Journal of Banking & Finance 65 (2016) 134–155

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

A new approach to statistical arbitrage: Strategies based on dynamic factor models of prices and their performance

Sergio M. Focardi [a,b], Frank J. Fabozzi [c,*], Ivan K. Mitov [d]

[a] Stony Brook University-New York (SUNY), United States
[b] De Vinci University, Paris, France
[c] EDHEC Business School, France
[d] FinAnalytica Inc., United States

ARTICLE INFO

Article history:
Received 25 June 2013
Accepted 3 October 2015
Available online 14 October 2015

JEL classification:
G11
G12
G17

Keywords:
Statistical arbitrage
Return-based factor models
Price-based factor models
Diebold–Mariano framework
Long-short strategies
Long-only strategies

ABSTRACT

Statistical arbitrage strategies are typically based on models of *returns*. We introduce a new statistical arbitrage strategy based on dynamic factor models of *prices*. Our objective in this paper is to exploit the mean-reverting properties of prices reported in the literature. We do so because, to capture the same information using a return-based factor model, a much larger number of lags would be needed, leading to inaccurate parameter estimation. To empirically test the relative performance of return-based and price-based models, we construct portfolios (long-short, long-only, and equally weighted) based on the forecasts generated by two dynamic factor models. Using the stock of companies included in the S&P 500 index for constructing portfolios, the empirical analysis statistically tests the relative forecasting performance using the Diebold–Mariano framework and performing the test for statistical arbitrage proposed by Hogan et al. (2004). Our results show that prices allow for significantly more accurate forecasts than returns and pass the test for statistical arbitrage. We attribute this finding to the mean-reverting properties of stock prices. The high level of forecasting accuracy using price-based factor models has important theoretical and practical implications.

1. Introduction

Statistical arbitrage strategies have as their objective to produce positive, low-volatility returns that are uncorrelated with market returns. In this paper, we present a new econometric framework for statistical arbitrage based on dynamic factor models using the logarithm of prices (logprices). The main contribution of this paper is to show that dynamic factor models of logprices are a powerful method to implement and study statistical arbitrage strategies. We show that our strategies based on dynamic factor models of prices worked well even during the financial crisis that started in 2007.

We implement statistical arbitrage strategies based on dynamic factor models of returns and on dynamic factor models of prices and then analyze and compare performance. Using historical equity price data, we show that strategies based on factor models of logprices largely outperform those based on factor models of returns. In addition, long-only strategies based on dynamic factor models of logprices outperform benchmark strategies proposed in the literature.

Factor models of logprices and factor models of returns can coexist. Although factor models of logprices and factor models of returns should theoretically offer similar performance, in order to capture the mean-reverting properties of prices, factor models of returns would require a larger number of factors and would therefore require estimating a larger number of parameters. As a consequence, the performance of factor models of returns is limited by the size of available samples. The intuition behind this finding is that in order to capture mean reversion, factor models of returns would require a long series of lagged factors while prices carry more information and allow for more parsimonious models.

The plan of the paper is the following. In Section 2, we present a brief review of the literature on pairs trading and statistical arbitrage. Our database and models, our portfolio strategies, and the methodologies we employed to evaluate the quality of forecasts are described in Sections 4–6, respectively. In Section 7 we illustrate how the number of factors is determined. In Section 8, we discuss the coexistence of factor models of prices and returns.

* Corresponding author.
E-mail address: fabozzi321@aol.com (F.J. Fabozzi).

http://dx.doi.org/10.1016/j.jbankfin.2015.10.005
0378-4266/

The trade-off implied in making forecasts at different time horizons and issues related to the bid-ask spread is the subject of Section 9. Our results are presented in Section 10, followed by our conclusions in Section 11.

2. Pairs trading and statistical arbitrage

The intuition behind statistical arbitrage is to exploit the spread of expected returns of large portfolios of financial assets. Statistical arbitrage strategies go long in the portfolio of assets with the highest expected return and short the portfolio of assets with the lowest expected return. Because the spread is considered to be uncorrelated with market returns, a statistical arbitrage is a market-neutral strategy. The forerunners of statistical arbitrage strategies were pairs trading strategies. Pairs trading strategies work by finding pairs of assets whose prices approximately follow the same path. When the spread between the two assets widens, traders will take long positions in the asset with the lower price and short positions in the other asset. When the prices converge again, the trader will liquidate positions and make a profit.

There are two main issues in implementing a pairs trading strategy. The first is identifying those assets whose prices move close together. The second is determining a trading strategy to decide in what asset to take a long-short position and when. Vidyamurthy (2004) and Pole (2007) provide general presentations of pairs trading. Several papers describe trading strategies and methods for identifying suitable pairs. Alexander (1999) and Alexander and Dimitriu (2002, 2005a,b) propose the framework of cointegration for identifying optimal pairs for trading. Bossaerts (1988) finds evidence of cointegration in stock prices. Modeling the stochastic spread process between the two assets in a pair as a mean-reverting Ornstein–Uhlenbeck process, Elliot et al. (2005) propose different strategies for estimation of the process and optimal trading. Gatev et al. (2006) present a detailed analysis of the performance of pairs trading strategies within a global framework of cointegration and find evidence of pairs trading profit. Triantafyllopoulos and Montana (2011) use state-space models to forecast spreads in pairs trading.

Do et al. (2006) propose modeling the stochastic spread between two stocks (i.e., the stochastic difference between their logprices) in the framework of a theoretical equilibrium asset pricing model. Boguslavsky and Boguslavskaya (2003), Mudchanatongsuk et a. (2008) and Montana et al. (2009) provide methods and algorithms for optimal strategies.

Pairs trading has been generalized to statistical arbitrage, a strategy based on constructing zero-cost portfolios, going long a portfolio of assets with the highest expected return and going short a portfolio of assets with the lowest expected return. Statistical arbitrage strategies are not true arbitrage strategies since they only generate, for any finite time horizon, an *expected* positive return, not a sure return. These strategies typically tend to make a large number of individual independent trades with a positive expected return, thereby reducing the risk of the strategy.

Although statistical arbitrage strategies can be applied to any asset type that permits shorting, our focus in this paper is on stocks and therefore the discussion to follow is cast in terms of this asset class. Extension of our results to other asset classes could be the subject of further research but is beyond the scope of this paper. Given a set of stocks, we can create a statistical arbitrage strategy by forming all possible pairs, choosing with appropriate tests those that are suitable for pairs trading and create all individual pairs trading strategies. In this way, we generate a number of zero-cost long-short positions. Aggregating all long positions and all short positions, we can generate a zero-cost long-short portfolio that is rebalanced every time a new pair is created or an old pair is liquidated.

Avellaneda and Lee (2010) analyze statistical arbitrage strategies using stocks included in the S&P 500 universe as candidates for statistical arbitrage strategies. For selecting stocks, they propose a novel method based on using either exchange-traded funds or principal components of the S&P 500. More specifically, their strategy uses continuous-time mean-reverting processes to model the difference between stock returns and the returns of exchange-traded funds or principal components of the S&P 500.

A strain of the literature uses statistical arbitrage to test the efficient market hypothesis. The key idea is that if such a strategy generates a sequence of positive expected profits then, in the limit of infinite time, statistical arbitrage approaches true arbitrage. For example, Bondarenko (2003) introduces the concept of a "statistical arbitrage opportunity" and finds that the existence of a pricing kernel implies that no such opportunities exist. After providing a precise definition of statistical arbitrage, Hogan et al. (2004) propose a test of market efficiency based on their definition. This test methodology was extended and generalized by Jarrow et al. (2005, 2012). Fernholz and Maguire (2007) discuss statistical arbitrage in the framework of stochastic portfolio theory.

Our objective in this paper is not to discuss questions of market efficiency (a controversial subject which would deserve a separate paper), nor to discuss the implications of statistical arbitrage, but to show that statistical arbitrage opportunities exist and to propose an efficient method to construct statistical arbitrage portfolios. In so doing, we follow the development of the literature on momentum where the phenomenon of momentum was first firmly established and then discuss its implications. In this paper, we define statistical arbitrage as a self-financing trading strategy with a positive expected return for each trading period. The trading strategy is formed selecting a long-short portfolio using the forecasted returns obtained through dynamic factor models. We test the statistical significance of the results but do not discuss market efficiency.

