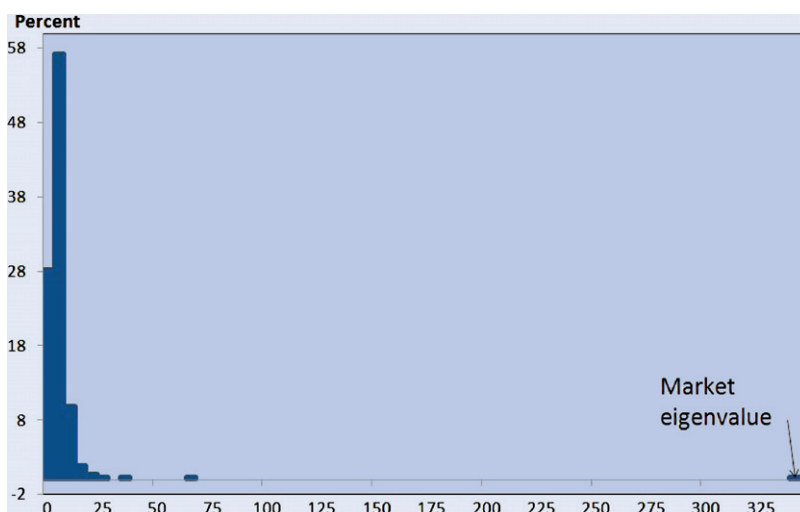

Figure 2. The density of states for 1 May 2007 estimated using a one-year window, corresponding to the same data used to generate figure 1. Notice that there are some 'detached eigenvalues', and a 'bulk spectrum'. The relevant eigenvalues include the detached eigenvalues as well as a eigenvalues in the edge of the bulk spectrum.

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 $v_i^{(1)}$, $i = 1, 2, \ldots, N$, are positive. Thus, the eigenportfolio has positive weights $Q_i^{(1)} = v_i^{(1)}/\overline{\sigma}_i$. 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,† 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 $v^{(1)}$; (ii) given that there is no natural order in the stock universe, the 'shape analysis' that is used to interpret the PCA of

---

†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.
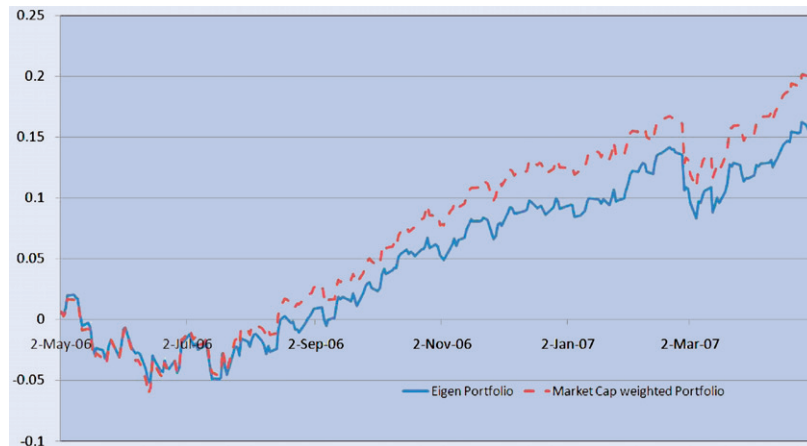
M. Avellaneda and J.-H. Lee

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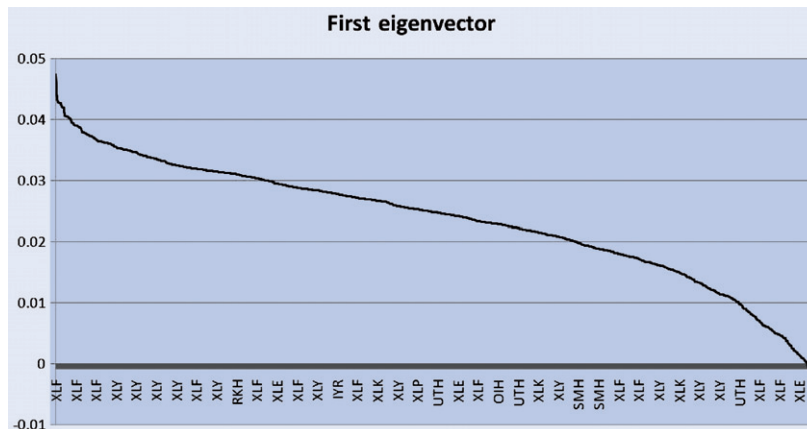

Figure 4. First eigenvector sorted by coefficient size. The *x*-axis shows the ETF corresponding to the industry sector of each stock.

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

Top 10 stocks energy, oil and gas	Bottom 10 stocks 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 2. The top 10 stocks and bottom 10 stocks in third eigenvector.

Top 10 stocks Utility	Bottom 10 stocks 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.

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 analysed 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:

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

the sequence n_i representing a re-labelling of the companies. In this new ordering, we notice that the 'neighbours'

Table 3. Trading universe on 1 January 2007, broken down by sector. The market is therefore partitioned into 15 sectors and each stock is associated with an ETF.

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	**XU**	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	**XIV**	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

of a particular company tend to be in the same industry group (see tables 1 and 2 and figures 4, 5 and 6). 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 long–short portfolios at the level of industries or sectors.

2.3. The ETF approach: using the industries

Another method for extracting residuals consists in using the returns of sector ETFs as factors. Table 3 shows a sample of industry sectors' numbers of stocks of companies with capitalization more than US$1 billion at the beginning of January 2007, classified by sector. It 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).

Unlike the case of eigenportfolios, which are uncorrelated by construction, ETF returns are correlated. This can lead to redundancies in the factor decomposition: strongly correlated ETFs sometimes give rise to large factor loadings with opposing signs for stocks that belong to or are strongly correlated with different ETFs. There are several approaches that can be used to remedy this: one is a robust version of multiple regression aiming at 'sparse' representations. For example, the *matching pursuit* algorithm (Davis *et al.* 1997) which favours sparse representations is preferable to a full multiple regression. Another class of regression methods known as *ridge regression* would achieve a similar goal (see, for instance, Jolliffe 2002).

In this paper we use a simple approach. We associate to each stock a single sector ETF (following the partition of the market shown in table 3) and perform a regression of the stock returns on the corresponding ETF returns, i.e.

$$R_i = \beta R_{\text{ETF}_i} + \tilde{R}_i,$$

where ETF_i is associated with stock i.

3. A relative-value model for equity valuation

We propose a quantitative approach to stock valuation based on its relative performance within industry sector ETFs or, alternatively, with respect to the constructed PCA factors. In section 4, we present a modification of this approach which takes into account the trading volume in the stocks, within a similar framework. Our investment model is purely based on price and volume 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 $S_i(t), \ldots, S_N(t)$, 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

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

where the term

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

represents the systematic component of returns (driven by the returns of the eigenportfolios or the ETFs). The coefficients β_{ij} are the corresponding factor loadings.