We implement the test for statistical arbitrage described in Hogan et al. (2004). This test is based on a definition of statistical arbitrage that considers the asymptotic behavior of a trading strategy. According to Hogan et al. (2004), a zero-cost, self-financing trading strategy that starts at some time and extends in the infinite future is a statistical arbitrage if, when time tends to infinity, (1) the expectation of gains is positive, (2) the limit of the variance of gains tends to zero, and (3) the probability of losses tends to zero. Because only a finite sample is available, Hogan et al. (2004) estimate a model of gains and formulate their test as a test on the parameters of the estimated model.

3. The development of dynamic factor models

We now review the basic facts about dynamic factor models that are relevant for our models. Several reviews of the literature on dynamic factor models and forecasting are available: Bai and Ng (2008) survey the econometric theory of large dimensional factor analysis; Stock and Watson (2004) survey methods for forecasting large panels of time series data; Eklund and Kapetanios (2008) review forecasting methods for large data sets; Barhoumi et al. (2013) provide a review of the literature on dynamic factor models; Banbura, et al. (2010) survey methods for "nowcasting", that is the forecasting of not-yet-released economic data, using large panels of available data.

Dynamic factor models were introduced in the 1970s by Geweke, (1977) and Sargent and Sims (1977). These researchers define dynamic factor models as where stationary economic variables are linear combinations of stationary factors and their lags. The dynamics of the model are generated by lagged factors.

Peña and Box (1987) formalize dynamic factor models in a state-space framework, where the variables to be modeled are regressed on factors and the factors follow a vector autoregressive moving average (VARMA) process. They define a simplifying transformation representing dynamic factors as linear combinations of observed variables. Models based on lagged factors are particular cases of the more general model defined by Peña and Box (1987). Note that adding dynamics through a VARMA process allows one to represent factors as linear combinations of observed variables.

Forni et al. (2000) propose a dynamic factor model with infinite T and N. Their estimation methodology is based on Brillinger's dynamic principal components analysis. As observed in Stock and Watson (2005), this methodology uses both leads and lags and is not suitable in a forecasting framework. Forni et al. (2009) propose a methodology based on fitting a vector autoregressive (VAR) model to principal components.

Several extensions of the basic dynamic factor model have been proposed. For example, So et al. (1998) and Lopes and Carvahlo (2007) propose stochastic volatility and Markov switching models. However, the parameter count of our models dissuades us from using additional modeling complications. For example, with weekly data, there are approximately 200 data points per return series, and therefore a total of 2000 data points in total for the factors. Given that we have to estimate more than 100 parameters, we have less than 10 data points per estimate. Thus it is not advisable that we increase the number of parameters. Another strain of the literature discusses factor models with integrated factors. Stock and Watson (1988) introduce the notion of common trends in cointegrated systems and propose estimating cointegrating relationships using principal components analysis (PCA). After showing that principal components estimators of cointegrating relationships are consistent, Harris (1997) proposes an asymptotically efficient estimator of cointegrating vectors based on PCA.

Under the assumption that factors are linear combinations of observed series, Gonzalo and Granger (1995) define conditions that distinguish permanent and transitory components. Engle and Vahid (1993) introduce the notion of common cycles and study the relationships between common trends and common cycles.

Peña and Poncela (2004) formalize dynamic factor models with integrated factors along the same lines as Peña and Box (1987), while Escribano and Peña (1994) establish the equivalence between cointegration, common trends, and common factors.

Although Ng et al. (1992) discuss static and dynamic factors, their notion of dynamic factors is different from the notion of dynamic factors in the dynamic factor model literature and does not include any dynamic modeling. Deistler and Hamann (2005) formally describe dynamic factor models applied to stock returns forecasting.

4. Data and models

Let's now discuss the data and the models used to produce a vector of forecasted returns.

4.1. Data

Our database includes the S&P 500 universe of daily stock prices in the 23-year period from January 1989 to December 2011, a period which includes 5799 trading days. In this period, 1127 stock price series appeared at least once as constituents of the S&P 500. Prices are adjusted for (1) splits, mergers, and other corporate actions and (2) the payment of dividends. If a stock pays a dividend, then the dividend is assumed to be reinvested in that stock and its price is adjusted accordingly. The choice of the S&P 500 universe allows us to test different models on liquid stocks, thereby minimizing spurious effects due to non-trading.

4.2. Factor models based on prices

Models based on prices exploit the fact that logprices can be represented as factor models with integrated I(1) factors plus stationary I(0) factors. More specifically, in Section 7.2 we show that prices for the S&P 500 — the universe that we consider in this paper — can be represented with one single integrated factor plus additional stationary factors. Our first model is an application of these findings.

Suppose there are N stocks and that $p_i(t) = \log(P_i(t))$ is the logprice of the ith stock at time t. Let's consider a moving window with ew time steps: $(t - ew + 1 : t)$. At time t, we consider only those stocks that are components of the S&P 500 index in the window $(t - ew + 1 : t)$. At time t, it is known what stocks are in the time window $(t - ew + 1 : t)$ and therefore, in backtesting, if we estimate models in this time window, there is no look-ahead bias due to anticipation of information.

Using the methodology proposed by Stock and Watson (1988), and analyzed by Escribano and Peña (1994) and Harris (1997), we can represent the observed logprices as follows:

$$p(t) = \beta_\perp \beta'_\perp p(t) + \beta\beta' p(t) \tag{1}$$

where $p(t)$ is the vector of logprices, β is the $N \times K$ matrix of cointegrating vectors, and $\beta_\perp$ is its orthogonal complement. If we estimate cointegrating relationships using PCA, we can estimate (1) the matrix β with the eigenvectors of the covariance matrix of the $p(t)$ corresponding to the K smallest eigenvalues and (2) estimate $\beta_\perp$ with the remaining eigenvectors corresponding to the largest N–K eigenvalues.

The expression (1) is the trend plus cycle decomposition of the logprices $p(t)$ in the sense of Engle and Vahid (1993). Assuming there are common cycles, we can write the following theoretical model:

$$\begin{aligned} &p_i(t) = a_{i1}F_1(t) + \cdots + a_{iK}F_K(t) + \varepsilon_i(t), \quad i = 1, \ldots, N \\ &F(t) = B_1F(t-1) + \cdots + B_qF(t-q) + v(t), \quad F(t) = [F_1(t)\ldots F_K(t)] \end{aligned} \tag{2}$$

where the Fs are estimated by the principal components corresponding to the largest Q eigenvalues. This model is a special case of the model described in Peña and Poncela (2004) insofar as we do not include a moving average term in modeling factors F. We assume that (1) the residuals $\varepsilon_i(t)$ have a nonsingular covariance matrix Σ_ε and are serially uncorrelated, (2) residuals $v(t)$ have a nonsingular covariance matrix Σ_v and are serially uncorrelated, and (3) residuals $\varepsilon_i(t)$ and $v(t)$ are mutually uncorrelated and serially uncorrelated at all times and lags. It is further assumed that the VAR model for F is invertible. If p would also depend on a finite number of lagged factors, it is well known that it could be transformed into the model given by (2) by adding lagged factors. In this case, some of the equations of the VAR model given by (2) would reduce to identities.

We make the additional simplifying assumption that the matrices B are diagonal; that is, we estimate each factor independently as an autoregressive model. This assumption is suggested by the fact that factors are uncorrelated by construction. Empirically, we find that cross autocorrelations between factors are minimal. By assuming diagonal Bs, we significantly reduce the number of parameters to estimate.

The number of lags is determined by applying the Akaike information criterion (AIC) and Bayesian information criterion (BIC). In our samples, both criteria suggest using only one lag when the number of factors varies from 5 to 15. Using the methodology proposed by Nyblom and Harvey (2000) to estimate the number of integrated factors, in Section 7.2 we find that there is only one integrated factor. In Section 7.1, we discuss the determination of the number of stationary factors. Anticipating the conclusions of Section 7.1, we will adopt models with a total of 15 factors.

After estimating an autoregressive model for each factor, we forecast factors and obtain $F(t+fh)$ where as defined earlier, fh is the forecasting horizon. In backtesting, if $fh = 1$, we use information up to time t and write: $F(t+1) = B_1F(t) + \cdots + B_qF(t-q+1)$. If $fh > 1$, we create forecasts recursively, rolling over our forecasts.