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.

In the case of ETF factors, we work with the model

$$\frac{dS_i(t)}{S_i(t)} = \alpha_i \, dt + \beta_i \frac{dI(t)}{I(t)} + dX_i(t), \qquad (11)$$

where $I(t)$ is the ETF corresponding to the stock under consideration.†

In both cases, the idiosyncratic component of the return is

$$d\tilde{X}_i(t) = \alpha_i \, dt + dX_i(t).$$

Here, 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 the 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 which are not driven by the systematic factors (PCA or ETF).

Based on these assumptions, we introduce a simple parametric model for $X_i(t)$ which can be estimated easily, namely the Ornstein–Uhlenbeck process:

$$dX_i(t) = \kappa_i(m_i - X_i(t))dt + \sigma_i \, dW_i(t), \quad \kappa_i > 0. \qquad (12)$$

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

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

The conditional mean, or forecast of expected daily returns of the residual process, 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 *de facto* to vary slowly in relation to the Brownian motion increments $dW_i(t)$, in the time-window of interest. In the simulations, we estimate the residual processes for each stock on a window of length 60 days, assuming implicitly that the parameters are constant over the window. We accept this hypothesis on stocks for which the speed of mean-reversion (the estimate of κ) is sufficiently high and reject it for stocks having a slow

---

†In other words, we analyse a 'pair-trade' between each stock and its assigned ETF.
‡Among the myriad of mean-reverting processes, we choose the simplest one to model residuals. We encourage the reader interested in practical implementations to experiment with other models, as we have in the course of this research.

Table 4. 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. The best years are 1996–2002. Sector Sharpe ratios assume beta-neutrality with respect to the S&P 500 index.

	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

Table 5. Sharpe ratios for actual 15 ETFs as factors: 2002–2007. Industry Sharpe ratios assume beta-neutrality with respect to the corresponding ETF. We observe, for the purpose of comparison, that the average Sharpe ratio from 2003 to 2007 was 0.6. Sharpe ratios above 1.0 where obtained in 2002 and 2004.

	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

speed of mean-reversion. Details on how to estimate the model are given in the next section and in Appendix A.

It follows from (12) that

$$X_i(t_0 + \Delta t) = e^{-\kappa_i \Delta t} X_i(t_0) + (1 - e^{-\kappa_i \Delta t}) m_i$$
$$+ \sigma_i \int_{t_0}^{t_0 + \Delta t} e^{-\kappa_i(t_0 + \Delta t - s)} dW_i(s). \quad (13)$$

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

$$E\{X_i(t)\} = m_i \quad \text{and} \quad Var\{X_i(t)\} = \frac{\sigma_i^2}{2\kappa_i}. \quad (14)$$

According to equation (10), an 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 factor (or the ETF, in the case of the ETF framework) has an expected one-day return

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

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

$$\tau_i = 1/\kappa_i$$

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. $T_1 = 60/252$. 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 table 5. For the details of the estimation of the Ornstein–Uhlenbeck process and more statistical details on signal generation, see Appendix A.

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

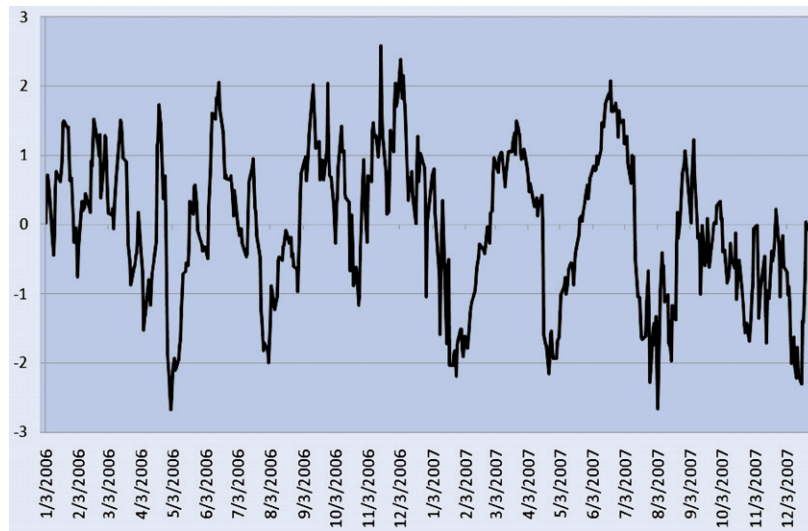$$s_i = \frac{X_i(t) - m_i}{\sigma_{eq,i}}. \quad (15)$$

Figure 7. Evolution of the *s*-score of JPM (versus XLF) from January 2006 to December 2007.

We call this variable the *s-score*.† See figure 7 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

$$\begin{aligned} &\text{buy to open if } s_i < -\bar{s}_{bo} \\ &\text{sell to open if } s_i > +\bar{s}_{so} \\ &\text{close short position if} s_i < +\bar{s}_{bc} \\ &\text{close long position } s_i > -\bar{s}_{sc} \end{aligned} \quad (16)$$

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 $\bar{s}_{bo}$, $\bar{s}_{bo}$, $\bar{s}_{bc}$, $\bar{s}_{sc}$ 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

$$\begin{aligned} &\bar{s}_{bo} = \bar{s}_{so} = 1.25 \\ &\bar{s}_{bc} = 0.75 \quad \text{and} \quad \bar{s}_{sc} = 0.50. \end{aligned}$$

Thus, we enter opening trades when the *s*-score exceeds 1.25 in absolute value. We close long trades when the *s*-score reaches -0.50. Closing short trades sooner, at $s = 0.75$, gives slightly better results than 0.50 in the training period of 2000–2002, so we use this slightly asymmetric rule in back-testing.

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 section, the presence of the drift α was ignored. That is, we assumed that the drift was statistically insignificant in comparison with the amplitude of excursions $\sigma_{eq} = \sigma/\sqrt{2\kappa}$. In this section, we show how to incorporate the drift. This leads to a modified *s*-score.

We consider the conditional expectation of the residual return over a period of time d*t*, namely,

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

This suggests that the dimensionless decision variable is the 'modified *s*-score'

$$s_{mod, i} = s_i - \frac{\alpha_i}{\kappa_i \sigma_{eq,i}} = s_i - \frac{\alpha_i \tau_i}{\sigma_{eq,i}}. \quad (17)$$

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

---

†See Appendix A for practical details on estimating the *s*-score.

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 the cutoffs defined in the previous strategy are also acceptable for this one.