Next, we regress prices on factors in-sample and obtain the coefficients $a_{i1}, \ldots a_{iK}$. Note that these coefficients can also be obtained from the matrix of the eigenvectors of the covariance matrix of logprices. The final step is to forecast logprices as follows: $fp_i(t+fh) = a_{i1}F_1(t+fh) + \cdots + a_{iK}F_K(t+fh)$. Given forecasted logprices, we obtain forecasted returns:

$$fR_i(t) = \frac{\exp(fp_i(t+fh)) - P_i(t)}{P_i(t)}.$$

Note that the backtesting results obtained with this procedure are strictly out-of-sample: At time *t*, we use only information available at time *t* or earlier. At no step is there any implicit anticipation of information.

4.3. Models based on returns

Our return-based strategies are based on the following dynamic factor model of returns described in Peña and Box (1987):

$$\begin{aligned} &R_t = \beta F_t + \varepsilon_t \\ &\Phi(L)F_t = \Theta(L)\eta_t \\ &\Phi(L) = I - \Phi_1 L - \cdots - \Phi_p L^p \\ &\Theta(L) = I - \Theta_1 L - \cdots - \Theta_q L^q \end{aligned} \tag{3}$$

where factors F are stationary processes, L is the lag operator, ε_t is white noise with a full covariance matrix but is serially uncorrelated, η_t has a full-rank covariance matrix and is serially uncorrelated and ε_t, and η_t are mutually uncorrelated at all lags. That is, the common dynamic structure comes only from the common factors while the idiosyncratic components can be correlated but no autocorrelation is allowed.

Assume that factors are normalized through the identification conditions $\beta'\beta = I$. Consider the covariance matrices $\Gamma_r(k) = E(R_tR_{t-k})$, $k = 0,1,2,\ldots$ and $\Gamma_f(k) = E(F_tF_{t-k})$, $k = 0,1,2,\ldots$ Peña and Box (1987) show that the following relationships hold:

$$\begin{aligned} &\Gamma_r(0) = \beta\Gamma_f(0)\beta' + \Sigma_\varepsilon, \quad k = 0 \\ &\Gamma_r(k) = \beta\Gamma_f(k)\beta', \quad k \geqslant 1 \end{aligned} \tag{4}$$

and that the number of factors is the common rank Q of the matrices. $\Gamma_r(k) > 1$.

If N is small, these models can be estimated using the maximum likelihood method and the Kalman filter. If N and T tend to infinity, we can consider models in their static form and estimate factors with PCA as well as fit a VAR model to principal components. While this procedure does not allow one to disentangle factors from their lagged values, it can be used to forecast.

As mentioned above, this procedure is consistent with the recent literature on factor models. The crucial simplifying method is the use of PCA to identify factors. Adding additional features such as ARCH/GARCH or stochastic volatility would increase the parameter count and render the estimation process unreliable.

From the above it is clear that, with realistic N and T, the definition of "factors" depends on how we fix the threshold between "large" and "small" eigenvalues. For our purpose, in order to make a reasonable comparison of different strategies, we use the model:

$$\begin{aligned} &R_t = \beta F_t + \varepsilon_t \\ &\Phi(L)F_t = \eta_t \end{aligned} \tag{5}$$

where we estimate factors using PCA. Estimated factors are then fitted to a VAR model and factors are forecasted to produce fF_{t+fh}. The last step is to forecast expected returns using the first equation in (5):

$$fR_{t+fh} = \beta fF_{t+fh}. \tag{6}$$

5. Portfolio strategies for statistical arbitrage

In order to implement statistical arbitrage strategies, we create long-short portfolios based on forecasted returns; that is, we assume that our models create for each time t a vector of forecasted returns, denoted by fR_t, as follows

$$\begin{aligned} &fR_t = [fR_1(t), \ldots, fR_N(t)] \\ &fR_i(t) = \tfrac{fP_i(t+fh) - P_i(t)}{P_i(t)} \end{aligned} \tag{7}$$

where $fP_i(t+fh)$ are the forecasted prices at time $t + fh$. Portfolios are rebalanced using only the information contained in these forecasted returns.

We adopt the following notation: fh is the forecasting horizon, ti is the interval between trading days, and ew is the length of the estimation window. After performing a thorough analysis of different parameterizations (as detailed in Section 10 below) for the portfolio strategies we investigate in this paper, we use $ew = 1000$ and $fh = 5$; that is, every five days we estimate our models on those stock processes that existed in the prior 1000 trading days and forecast five days ahead. The trading interval is set equal to the forecasting horizon (i.e., $ti = fh$) because we believe it only makes sense to align the trading interval with the model forecast horizon. The choice of the model parameters ew and fh is discussed in Section 10.

In addition to long-short statistical arbitrage strategies, we also create long-only trading strategies and compare them to benchmark strategies. Next we discuss each of these strategies

5.1. Long-short strategies

We implement our statistical arbitrage strategies as long-short zero-cost portfolios as follows. The first long-short strategy invests in a long-short portfolio where the portfolio is long in the stocks corresponding to the k% highest forecasted returns and short in the stocks corresponding to the k% lowest returns. We denote this long-short strategy as "LS_*k*_*k*". For example, if $k = 15\%$, then the respective long-short strategy will be denoted as "LS_15_15". Since we implement two long-short strategies based on a factor model of prices and a factor model of returns we have the following two strategies: "PF_LS_*k*_*k*" – the *price*-based LS strategy and "RF_LS_*k*_*k*" – the *return*-based LS strategy. Those two strategies allocate equal weights among the forecasted winners and among the forecasted losers. We call them naïve strategies. We use $k = 10\%$; that is, "PF_LS_10_10" and "PF_LS_10_10" strategies are investigated in detail. The choice of the upper/lower bound k is discussed in Section 10.

The naïve strategies allocate equal weights among the selected winners ($L10$) and losers, ($S10$); We also run two types of optimizations – utility type optimization and maximum Sharpe ratio optimization.

The second long-short strategy optimizes as follows. For each time t we run a utility type optimization of the form

$$\max\{w : E(R_P(t+fh)) - \lambda \times ETL_\alpha(R_P(t+fh))\}$$

s.t.

$$\sum_{j\in L10} w_j = 1;\ \sum_{j\in S10} w_j = -1;\ w_j \leqslant w^*,\ j \in L10;\ w_j \geqslant -w^*,\ j \in S10.$$

Here the portfolio's expected return at time t for horizon $t + fh$ is given by $E(R_P(t+fh)) = \sum_{i=1}^{N} w_i fr_i(t)$ and the portfolio's Expected Tail Loss (ETL) at a confidence level α is given by $ETL_\alpha(R_P(t+fh)) = -E[R_P(t+fh)|R_P(t+fh) < q_\alpha(R_P(t+fh))]$. We set the risk-aversion parameter equal to five (i.e., $\lambda = 5$), the ETL confidence level α equal to 0.05, and the weight constraint w^* equal to 5%. The value of the risk-aversion parameter is empirically tested to provide a reasonable trade-off between expected return and risk. We call this optimization "*mean-ETL*". The choice of the weight constraint (w^*), confidence level (α), and risk aversion parameter (λ) is discussed in Section 10.

The third long-short strategy optimizes as follows. For each time t we run a maximum Sharpe ratio optimization

$$\max\left\{w : \frac{E(R_P(t+fh))}{\sigma(R_P(t+fh))}\right\}$$

s.t.

$$\sum_{j\in L10} w_j = 1; \sum_{j\in S10} w_j = -1; w_j \le w^*,\ j \in L10;\ w_j \geqslant -w^*,\ j \in S10,$$

where $\sigma_P(R_P(t+fh))$ is the portfolio forecasted volatility at time t and horizon fh. We refer to this optimization as "*max-Sharpe*".

The long-short "*mean-ETL*" optimization with weights constraint $w^* = 5\%$ will be denoted by "5_LS_O1_10_10". The *price*-based "*mean-ETL*" LS strategy is denoted by "PF_5_LS_O1_10_10" and the *return*-based "*mean-ETL*" LS strategy is denoted by "RF_5_LS_O1_10_10". The respective long-short "max-Sharpe" optimization with $w^* = 5\%$ will be denoted by "5_LS_O2_10_10". The *price*-based "max-Sharpe" LS strategy is denoted by "PF_5_LS_O2_10_10" and the *return*-based "max-Sharpe" LS strategy is denoted by "RF_5_LS_O2_10_10".