We observed empirically, through the calculation of signals, that the drift α has values of the order of 15 basis points, the average expected reversion time is 7 days, and the equilibrium volatility of residuals is of 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 back-testing simulations, the effect of incorporating a drift in these time-scales of a few days is minor. Thus, for the sake of brevity, we do not present back-testing results with the modified *s*-scores.

In essence, we claim that the residual process has no significant drift, or at least that trading based on the simple constant drift α estimated from the model does not improve trading performance over assuming that $\alpha = 0$. We could say that, in aggregate, stock returns have negligible momentum after controlling for industry/size factors, on the trading scale of interest.

5. Back-testing results

The back-testing experiments consist in running the signals through historical data, simulating trades in all the stocks in the universe according to the signals in (16). Estimation of parameters (betas, residuals) and signal evaluations are performed daily. Estimation of the parameters sometimes leads to values of $\kappa_i < 8.4$. When κ_i crosses this threshold, we reject the model and (i) do not open trades or (ii) close open trades.

We assume that all trades are done at the closing price of that day. As mentioned previously, we assume 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:

$$E_{t+\Delta t} = E_t + E_t r \Delta t + \sum_{i=1}^{N} Q_{it} R_{it} - \left(\sum_{i=1}^{N} Q_{it} \right) r \Delta t$$
$$+ \sum_{i=1}^{N} Q_{it} D_{it} / S_{it} - \sum_{i=1}^{N} |Q_{i(t+\Delta t)} - Q_{it}| \epsilon,$$
$$Q_{it} = E_t \Lambda_t,$$

where Q_{it} represents the 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 lending and borrowing 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 expect to have 100 stocks long and 100 short and we wish to have a '2 + 2' leverage, then $\Lambda_t = 2/100$.† Another way to think of Λ_t is as the maximum fraction of the equity that can be invested in any given stock.‡ The choice of $2 + 2$ leverage was made to target a volatility of approximately 10% by back-testing in the period 2002–2004. Of course, the choice of leverage does not affect the Sharpe ratio and other choices would give comparable results after standardization (Khandani and Lo 2007 run a '1/2 + 1/2' contrarian strategy over a different, broader, stock universe).

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, probably due to model mis-specification.

5.1. Synthetic ETFs as factors

The first set of experiments was done using 15 *synthetic capitalization-weighted industry-sector indices* as risk-factors (see table 3). The reason for using synthetic ETFs was that most sector ETFs were launched after 2000. In order to be able to back-test strategies going back to 1996, when most ETFs did not exist, and to compare the strategies with PCA, we decided to construct capitalization-weighted sector indices and to construct residuals based on these indices.

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 then perform a regression of the stock returns on the returns of the associated sector index and extract the corresponding residual series and trading signals.

To ensure market-neutrality, we added to the portfolio an S&P 500 index hedge using the SPY, which was adjusted daily and kept the overall portfolio beta-neutral. In other words, due to the fact that the synthetic ETFs are

---

†2 + 2 leverage means 2 dollars long and 2 dollars short per dollar of equity in the portfolio. In practice, Λ_t is adjusted only for new positions, so as not to incur transaction costs for stock which are already held in the portfolio.
‡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.

Figure 8. Historical PNL for the strategy using synthetic ETFs as factors from 1996–2007. The strategy does not produce significant returns after 2004.
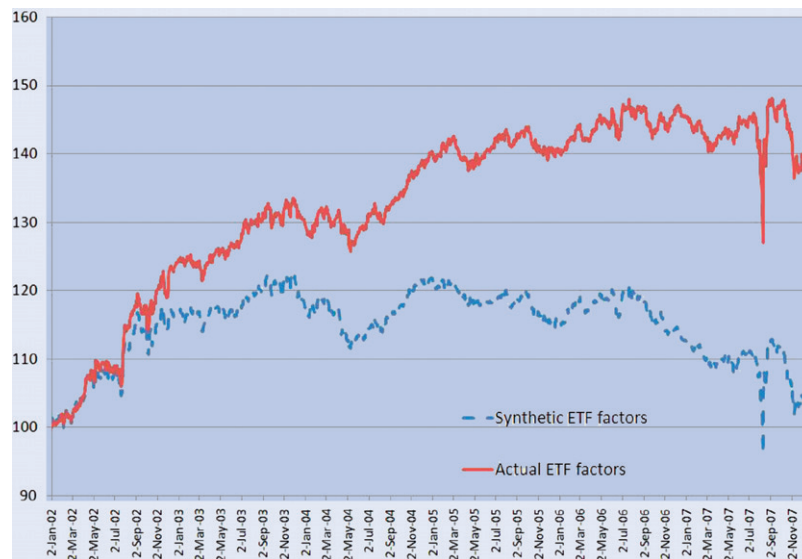

Figure 9. Historical PNL for the strategy using actual ETFs as factors, compared with the one using synthetic ETFs: 2002–2007. Notice the strong out-performance by the strategy which uses actual ETFs.

not traded instruments, we trade the stocks according to the signal and buy or sell the S&P 500 Index ETF in the appropriate amount so as to be beta-neutral.

Since we expect that, on average, stocks are correctly priced, we experimented with adjusting the mean of the OU processes exogenously. We introduced the adjusted means for the residuals

$$\overline{m}_i = m_i - \frac{1}{N} \sum_{j=1}^{N} m_j, \quad i = 1, 2, \ldots, N. \qquad (18)$$

This modification has the effect of removing 'model bias'; it expresses that we expect that on average the mean of the residual is zero. We obtained consistently better results in back-testing when using $\overline{m}_i$ instead of m_i. Therefore, we adopted this centring approach in all the simulations (synthetic ETFs, ETFs and PCA).

The results of back-testing with synthetic ETFs are shown in figures 8 and 9, and in table 4.

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 before. We back-tested the strategy going back to 2002, using regression on the ETF assigned to each stock to generate residuals. The results are displayed on figure 9 and table 5.