We compare the six long-short trading strategies described above with a benchmark strategy with weights proportional to the differences

$$R(t) - \bar{R}(t), \bar{R}(t) = \frac{1}{N}\sum_{i=1}^{N} R_i(t)$$

This benchmark strategy is the one adopted in Lehmann (1990) and Lo and MacKinlay (1990) that we will refer to as the "LS_REV" strategy because it is based on reversals of returns.

Theoretically, all our long-short strategies are zero-cost portfolios and do not require any investment. In practice, however, investors cannot use the entire proceeds of the short portfolio to buy the long portfolio because of margin requirements imposed by regulators. In addition, clients or regulatory imposed investment requirements may specify that the portfolio manager's strategy does not exceed a certain level of leverage. The strategy's return therefore depends on the amount of capital allocated to the strategy.

We compute the returns of long-short strategies under the assumption that we use capital, denoted by C, to buy a long portfolio, we short an amount equal to C without reinvesting the proceeds, and compute the strategy's return by dividing the sum of the profit of the long and short portfolios by C. To this end, we normalize a portfolio's weights so that the sum of the weights of the long portfolio is equal to 1 and the sum of those of the short portfolio is equal to –1. Then we compute the return of each strategy as the average return of the long portfolio minus the average return of the short portfolio.

5.2. Long-only strategies

In long-only strategies, only positive weights are allowed. The first long-only strategy invests in a long portfolio formed with the stocks corresponding to the k% highest forecasted returns. We denote this long only strategy as "LO_k". Further, "PF_LO_k" stands for the price-based naïve LO strategy and "RF_LO_k" stands for the return-based naïve LO strategy. After performing a thorough analysis of different parameterizations (as detailed in Section 10 below), we use the same combination of parameters $ew = 1000, fh = 5$, and $k = 10\%$. Therefore the following two strategies are investigated: "PF_LO_10" and "RF_LO_10". Again we try to improve the performance of the naïve strategies optimizing among the selected stocks.

The second strategy optimizes as follows. For each time t we run a utility type optimization of the form

$$\max\{w : E(R_P(t+fh)) - \lambda \times ETL_\alpha(R_P(t+fh))\}$$

s.t.

$$\sum_{j\in L10} w_j = 1; w_j \leqslant w^*, \quad j \in L10.$$

The third long only strategy optimizes as follows. For each time t we run maximum Sharpe ratio optimization

$$\max\left\{w : \frac{E(R_P(t+fh))}{\sigma(R_P(t+fh))}\right\}$$

s.t.

$$\sum_{j\in L10} w_j = 1; w_j \leqslant w^*, \quad j \in L10.$$

The long-only "*mean-ETL*" optimization with weights constraint $w^* = 5\%$ will be denoted by "5_LO_O1_10". The *price*-based "*mean-ETL*" LO strategy is denoted by "PF_5_LO_O1_10" and the *return*-based "*mean-ETL*" LO strategy is denoted by "RF_5_LO_O1_10". The respective long-only "max-Sharpe" optimization with $w^* = 5\%$ will be denoted by "5_LO_O2_10". The *price*-based "max-Sharpe" LO strategy is denoted by "PF_5_LO_O2_10" and the *return*-based "max-Sharpe" LO strategy is denoted by "RF_5_LO_O2_10".

We compare long-only strategies with a benchmark of an equally weighted portfolio. We refer to this strategy as LO_EW.

6. Methodologies to evaluate forecasts

We evaluate the forecasting ability of different families of dynamic factor models in two ways: (1) comparing the squared errors of forecasts obtained with the models under analysis and (2) applying the test for statistical arbitrage in Hogan et al. (2004).

6.1. The Diebold and Mariano framework

When considering the squared errors of forecasts, there is, of course, the need for a procedure to test the statistical significance of the squared errors of different strategies. We adopt the methodologies for testing the significance of prediction that Diebold and Mariano (1994) propose.

Diebold and Mariano test the null of no difference in the expectation of any function of forecasts. In our case, we want to test the null that there is no difference in the squared errors generated by our strategies. Because the squared error is a function of forecasts, we can use the Diebold–Mariano framework. Consider two different strategies j and i and their respective forecasts fR_t^j, fR_t^i at time t. If there are N stock price processes in the time window that ends at time t, the forecasts fR_t^j, fR_t^i are two N-vectors of N forecasted returns. The error of the strategy l in the period $(t, t+fh)$ is given by:

$$e_{t,l} = w_t^l(fR_t^l - R_t)$$

where w_t^l is the time t vector of loadings of strategy l; the vector is a nonlinear function of the forecasts fR_t^l (see Section 6.2 below) of strategy l. We consider the differentials of the squared errors: $k_t = e_{t,i}^2 - e_{t,j}^2$. Following Diebold and Mariano, we test the null that $E(k_t) = 0$ by first computing the average $\frac{1}{T}\sum_{t=1}^{T} k_t$ and considering the test statistics

$$S = \frac{m}{\sqrt{Ts^2(Q)}} \tag{8}$$

where $s^2(Q)$ is the long-run variance of k_t computed as a linear function of the autocovariances of k. We use the Bartlett weighting functions as follows:

$$s^2(Q) = s^2(0) + 2\sum_{\tau=1}^{Q} w_\tau \gamma_\tau$$

$$\gamma_\tau = \frac{1}{T-\tau}\sum_{l=\tau+1}^{T} k(l)k(l-\tau)$$

$$w_\tau = 1 - \frac{\tau}{Q+1}, \quad \tau = 1,\ldots,Q$$

The statistic S is distributed as a standard normal variable $S \cong N(0,1)$; we can therefore easily compute significance tests.

6.2. The statistical arbitrage test

We implemented the test for statistical arbitrage described in Hogan et al. (2004). This test explores the asymptotic behavior of profits from a statistical-arbitrage strategy when time goes to infinity. As no infinite time series of profits/losses is observable, these authors model the increments of a statistical arbitrage strategy.

To implement Hogan et al.'s test, at every trading date, we invest in our risky long-short portfolio (which is a zero-cost portfolio) and invest the profit or borrow the loss from the previous date in a risk-free money account. We then do the following: (1) compute the cumulative profits and the increments of cumulative profits, (2) discount increments, and (3) model the differences using Eq. (12) in Hogan et al. (2004) as follows: Let the discounted incremental trading profits satisfy: $\Delta v_i = \mu i^{\vartheta} + \sigma i^{\lambda} z_i$ for $i = 1, 2, \ldots, n$, where z_i are i.i.d $N(0,1)$ random variables with $z_0 = 0$ and $\Delta v_i = 0$. Next we compute the min t-statistics. We applied this test to all long-short strategies based both on return factors and price factors. All strategies pass the test in that the min t-statistics exceed the critical values.

7. Determining the number of factors

Dynamic factor models can be applied only under the condition that returns or prices exhibit a number of factors much smaller than the number of series themselves. In Section 7.2 below we argue that prices are mean-reverting around a single integrated factor. We therefore need to determine the number of stationary factors of models of returns and of prices. We will first determine the number of stationary factors of returns and then the number of stationary factors of prices.

7.1. Number of stationary factors of returns

We assume that our sample can be described by an approximate factor structure as described in Bai and Ng (2002). When discussing the number of factors of stationary processes, we need to distinguish between stationary and dynamic factors. The number of dynamic factors is less than or equal to the number of static factors. Therefore, the number of static factors is an upper bound for the number of factors. Various criteria have been proposed for determining the number of static factors of factor models of stationary processes. In particular, for large factor models when N and T are of comparable size, Bai and Ng (2002) suggest a criterion based on extending the AIC and BIC to an infinite N, infinite T framework; Onatski (2009) and Kapetanios (2010) suggest criteria based on the theory of random matrices.

Proposing a criterion which holds in the limit of infinite N and infinite T, Bai and Ng (2002) show that in the limit of infinite N and infinite T, factors can be estimated using PCA. The number of factors is determined as a model selection problem defining information criteria in the nonstandard asymptotic limit of infinite N and infinite T. Among the different criteria, we choose $IC_{p2}(k)$, which minimizes the following quantities over the number of factors k:

$$IC_{p2}(k) = \log(V(k)) + k\left(\frac{N+T}{NT}\right)\log(\min(N,T))$$

$$V(k) = \frac{1}{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \sum_{q=1}^{k}\lambda_{iq}F_{q,t}\right)\right)$$

The Bai–Ng criteria hold asymptotically assuming that both the number of processes N and the time T tend to infinity. In finite samples, we can interpret the Bai–Ng criteria as criteria to choose the optimal number of principal components.