The simulations suggest that using actual ETFs improves performance considerably in relation to synthetic ETFs. A possible explanation for this improvement is that ETFs are traded instruments, whereas the synthetic ETFs are not, thus providing better price information. Another possible explanation is hedging: in the case of
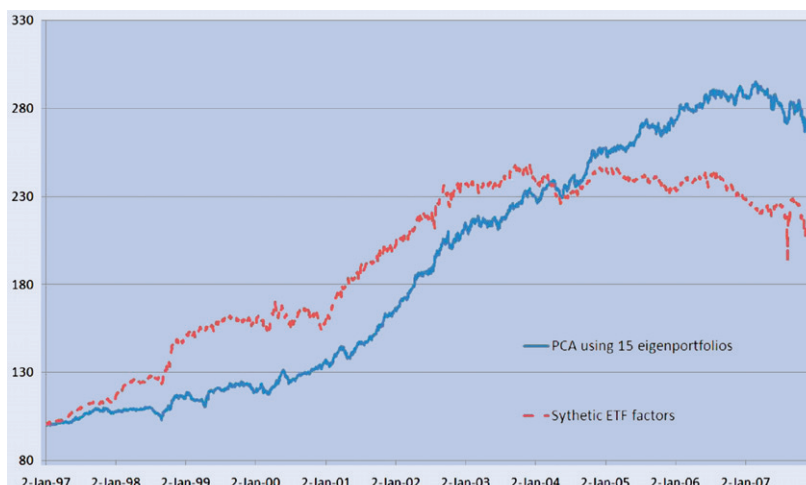
Figure 10. PNL corresponding 15 PCA factors, compared with synthetic ETFs from 1997–2007. The reason for starting in 1997 is that we need one year of data to compute the initial correlation matrix. The PCA strategy produces superior results, particularly after 2002.
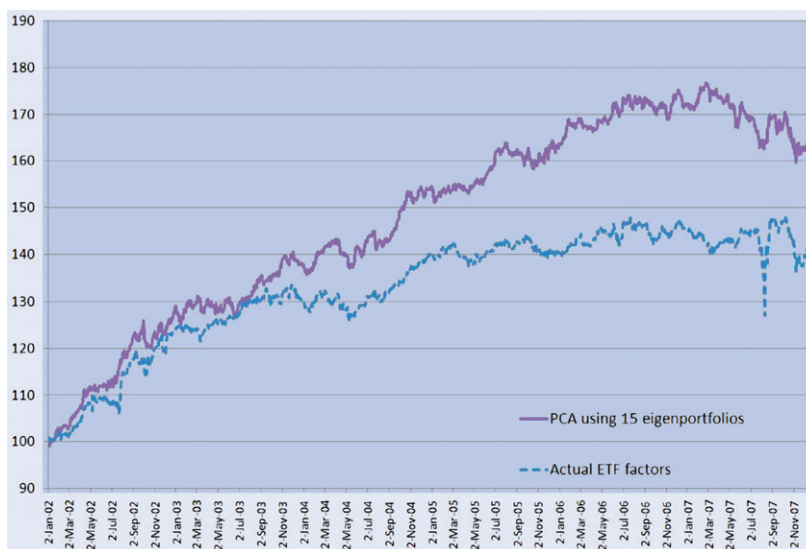

Figure 11. Comparison of strategies with 15 PCA factors and using actual ETFs in the period 2002–2007. 15-PCA outperforms significantly the strategy with actual ETFs.

actual ETFs we were able to neutralize the portfolio across each industry using ETFs as hedges, whereas the back-testing with synthetic ETFs is beta-neutral with respect to the SPY.

5.3. PCA with 15 eigenportfolios

The back-testing results for signals generated with 15 PCA factors are shown in figures 10 and 11 and table 6. In the case of PCA factors, we trade the signals as explained in the previous section. To ensure market neutrality, we hedge daily with the S&P 500 tracker as in the case of synthetic ETFs.

The first set of results, displayed in figure 10, shows the progression of the equity in the two portfolios corresponding to synthetic ETFs and PCA with 15 eigenportfolios.

The second set of results, displayed in figure 11, compare the performances of PCA with 15 eigenportfolios with the strategy with actual ETFs. We observe that the 15-PCA strategy out-performs the actual ETF strategy

since 2002. It is noteworthy that this is the case even if the actual ETF strategy is hedged sector-by-sector but the 15-PCA strategy is hedged only at the level of SPY.

5.4. Using a variable number of PCA factors

We also back-tested strategies based on a variable number of factors, with the number of factors chosen so as to explain a given level of variance. In this approach, we retain a certain number of eigenportfolios (factors) such that the sum of the corresponding eigenvectors is equal to a set percentage of the trace of the correlation matrix.

The number of eigenvalues (or eigenvectors) which are needed to explain 55% of the total variance varies in time. This variability is displayed in figures 12 and 13. We also looked at other cutoffs and report similar results in figure 15. The periods over which the number of eigenvectors needed to explain a given level of variance is low appear to be those when the risk-premium for equities is relatively high. For instance, the latter parts of 2002 and

Table 6. Sharpe ratios for 15 PCA factors: 1997–2007. We start in 1997 due to the fact that we need one year's worth of data to compute the initial correlation matrix. The best performances, with Sharpe ratios above 2.0, were 2000, 2001, 2002 and 2004. Sectors are beta-neutral with respect to the S&P 500 index.

|	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
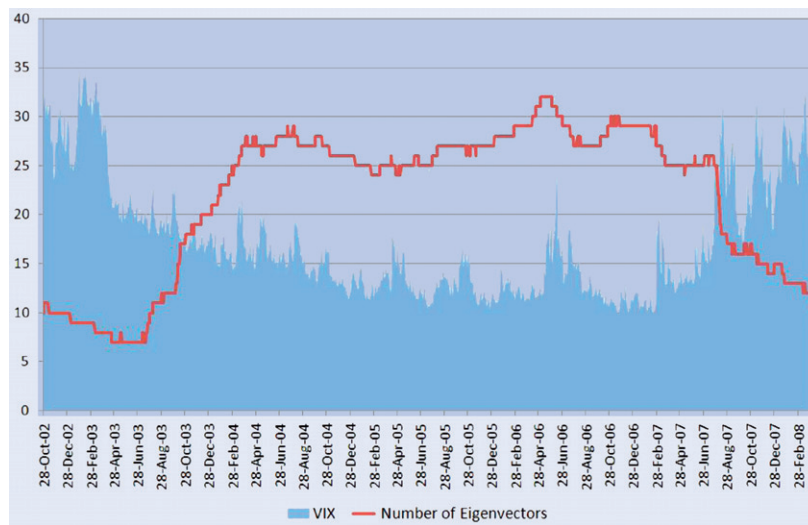
Figure 12. Number of significant eigenvectors needed to explain the variance of the correlation matrix at the 55% level, from 2002 to February 2008. The estimation window for the correlation matrix is 252 days. The boundary of the shaded region represents the VIX CBOT Volatility Index (measured in percentage points).
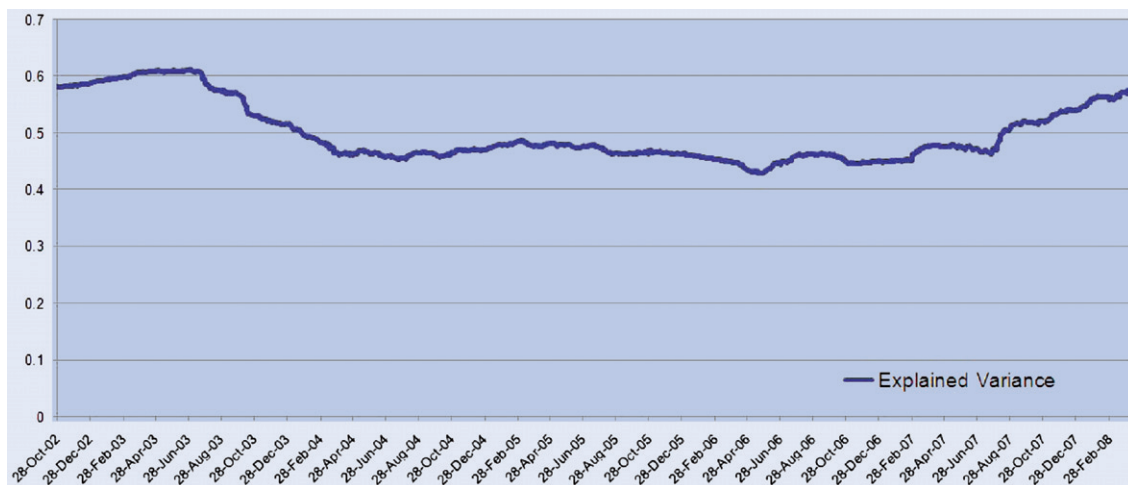