Kapetanios (2010) takes a different approach based on random matrix theory based on the following reasoning. The eigenvalues of the correlation matrix of independent standard normal variables are all equal to 1. However, the eigenvalues of the empirical correlation matrix of a series of independent samples occupy a region around unity. It is well known from random matrix theory that the limit distribution of these eigenvalues when N tends to infinity follows the Marčenko–Pastur law. In particular, the largest and the smallest eigenvalue are:

$$\lambda_{\max} = \left(1+\sqrt{\frac{N}{T}}\right)^2, \lambda_{\min} = \left(1-\sqrt{\frac{N}{T}}\right)^2$$

Kapetanios (2010) adopts a sequential procedure as follows. First, compute the correlation matrix C of the data and compare the largest eigenvalue $\lambda_{C,\max}$ of the correlation matrix C with a number larger than $\left(1+\sqrt{\frac{N}{T}}\right)^2$. Kapetanios selected $\left(1+\sqrt{\frac{N}{T}}\right)^2 + 1$ so in this case, if $\lambda_{C,\max} > \left(1+\sqrt{\frac{N}{T}}\right)^2 + 1$, then there is at least one factor. Then regress data on the first principal component, normalize residuals, compute the eigenvalues of the correlation matrix of residuals, and repeat the same procedure until the largest eigenvalue at step j is less than $\left(1+\sqrt{\frac{N}{T}}\right)^2 + 1$.

This procedure compares the eigenvalues of the empirical correlation matrix with the benchmark distribution of eigenvalues relative to independent and identically distributed variables. However, there are local correlation structures that cannot be associated with pervasive factors but still do not produce eigenvalues equal to 1 and whose limit distribution does not follow the Marčenko–Pastur law. To overcome these limitations, Kapetanios (2010) and Onatski (2009) propose more complicated tests based on random matrix theory. However, Kapetanios (2010) demonstrates that these new tests produce results that are quite similar to the test we described above.

Applying the two tests just described, we obtain the numbers of factors identified in the following table:

	Time window				
Number of factors based on	1	2	3	4	5
Kapetanios (2010)	8	11	16	43	57
Bai and Ng (2002)	3	3	5	7	6

The two methods yield a significantly different number of factors; actually the two methods estimate different quantities. The Bai–Ng method computes the optimal number of principal components as a model selection problem. It implements a trade-off between model complexity and the average size of residuals. Asymptotically this method yields the correct number of factors. However, in finite samples it gives the optimal number of principal components. In contrast, the method proposed by Kapetanios establishes how many factors cannot be considered the result of pure noise.

7.2. Number of integrated factors

In principle, the number of integrated factors can be determined with any of the methods proposed to determine the number of cointegrating relationships. However, only a few methods can be applied to a large number of stock price processes. We adopt the methodology presented in Nyblom and Harvey (2000).

The starting point of the Nyblom–Harvey methodology is a test proposed by Kwiatkowski et al. (1992) – hereafter referred to as "KPSS" – to determine if a univariate series is I(0) against the alternative I(1). The model for the univariate series is the sum of a random walk u_t plus a stationary component η_t

$$\begin{aligned} &X_t = \mu + u_t + \eta_t, \eta_t \approx N(0, \sigma_\eta^2) \\ &u_t = u_{t-1} + \varepsilon_t, \varepsilon_t \approx N(0, \sigma_\varepsilon^2) \end{aligned} \tag{9}$$

Assume for the moment that both η_t and ε_t are i.i.d. normally distributed and mutually independent variables. Defining the deviation from the mean as $e_t = X_t - \bar{X}$, the KPSS statistic is:

$$S = \frac{1}{T^2}\sum_{i=1}^{T}\left[\sum_{t=1}^{i} e_t\right]^2 \Bigg/ s^2 \tag{10}$$

where $s^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$ is an estimate of the variance of e_t. KPSS show that the asymptotic distribution of S is a functional of Brownian motions whose distribution belongs to the family of Cramèr-von Mises distributions. If $\sigma_\eta^2 = 0$, then the random-walk component becomes a constant and the process X_t is stationary. Hence, the KPSS test for $\sigma_\eta^2 = 0$ against the alternative $\sigma_\eta^2 > 0$ using the upper tail test $S > c$ where c is a critical value free from nuisance parameters.

The assumption that ε_t is an i.i.d. sequence is too strong in applied work. KPSS show that if residuals are autocorrelated, the test remains the same provided that the estimate of the variance $s^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$ is replaced with the Newey–West estimator of the long-run variance which KPSS write as follows:

$$s^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 + 2\left[\sum_{i=1}^{s}(\psi(i,s)\sum_{t=i+1}^{T} e_t e_{t-i})\right] \tag{11}$$

The weighting functions might have different forms. KPSS use the following expression: $\psi(i,s) = [1 - i/(s+1)]$.

Nyblom and Harvey (2000) generalize this test in a multivariate setting by considering the multivariate model:

$$\begin{aligned} &X_t = \mu + u_t + \eta_t, \eta_t \approx N(0, \Sigma_\eta) \\ &u_t = u_{t-1} + \varepsilon_t, \varepsilon_t \approx N(0, \Sigma_\varepsilon) \end{aligned} \tag{12}$$

where the process X is a N-variate process and both η_t and ε_t are serially uncorrelated and mutually uncorrelated at every lag. Initially, Nyblom and Harvey assume that $\Sigma_\eta = q\Sigma_\varepsilon$ and test the null $q = 0$ against the alternative $q > 0$. The test statistic is:

$$\xi_N = tr(S^{-1}C) \tag{13}$$

which is equal to the sum of the eigenvalues of $S^{-1}C$, where

$$S = \frac{1}{T}\sum_{t=1}^{T} e_t' e_t + \frac{2}{T}\left[\sum_{i=1}^{s}\left(\psi(i,s)\sum_{t=i+1}^{T}(e_t' e_{t-i})\right)\right] \tag{14}$$

is an estimate of the long-run covariance of the process X_t and the matrix C is defined as:

$$C = \frac{1}{T^2}\sum_{i=1}^{T}\left[\sum_{t=1}^{i} e_t\right]\left[\sum_{t=1}^{i} e_t\right]' \tag{15}$$

The proposed test is $\xi_N = tr(S^{-1}C) > c$, where c is a critical value free from nuisance parameters. This test is based on the surprising property that the limiting distribution of a number of statistics related to the matrix $(S^{-1}C)$ depends only on the rank of this matrix.

Nyblom and Harvey demonstrate that their test has maximum power against the alternative $\Sigma_\eta = q\Sigma_\varepsilon$ and that it is consistent for any non-zero Σ_η. They also show that the model will exhibit $K < N$ trends if the rank of the matrix Σ_η is K. In this case, the limiting distribution of the statistic $\xi_N = tr(S^{-1}C)$ will be dominated by the K largest eigenvalues.

Based on this property, the methodology that Nyblom and Harvey propose to test the hypothesis that the rank of Σ_η is $K < N$ against the alternative that the rank of $\Sigma_\eta > K$ is based on the sum $\xi_{K,N}$ of the smallest $N - K$ eigenvalues of the matrix $S^{-1}C$ defined as above. Harvey and Nyblom show that, in all cases, the test statistic $\xi_{K,N}$ converges to functionals of Brownian motions and determine analytic expressions for their limit distributions as infinite series:

$$\begin{aligned} &T^{-1}\xi_N \xrightarrow{d} tr\left(\sum_{k=1}^{\infty}(\pi k)^{-2} v_k v_k'\right)^{-1}\left(\sum_{k=1}^{\infty}(\pi k)^{-4} v_k v_k'\right) \\ &v_k \propto \mathrm{NID}(0, I_K) \text{ is a } K \times 1 \text{ vector} \\ &\xi_{K,N} = \sum_{i=1}^{N-K}\lambda_i \xrightarrow{d} \sum_{k=1}^{\infty}(\pi k)^{-2} u_k' u_k tr\left(\sum_{k=1}^{\infty}(\pi k)^{-3} u_k v_k'\right)\left(\sum_{k=1}^{\infty}(\pi k)^{-4} v_k v_k'\right)^{-1} \\ &\quad \times \left(\sum_{k=1}^{\infty}(\pi k)^{-3} v_k u_k'\right) \\ &v_k \propto \mathrm{NID}(0, I_K) \text{ is a } K \times 1 \text{ vector, } u_k \propto NID(0, I_K) \text{ is a } r \times 1 \text{ vector,} \\ &\quad r = N - K \\ &[\lambda] = eig(S^{-1}C), \lambda_N \geqslant \cdots \geqslant \lambda_1 \end{aligned} \tag{16}$$

Let's now discuss the results of applying the Harvey–Nyblom test to the S&P 500 universe for the five 4-year windows. For each time window, we consider only those components that were in the S&P 500 and that existed throughout the entire time window.