Figure 13. Percentage of variance explained by the top 15 eigenvectors: 2002–February 2008. Notice the increase in the Summer of 2007.
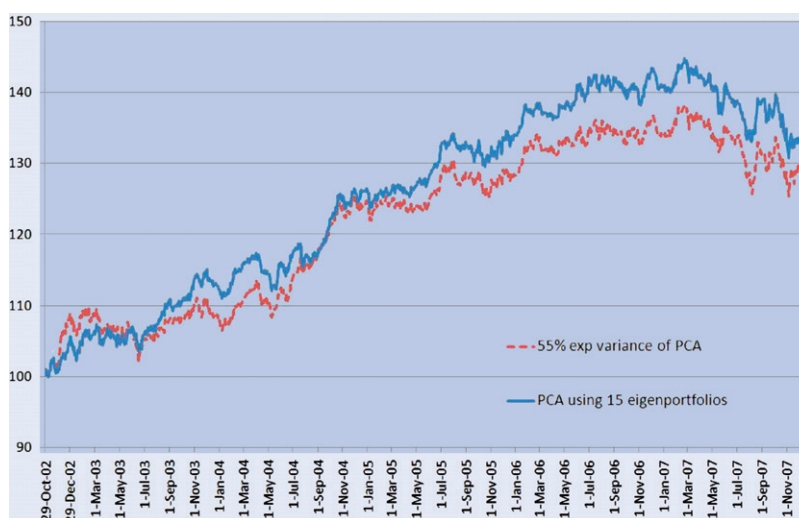
Figure 14. Comparison of the PNLs for the fixed explained variance (55%) of PCA and the 15 PCA strategy: 2002–2007. The performance of the 15 PCA strategy is slightly superior.
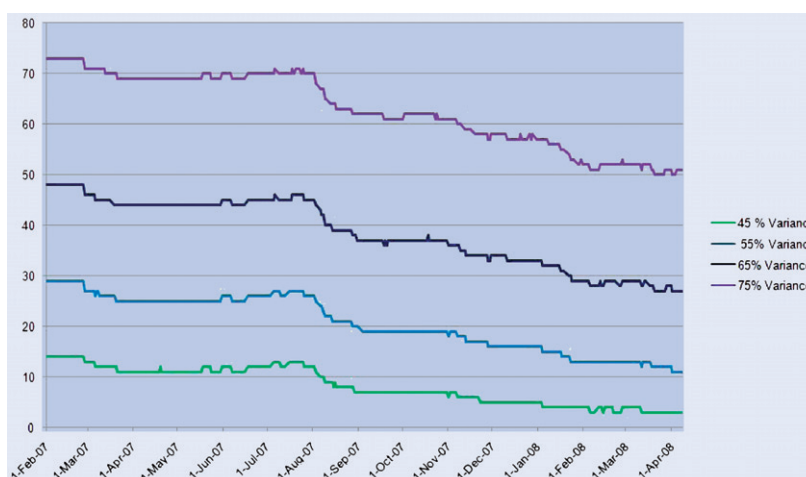

Figure 15. Time-evolution of number of PCA factors for different levels of explained variance: 2002–2007. The decay in the number of factors is associated with the onset of the subprime crisis in the summer of 2007.

Table 7. Sharpe ratios for the fixed explained variance (55%) of PCA: 2003–2007. Sectors are neutral with respect to the S&P 500 index.

	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

2007, which correspond to 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. The equity risk premium, as represented by the VIX, reaches a historical low in early 2007.

Back-testing the strategy with 55% explained variance shows that the strategy is comparable but slightly inferior to taking 15 eigenvectors (see figure 14 and table 7).

The corresponding results are presented in figure 17 and table 9. As before, we trade daily according to the *s*-score levels and keep the portfolio beta-neutral with respect to the S&P 500 index.

In the same vein, we studied the performances of other strategies with a variable number of eigenportfolios explaining different levels of variance. In table 8 and figure 16, we display the performances of strategies using 45, 55 and 65% compared with the PCA strategies with 1 eigenportfolio and with 15 eigenportfolios.
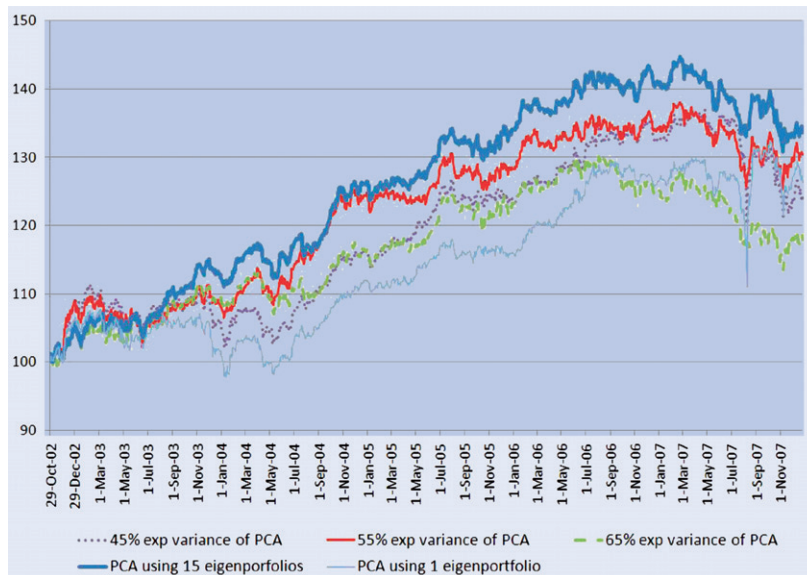
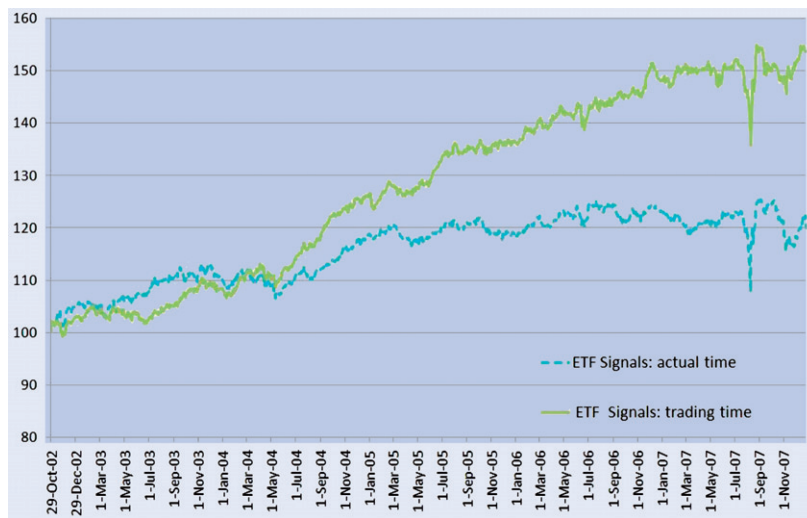Figure 16. PNL for different variance truncation levels: 2002–2007.

Figure 17. Comparison of strategies corresponding to signals generated using trading time versus using calendar time, using actual ETFs as factors: 2002–2007.

Table 8. Sharpe ratios for variable PCA strategies: 2002–2007.

	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

Table 9. Sharpe ratios for signals in trading time using actual ETFs as factors: 2002–2007. Sector portfolios are beta-neutral with respect to the corresponding ETF.

	HHH	IYR	IYT	OIH	RKH	RTH	SMH	UTH	XLE	XLF	XLI	XLK	XLP	XLV	XLY	Porfoto
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	0.20	(0.12)	0.54	0.88	0.28	1.26	1.91	(0.28)	0.87	0.76	1.51

Table 10. Sharpe ratios for signals in trading 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

The conclusion is that 55% PCA performs best among the three strategies and underperforms slightly the 15 PCA strategy. We also observe that taking a high cutoff such as 75% of explained variance leads invariably to steady losses, probably due to the fact that transaction costs dominate the small residual variance which remains in the system after 'defactoring'. Using too many factors lead to 'noise trading'!

On the opposite side of the spectrum, using only one eigenportfolio, as in the Capital Asset Pricing Model, gives rise to lower speeds of mean-reversion, higher residual volatilities and worst Sharpe ratios.

6. Taking trading volume into account

In this section, we incorporate 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

$$S_{t+\Delta t} - S_t = P_{V(t+\Delta t)} - P_{V(t)}$$
$$= \frac{P_{V(t+\Delta t)} - P_{V(t)}}{V(t + \Delta t) - V(t)}[V(t + \Delta t) - V(t)]. \quad (19)$$

Thus, the *price change per share traded* 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 returns, we use the modified returns

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

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 'rescaling' stock returns as in (20).

The modified returns $\overline{R}_t$ are equal to the classical returns if the daily trading volume is typical. If the trading volume is low, the factor on the right-hand side of the last equation is larger than unity and $\overline{R}_t > R_t$. Conversely, 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 ETF and PCA strategies using the trading-time approach. The window for calculating the trading volume was taken to be ten trading days, a choice dictated by the fact that the measured volume should be over a period smaller than the estimation window for the residuals but long enough to average spikes.†

We found that the technique increases the PNL and the Sharpe ratios unequivocally for strategies with ETF-generated signals (figure 17 and table 9). For PCA-based strategies, a trading-time framework does not seem to produce a significant improvement (figure 18 and table 10). Finally, we find that the ETF strategy using trading time is comparable in performance to the 15-PCA/55% PCA strategies until 2006 and performs slightly better after that, if we exclude the August 2007 drawdown (figure 19 and table 10).

7. A closer look at 2007

It is well-known that 2007 was a very challenging year for quantitative hedge funds – see Khandani and Lo (2007), Barr (2007), the Associated Press (2007) and Rusli (2007). After a mediocre performance in the first half of the year, statistical arbitrage strategies experienced a large drawdown followed by a partial recovery in the second week of August 2007. Unfortunately for many managers, the size of the drawdown was so large that many had to de-leverage their portfolios and did not recover to pre-August levels (see the above references for details).

---

†This value was chosen as an educated first guess. We did not optimize the trading-time window length in this study.
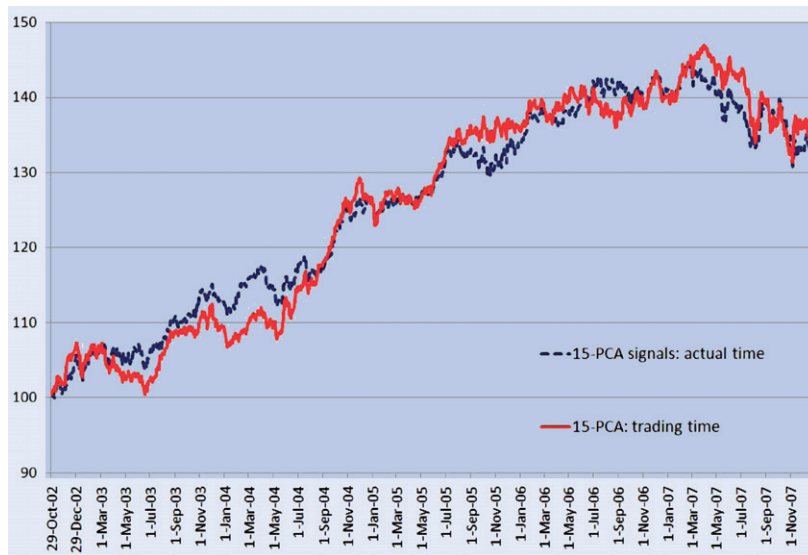
Figure 18. Comparison of signals in trading time versus actual time using 15 PCAs as factors: 2002–2007.
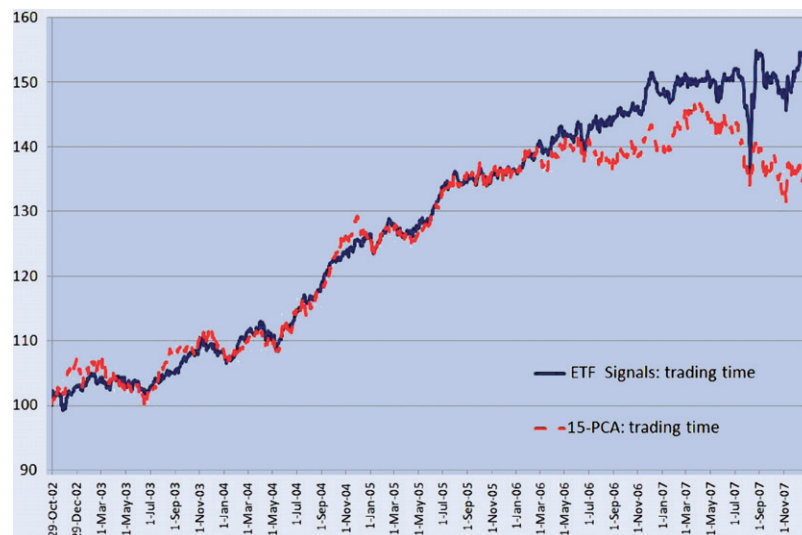

Figure 19. Comparison of ETF and PCA strategies using trading time. The ETF strategy in trading time clearly outperforms the PCA strategy.