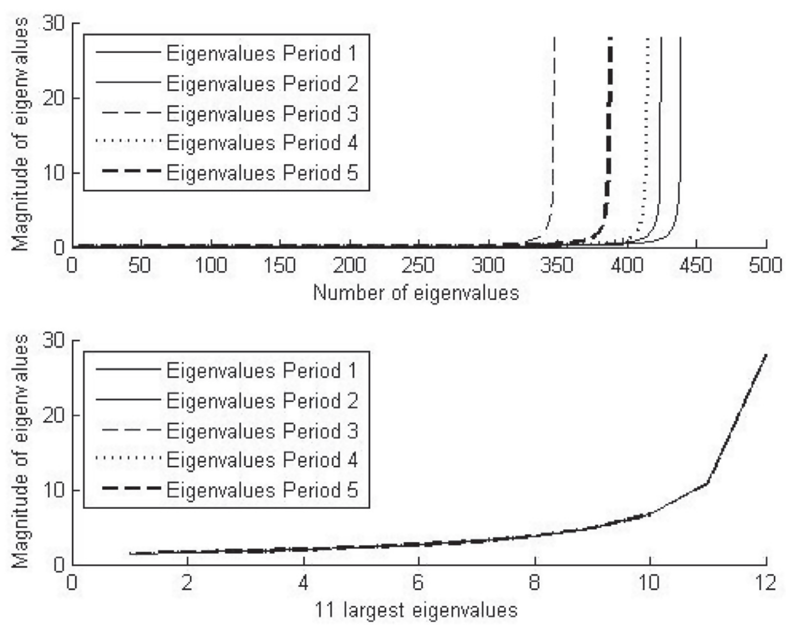
For each time window, we compute the test statistic ξ_N and the test statistic $\xi_{1,N}$. We compute the long-run covariance matrix using equation (14) with the number of lags $s = 6$,

As expected, ξ_N, which is equal to the trace of the matrix $S^{-1}C$, is approximately equal for all time windows. For each time window, Panel A of Table 1 gives the test statistics ξ_N while Panel B of Table 1 gives the test statistics $\xi_{1,N}$ which is equal to the sum of the N-1 smallest eigenvalues:

$$\sum_{i=1}^{N-1}\lambda_i, [\lambda] = eig(S^{-1}C), \lambda_N \geqslant \cdots \geqslant \lambda_1$$

Table 1
Test statistics ξ_N and test statistics $\xi_{1,N}$ and the relative critical values.

Period	03/Jan/1989–14/Dec/1992	15/Dec/1992–26/Nov/1996	11/26/1996–13/Nov/2000	11/14/2000–09/Nov/2004	11/10/2004–10/Nov/2004
Panel A. Test statistics ξ_N and the relative critical values $\xi_N > crit.val.$ rejects null of joint stationarity. The null of joint stationarity can be rejected for each time window.					
ξ_N	129.9894	129.956	129.8191	129.9496	129.8976
Crit. Val. 99%	76.6591	74.1949	63.0397	72.704	68.1744
Crit. Val. 95%	74.3675	71.905	60.9613	70.4272	66.0304
Crit. Val. 90%	73.1476	70.7243	59.8663	69.2705	64.8832
N series	439	425	358	416	389
Panel B. Test statistics $\xi 1_{,N}$ and the relative critical values; $\xi_{1,N} > crit.val.$ rejects null of 1 common trend against more than one common trend. The null of 1 common trend against more than one common trend cannot be rejected for any time window.					
$\xi_{1,N}$	50.9736	50.9411	50.8047	50.9353	50.883
Crit. Val. 99%	63.0469	61.0805	51.544	59.7131	55.8781
Crit. Val. 95%	57.5796	55.6083	46.8482	54.4494	50.8655
Crit. Val. 90%	52.6516	50.985	42.782	49.8544	46.4682
N series	439	425	358	416	389



Period 1	Period 2	Period 3	Period 4	Period 5
03-Jan-1989-14-Dec-1992	15-Dec-1992-26-Nov-1996	27-Nov-1996-13-Nov-2000	14-Nov-2000-09-Nov-2004	10-Nov-2004-29-Oct-2008

Figure 1. Decay pattern of eigenvalues.

As no table of critical values is available for the number of series that we are considering, we compute critical values for each time window by simulating the time series expansion given by (16) with the number of time series of each time window. Panels A and B of Table 1 report the critical values for each time window respectively for ξ_N and $\xi_{1,N}$. The reported critical values are very similar but not equal for different time windows due to the different number of processes in each time window.

If we compare the test statistic ξ_N with critical values, we see that we can reject the null assumption of stationarity of logprice processes in each time window at confidence levels of 90%, 95%, and 99%. If we compare the test statistic $\xi_{1,N}$ with critical values, we see that we cannot reject the null assumption of one integrated trend against more than one integrated trend in any time window at confidence levels of 90%, 95%, and 99%.

The matrices C and S are computed by applying formulas (9) and (10) to logprice processes in the five non-overlapping 1000-step time windows. In Fig. 1, the upper panel shows the plots of all eigenvalues of matrix $S^{-1}C$ and the lower panel shows the 11 largest eigenvalues of the same matrix. Note that the 11 largest eigenvalues are practically equal for all five time windows.

As discussed in Section 7.1, there is no rigorous criterion to determine the number of factors in a finite sample. Therefore, estimating the "true" number of factors is essentially a model selection exercise. We thus opt for a model selection framework which allows us to optimize the entire model, choosing the optimal number of factors and time lags, and counting (in the optimization process) the global number of the model's parameters. More specifically, we create different models — either autoregressive models for each factor or full VAR models — with 5, 10, and 15 principal components, and using from 1 to 5 time lags. We select the optimal model using the BIC. Optimal models thus selected use 15 factors with independent autoregressive processes and one lag in the case of price factors, and 15 factors with full VAR and one lag for return factors. Out-of-sample testing confirmed that these models actually give the best performance in terms of forecasting.

8. Can factor models of prices and of returns coexist theoretically?

The question as to whether dynamic factor models of prices and dynamic factor models of returns can theoretically coexist is an instance of a more general question that, though important, has not received much attention in the literature. The argument is the following.

Consider the logprices and logarithm of returns (logreturns) – that is the first difference of logprices. For simplicity, we will refer to them as prices and returns. Suppose that returns can be represented through a dynamic factor model as described in Section 3. A dynamic factor model is a parsimonious way to write a VAR model.

It is well known (see, for example, Hamilton 1994, p. 573) that if a multivariate integrated process y_t is cointegrated, then the Wald representation of the first differences Δy_t is not invertible. It is generally believed that if a process is not invertible, no VAR model describing the process is properly specified and consequently cannot be estimated. If this is true, it implies that, if prices are cointegrated, no dynamic factor model of returns can be estimated.

But prices appear to be cointegrated. The assumption of cointegrated prices appears in Bossaerts (1988) and we show that the prices of stocks comprising the S&P 500 index admit one integrated factor. This implies that prices are indeed cointegrated. Therefore, according to generally accepted econometric principles, no factor model of returns can be reasonably used for forecasting returns using autoregressive techniques. The problem is not alleviated if factors are used as predictors. In fact, regressing returns over lagged factors is equivalent to estimating a dynamic factor model. Still, factor models of returns are described in the literature (see, e.g., Deistler and Hamann, 2005) and widely used in practice.

Do we have to conclude that using factor models of returns to forecast returns is fundamentally wrong? Not necessarily, according to Marcet (2005) who argues that applying and estimating VAR models to "overdifferenced" variables yields correct results provided that a sufficient number of lags are used. He concludes that estimating VARs in differences is an optimal strategy even in the presence of cointegration or when the level of integration is not certain. Based on Marcet's results, we can conclude that factor models of prices and factor models of returns can coexist and should ultimately give the same results.