Our back-testing results are consistent with the real-world events of 2007 and also show a strong drawdown in August 2007 (see below). This drawdown was first reproduced in back-testing by Khandani and Lo (2007) using their classical contrarian strategy.

We carefully analysed the performance for all the strategies in this paper in 2007, with the exception of synthetic ETFs (which are known to underperform the others). Figure 20 displays the results graphically.

First, we found that performance of all strategies was flat or slightly negative in the first part of the year. In early August, we found that all the 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) simulation after adjusting for leverage (KL used $4+4$ in 2007 leverage whereas we use $2+2$ in this paper). 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 20).

Khandani and Lo (2007) suggest that the events of 2007 could have been due to a liquidity shock caused by funds unwinding their positions. As we have seen, market-neutral statistical arbitrage strategies result in leveraged portfolios with hundreds of long and short positions. 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 21.

A closer look at the different sectors shows that the Technology and Consumer Discretionary sectors were strongly affected by the shock – and more so in terms of price movements than Financials and Real Estate; see figure 22, which gives a breakdown of the performance of the different industry sectors in August 2007. This apparently paradoxical result – whereby sectors that are uncorrelated with Financials experience large volatility – is
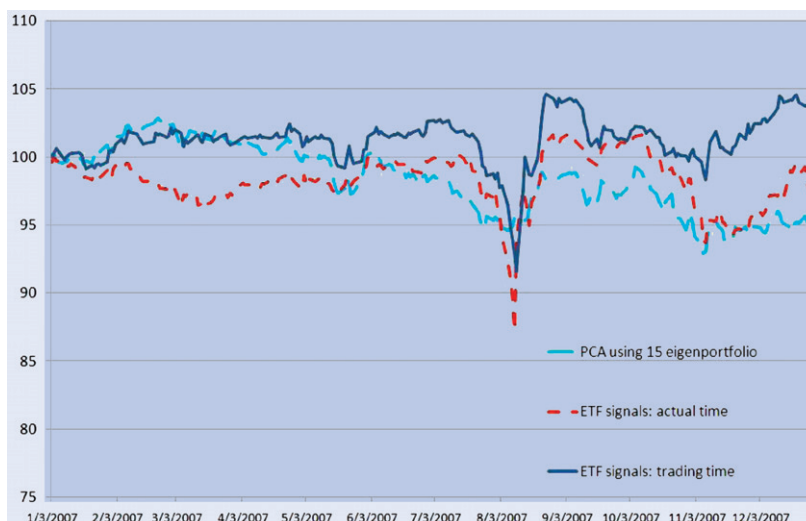
Figure 20. Zoom on 2007. Performance of strategies with ETF factors with trading time, ETF factor with calendar time and 15-PCA. All strategies exhibit a similar drawdown in August 2007, like the contrarian strategy of Khandani and Lo (2007).
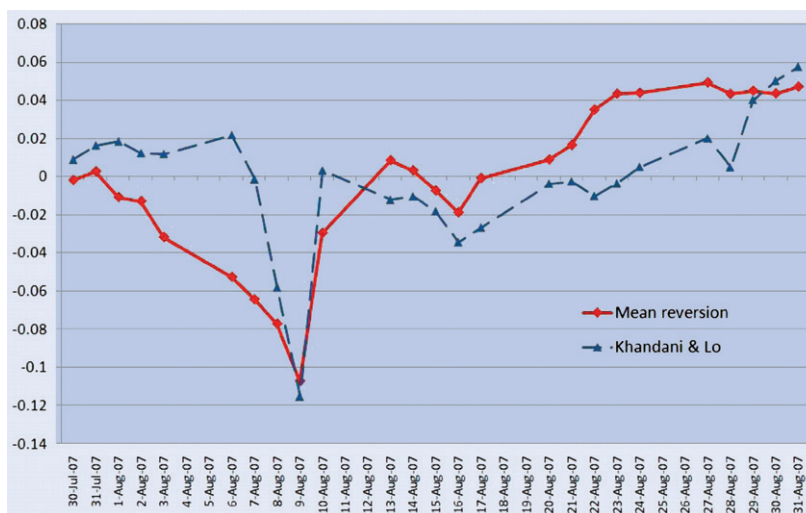

Figure 21. Comparison of the ETF strategy in trading time with Khandani and Lo during August 2007. The leverage is $2 + 2$ for both strategies.

consistent with the unwinding theory of Khandani and Lo. In fact, it shows a sudden 'divergence' of the mean-reversion strategy in sectors unrelated to financials and real estate.

8. Conclusions

We have presented a systematic approach for constructing market-neutral portfolio strategies based on mean-reversion. The approach is based on decomposing stock returns into systematic and idiosyncratic components. This extraction of 'residuals' is done using different definitions of risk-factors: (i) ETFs as proxies for industry factors or (ii) a PCA-based approach where we extract factors from eigenvectors of the empirical correlation matrix of returns.

We compared extensively the ETF and PCA methods. In the ETF method, we used 15 liquid ETFs to representing the systematic price fluctuations, and we

de-trended stock returns by regressing each of them on the associated ETF returns.

With this definition of systematic–idiosyncratic decomposition, the systematic component of stock returns explains between 40 and 60% of the variance. 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 this number to vary across time, and to lie somewhere between 10 and 30 factors. We also observe that this number varies inversely to the value of the VIX Option volatility 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.