However, given the size of available samples of prices and returns, we would argue that modeling prices yields better results than modeling returns. Detailed comparisons between the two classes of models will be given in Section 10. In fact, factor models of prices have a much smaller number of factors and use a smaller number of lags. In contrast, because factor models of returns use many more factors and lags, they therefore require estimation with a much larger number of parameters. As a consequence, estimates of factor models of returns are noisier than the estimates of factor models of prices.

9. Performance decay with the forecasting horizon and the bid-ask spread

Our models are based on daily returns. The reason is that longer spacing would have reduced the number of data points and, consequently, the ability to estimate autoregressive models. If the forecasting horizon is one day and if rebalancing is done daily, daily returns present two main problems: (1) the bid-ask spread might create illusory profits in backtesting and (2) transaction costs would be high.

The problem arising from the effects of the bid-ask spread is discussed in Lehmann (1990). The CRSP files do not distinguish if the closing prices are bids or offers. Lehman explains how the bid-ask spread will influence the reversal strategy by creating illusory profits, impossible to exploit in practice. One of the remedies that he proposes is to trade weekly rather than daily.

The influence of the bid-ask spread on factor-based strategies is difficult to gauge. However, one indication of the magnitude of the bid-ask spread on different strategies comes from the decay in performance when the forecasting horizons are longer. In this paper, we use a 5-day forecasting horizon. Results are reported and discussed in Section 10.

Another reason to explore the decay of performance associated with longer forecasting horizons is the impact of transaction costs. All our strategies produce a very high turnover. Lengthening the forecasting horizon generally reduces turnover and thus transaction costs, but performance decays.

10. Results

In this section we analyze the performance of our statistical arbitrage long-short strategies and compare their performance to that of long-only strategies. We test strategies with 15 factors and a five-day forecasting horizon and compare results of factor models with benchmark strategies. All models were estimated using daily prices. We compute the following quantities: (1) annualized return, (2) annual turnover, (3) annualized Sharpe ratio, (4) breakeven transaction cost (i.e., the transaction costs that would produce zero returns), and (5) maximum drawdown. The significance test for each pair of the three types of forecasts for stocks returns based on the Diebold–Mariano framework is also computed. A simulation experiment is performed in order to gain additional intuition about the model. We test the exposure of the strategies to the four-factor model formulated by Carhart (1997), and test for statistical arbitrage using the test suggested by Hogan et al. (2004).

10.1. Preliminary analysis of the models parameters

To test the robustness of our strategies to parameter changes, we test the performance of our strategies with 54 ($= 3 \times 3 \times 6$) combinations of parameters ew, fh and k: three different values for the estimation window, $ew = 750$, 1000, and 1250 days, three different values for the forecasting horizon, $fh = 1$, 5, and 10 days, and six values of k, $k = 5\%$, 10%, 15%, 20%, 25%, and 50%. In order to have more observations than variables, we use daily data.

Tables 2–4 report the performance of the naive long-short and long-only strategies with estimation windows $ew = 750$, 1000, 1250 respectively. Looking at a given estimation window and forecasting horizon, one can see that the performance of the price-based LS strategy (PF_LS_*k*_*k*) decreases when the bound k increases; the same is true for the return-based LS strategy (RF_LS_*k*_*k*) but the decrease is not so pronounced. This fact is, of course, expected since when k increases the importance of the forecasts ranking decreases.

Looking at a given estimation window and given bound k, one can see that the performance of the PF_LS_*k*_*k* strategy decreases when the forecasting horizon increases. However, the break-even transaction cost (BETC) for the daily forecasting horizon is almost one half of the BETC for the weekly forecasting horizon, and the BETC for weekly and bi-weekly horizons are almost equal. On the other hand, the annualized Sharpe ratio for the weekly forecasting horizon is approximately 50% higher than the annualized Sharpe ratio for the bi-weekly forecasting horizon having almost equal BETC. Furthermore, maximum drawdown decreases when increasing k for all the horizons. The effect of the estimation window on the overall results seems to be insignificant.

Table 2
Out-of-sample results for long-short and long-only naïve strategies using 15 factors and a 750-day estimation window: Results for performance tests, returns, turnover, annualized Sharpe ratio, breakeven transaction costs and maximum drawdown for k = 5%, 10%, 15%, 20%, 25%, and 50% and fh = 1, 5, and 10 days.

	ew = 750											
	LS strategies						LO strategies					
	Strategy name	Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)	Strategy name	Annualized return (%)	Annual turnover	Annualized sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
fh = 1	PF_LS_05_05	69.592	749.01	2.5502	0.0009	28.136	PF_LO_05	42.407	371.53	1.1809	0.0011	59.058
	RF_LS_05_05	37.926	1307.71	1.0667	0.0003	77.477	RF_LO_05	30.681	655.32	0.8690	0.0005	85.256
	PF_LS_10_10	51.527	721.68	2.8535	0.0007	17.108	PF_LO_10	36.919	359.48	1.1888	0.0010	55.187
	RF_LS_10_10	36.899	1267.41	1.2323	0.0003	70.913	RF_LO_10	31.974	635.30	0.9943	0.0005	79.675
	PF_LS_15_15	43.941	702.01	3.0321	0.0006	13.269	PF_LO_15	35.084	350.48	1.2085	0.0010	51.336
	RF_LS_15_15	34.642	1238.54	1.3522	0.0003	64.460	RF_LO_15	30.339	620.33	1.0179	0.0005	73.883
	PF_LS_20_20	37.228	685.29	2.9455	0.0005	12.063	PF_LO_20	32.117	342.54	1.1609	0.0009	51.148
	RF_LS_20_20	31.984	1210.76	1.4243	0.0003	53.949	RF_LO_20	28.915	605.99	1.0255	0.0005	68.666
	PF_LS_25_25	31.899	670.55	2.8753	0.0005	11.655	PF_LO_25	29.492	335.36	1.1064	0.0009	51.898
	RF_LS_25_25	28.976	1182.00	1.4447	0.0002	49.617	RF_LO_25	27.812	591.22	1.0282	0.0005	65.542
	PF_LS_50_50	18.889	611.14	2.5691	0.0003	9.024	PF_LO_50	22.356	305.99	0.9151	0.0007	48.668
	RF_LS_50_50	18.105	990.18	1.3744	0.0002	29.981	RF_LO_50	22.180	495.50	0.9150	0.0004	59.638
	LS_REV	15.018	1146.51	0.5046	0.0001	81.321	LO_EW	12.291	252.24	0.5020	0.0005	57.945
fh = 5	PF_LS_05_05	25.307	204.97	1.1811	0.0012	22.370	PF_LO_05	22.409	101.79	0.7208	0.0022	58.337
	RF_LS_05_05	15.001	265.25	0.4914	0.0006	51.341	RF_LO_05	24.626	132.94	0.7278	0.0019	54.339
	PF_LS_10_10	21.728	194.95	1.3373	0.0011	14.865	PF_LO_10	22.615	97.28	0.8027	0.0023	60.879
	RF_LS_10_10	13.268	257.66	0.4856	0.0005	45.404	RF_LO_10	21.987	129.08	0.7352	0.0017	55.572
	PF_LS_15_15	19.491	186.76	1.3962	0.0010	13.112	PF_LO_15	22.961	93.33	0.8613	0.0025	56.026
	RF_LS_15_15	11.503	252.22	0.4627	0.0005	40.206	RF_LO_15	19.606	126.07	0.7055	0.0016	52.102
	PF_LS_20_20	16.958	179.91	1.3706	0.0009	10.472	PF_LO_20	21.552	90.04	0.8403	0.0024	53.549
	RF_LS_20_20	10.769	246.90	0.4684	0.0004	33.548	RF_LO_20	19.117	123.42	0.7216	0.0015	52.348
	PF_LS_25_25	15.483	173.61	1.3705	0.0009	8.939	PF_LO_25	20.796	87.04	0.8308	0.0024	52.635
	RF_LS_25_25	9.852	241.26	0.4502	0.0004	29.750	RF_LO_25	17.793	120.60	0.6959	0.0015	52.770
	PF_LS_50_50	8.994	148.31	0.9722	0.0006	7.799	PF_LO_50	16.870	74.27	0.7221	0.0023	49.344
	RF_LS_50_50	4.998	201.83	0.1740	0.0002	23.600	RF_LO_50	14.851	101.02	0.6282	0.0015	53.856
	LS_REV	20.858	231.26	0.7503	0.0009	33.521	LO_EW	11.979	50.58	0.5077	0.0024	55.388
fh = 10	PF_LS_05_05	13.574	119.38	0.6515	0.0011	29.615	PF_LO_05	17.893	59.74	0.6252	0.0030	62.867
	RF_LS_05_05	3.977	130.17	0.1490	0.0003	67.313	RF_LO_05	15.828	65.47	0.5351	0.0024	79.202
	PF_LS_10_10	9.439	113.06	0.5123	0.0008	24.794	PF_LO_10	16.563	56.52	0.6305	0.0029	60.692
	RF_LS_10_10	3.979	126.36	0.1209	0.0003	50.787	RF_LO_10	14.949	63.68	0.5511	0.0023	73.778
	PF_LS_15_15	9.734	108.53	0.6285	0.0009	19.004	PF_LO_15	16.618	54.30	0.6607	0.0031	58.959
	RF_LS_15_15	2.232	123.45	0.0170	0.0002	50.173	RF_LO_15	14.063	62.22	0.5425	0.0023	64.563
	PF_LS_20_20	10.197	104.49	0.7611	0.0010	13.648	PF_LO_20	17.229	52.37	0.7069	0.0033	57.531
	RF_LS_20_20	2.128	120.94	−0.0145	0.0002	43.238	RF_LO_20	13.601	60.82	0.5467	0.0022	61.606
	PF_LS_25_25	9.655	100.46	0.7516	0.0010	12.208	PF_LO_25	16.988	50.34	0.7136	0.0034	55.668
	RF_LS_25_25	2.141	118.31	−0.0315	0.0002	40.414	RF_LO_25	13.078	59.44	0.5385	0.0022	59.058
	PF_LS_50_50	6.720	83.38	0.5817	0.0008	7.556	PF_LO_50	15.210	41.76	0.6763	0.0036	51.868
	RF_LS_50_50	0.982	99.33	−0.2579	0.0001	29.468	RF_LO_50	12.296	49.70	0.5356	0.0025	56.150
	LS_REV	4.322	114.64	0.1160	0.0004	42.161	LO_EW	11.582	25.38	0.5073	0.0046	55.097