A word on hedging. For the trading strategies based on actual ETFs, we hedged the portfolio daily, on each sector, with ETFs. Hence, the resulting portfolio is beta-neutral per sector. In the case of strategies generated using
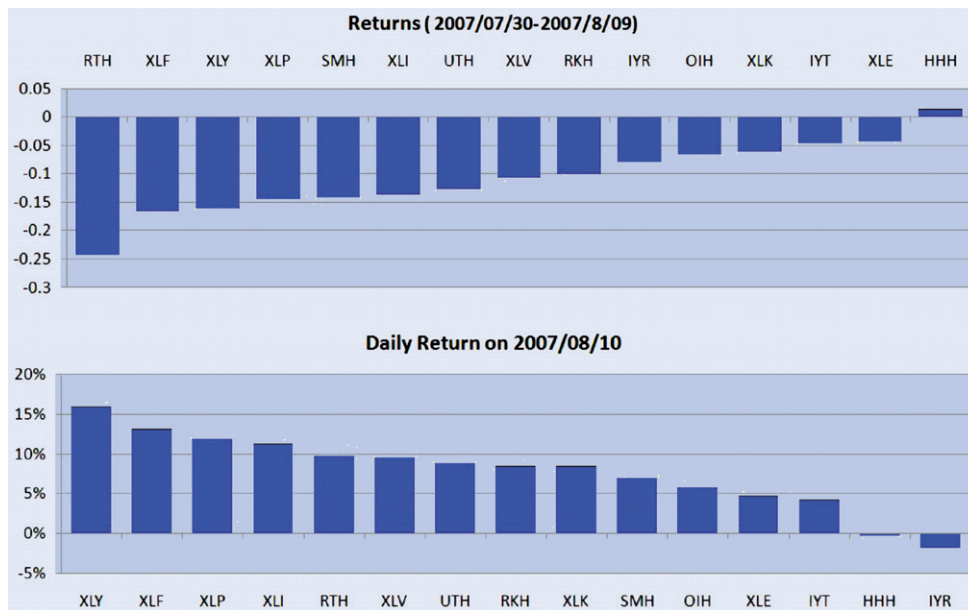
Figure 22. Breakdown of performance per sector in the ETF strategy, August 2007. All sectors are beta-hedged with ETFs.

'synthetic' ETFs or using the PCA factors, we did not enforce beta-neutrality per sector. Instead we traded all the signals and hedged the overall portfolio beta using SPY, the S&P 500 tracker. The reason for this is that we wanted to evaluate the performance of the synthetic ETF and the PCA strategies beginning in the mid-1990s, when there were few sector ETFs. (Hedging with ETFs was not possible then, with the exception of SPY, which was available also during that period.†)

We back-tested the strategies using data from 1996 to 2007 for synthetic ETFs and PCA. Back-testing with actual ETFs was only possible since 2002, after the sector ETFs were created. We then compared all strategies over the relevant time periods, and particularly after 2002, when ETFs were available. The best performing strategies were the ETF strategy in trading time and the 15-PCA strategy. A PCA strategy with a variable number of factors explaining 55% of the variance also performed relatively well.

Estimation of signals in trading time is equivalent, for EOD trading, to weighting returns inversely to the daily traded volume. This additional feature, which incorporates trading volume into our signals, appears to benefit particularly the ETF strategy and to make it competitive with PCA. Given its simplicity, this means that the ETF-based in trading-time strategy might merit further attention and refinements.

We also noted that, in general, the performance of mean-reversion strategies appears 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. We conjecture therefore that markets having many stocks (i.e. investment opportunities), and yet are in conditions in which they are driven by a small number of explanatory factors, may be the most propitious for this strategy. This conjecture might be useful for performance evaluation of this class of strategies and to determine when they may provide good investment opportunities, although more studies, particularly after 2007, may be required to ascertain this.

Finally, we 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 Discretionary than in Financials and Real Estate, confirming the plausibility of the 'unwinding theory' of Khandani and Lo.

†Of course, we could have hedged the PCA portfolios using sector ETFs *after* 2002, but we preferred not to do this in order to keep ideas simple. Other more sophisticated risk-management methods, such as trading with limits on each PCA factor exposure, would have been possible. We leave these more sophisticated portfolio strategies as avenues for future research.

comments and suggestions. All errors and omissions are ours, of course.

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*, 13 August, 2007.

Cont, R. and Da Fonseca, J., Dynamics of implied volatility surfaces. *Quant. Finan.*, 2002, **2**, 45–60.

Davis, G., Mallat, S. and Avellaneda, M., Adaptive greedy approximations. *Constr. Approx.*, 1997, **13**, 57–98.

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

Khandani, A.E. and Lo, A.W., What happened to the quants in August 2007? Social Science Research Network (SSRN) working paper, 2007. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 1015987 (accessed 18 July 2009).

Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J.P., Random matrix theory and financial correlations. *Int. J. Theor. & Appl. Finan.*, 2000, **3**, 391–397.

Lehmann, B., Fads, martingales, and market efficiency. *Quart. J. Econ.*, 1990, **105**, 1–28.

Litterman, R. and Scheinkman, J.A., Common factors affecting bond returns. *J. Fixed Income*, 1991, **1**, 54–61.

Lo, A.W. and MacKinlay, A.C., When are contrarian profits due to stock market overreaction? *Rev. Finan. Studies*, 1990, **3**, 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. E*, 2002, **65**, 066126-1–066126-18.

Pole, A., *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*, 2007 (Wiley Finance: New York).

Poterba, J.M. and Summers, L.H., Mean reversion in stock prices: evidence and implications. *J. Finan. Econ.*, 1988, **22**, 27–59.

Rusli, E.M., Goldman Sachs Alpha to Fail? *Forbes.com,* 9 August, 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. *Int. J. Theor. & Appl. Finan.*, 2002, **5**, 79–106.

Appendix A: Estimation of the residual process

We describe our approach for estimating co-integration residuals as Ornstein–Uhlenbeck 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 improved) by practitioners.

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

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

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

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

Next, we define auxiliary process

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

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.[†]

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

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

According to (13), we have

$$a = m(1 - e^{-\kappa \Delta t})$$
$$b = e^{-\kappa \Delta t}$$
$$\text{Variance}(\zeta) = \sigma^2 \frac{1 - e^{-2\kappa \Delta t}}{2\kappa}$$

whence

$$\kappa = -\log(b) \times 252$$
$$m = \frac{a}{1 - b}$$
$$\sigma = \sqrt{\frac{\text{Variance}(\zeta) \times 2\kappa}{1 - b^2}}$$
$$\sigma_{\text{eq}} = \sqrt{\frac{\text{Variance}(\zeta)}{1 - b^2}}. \tag{A1}$$

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 formulae 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

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

[†]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.

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

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

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

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

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

$$s = \frac{-\overline{m}}{\sigma_{eq}} = \frac{-a \times \sqrt{1 - b^2}}{(1 - b) \times \sqrt{\text{Variance}(\zeta)}} + \left\langle \frac{a}{1 - b} \right\rangle$$
$$\times \sqrt{\frac{1 - b^2}{\text{Variance}(\zeta)}}. \tag{A2}$$