PF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with price-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
RF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with return-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
PF_LO_*k*: Long-only strategy based on forecasts obtained with price-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
RF_LO_*k*: Long-only strategy based on forecasts obtained with return-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
LS_REV: Long-short reversal strategy. Weights are proportional to the opposite of the deviation of the corresponding returns from the mean in the previous period.
LO_EW: Long-only strategy. Portfolio is rebalanced at every trading date so that weights are all equal and sum to 1.

Table 3
Out-of-sample results for long-short and long-only naïve strategies using 15 factors and a 1000-day estimation window: Results for performance tests, returns, turnover, annualized Sharpe ratio, breakeven transaction costs and maximum drawdown for k = 5%, 10%, 15%, 20%, 25%, and 50% and fh = 1, 5, and 10 days.

	ew = 1000											
	LS strategies						LO strategies					
	Strategy name	Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)	Strategy name	Annualized return (%)	Annual turnover	Annualized sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
fh = 1	PF_LS_05_05	66.402	719.22	2.4041	0.0009	29.956	PF_LO_05	39.018	356.98	1.0854	0.0011	53.911
	RF_LS_05_05	32.676	1301.86	0.9376	0.0003	76.021	RF_LO_05	23.831	652.93	0.7096	0.0004	83.647
	PF_LS_10_10	51.891	695.18	2.8205	0.0007	18.252	PF_LO_10	36.254	347.35	1.1578	0.0010	51.895
	RF_LS_10_10	33.101	1259.95	1.1196	0.0003	71.489	RF_LO_10	26.299	632.27	0.8499	0.0004	79.083
	PF_LS_15_15	41.927	676.21	2.8621	0.0006	16.447	PF_LO_15	32.190	337.58	1.1103	0.0010	48.852
	RF_LS_15_15	33.962	1230.74	1.3256	0.0003	63.420	RF_LO_15	28.012	617.45	0.9612	0.0005	74.804
	PF_LS_20_20	35.027	660.98	2.7615	0.0005	13.879	PF_LO_20	28.816	330.28	1.0439	0.0009	48.846
	RF_LS_20_20	31.789	1203.49	1.4089	0.0003	55.216	RF_LO_20	28.002	602.91	1.0056	0.0005	69.707
	PF_LS_25_25	30.640	648.54	2.7213	0.0005	12.195	PF_LO_25	26.816	324.09	1.0034	0.0008	48.729
	RF_LS_25_25	28.996	1175.26	1.4362	0.0002	46.858	RF_LO_25	26.706	588.37	0.9960	0.0005	67.587
	PF_LS_50_50	16.666	595.67	2.1660	0.0003	9.402	PF_LO_50	20.173	298.26	0.8201	0.0007	48.835
	RF_LS_50_50	16.929	985.91	1.2553	0.0002	33.449	RF_LO_50	20.609	493.35	0.8489	0.0004	60.761
	LS_REV	16.055	1147.00	0.5306	0.0001	81.334	LO_EW	11.344	252.23	0.4573	0.0004	57.870
fh = 5	PF_LS_05_05	29.143	192.25	1.3225	0.0015	25.988	PF_LO_05	21.687	95.74	0.6826	0.0023	65.035
	RF_LS_05_05	10.386	262.29	0.3576	0.0004	53.728	RF_LO_05	19.969	131.39	0.6305	0.0015	63.482
	PF_LS_10_10	25.221	183.20	1.4994	0.0014	21.006	PF_LO_10	22.637	91.56	0.7883	0.0025	57.797
	RF_LS_10_10	11.365	255.26	0.4197	0.0004	41.515	RF_LO_10	19.577	127.46	0.6801	0.0015	64.498
	PF_LS_15_15	21.202	175.75	1.4893	0.0012	18.565	PF_LO_15	21.592	87.87	0.7967	0.0025	54.143
	RF_LS_15_15	10.903	249.71	0.4391	0.0004	34.381	RF_LO_15	18.563	124.77	0.6871	0.0015	62.496
	PF_LS_20_20	17.405	169.69	1.3858	0.0010	17.449	PF_LO_20	20.141	84.86	0.7742	0.0024	54.381
	RF_LS_20_20	10.251	244.43	0.4431	0.0004	27.166	RF_LO_20	17.265	122.23	0.6640	0.0014	61.417
	PF_LS_25_25	15.474	164.21	1.3509	0.0009	14.270	PF_LO_25	19.568	82.19	0.7729	0.0024	53.274
	RF_LS_25_25	9.199	239.06	0.4135	0.0004	23.290	RF_LO_25	16.336	119.69	0.6445	0.0014	60.561
	PF_LS_50_50	9.024	141.56	0.9762	0.0006	7.623	PF_LO_50	15.869	70.88	0.6709	0.0022	50.498
	RF_LS_50_50	5.079	201.03	0.1817	0.0003	20.775	RF_LO_50	14.019	100.61	0.5923	0.0014	55.130
	LS_REV	20.460	231.17	0.7142	0.0009	37.516	LO_EW	11.078	50.57	0.4631	0.0022	55.262
fh = 10	PF_LS_05_05	19.106	112.24	0.9297	0.0017	28.141	PF_LO_05	19.274	56.28	0.6749	0.0034	62.129
	RF_LS_05_05	1.592	130.32	0.0709	0.0001	69.827	RF_LO_05	12.041	65.35	0.4252	0.0018	76.238
	PF_LS_10_10	15.998	106.51	0.9834	0.0015	22.331	PF_LO_10	18.032	53.36	0.6854	0.0034	57.856
	RF_LS_10_10	0.704	126.44	−0.0158	0.0001	55.508	RF_LO_10	12.068	63.47	0.4532	0.0019	70.189
	PF_LS_15_15	13.480	101.94	0.9418	0.0013	20.241	PF_LO_15	16.848	51.05	0.6665	0.0033	58.193
	RF_LS_15_15	0.945	123.60	−0.0482	0.0001	51.633	RF_LO_15	12.160	62.02	0.4769	0.0020	65.334
	PF_LS_20_20	10.889	97.99	0.7999	0.0011	15.773	PF_LO_20	15.970	49.09	0.6456	0.0033	58.151
	RF_LS_20_20	2.425	120.94	0.0035	0.0002	44.393	RF_LO_20	12.652	60.63	0.5152	0.0021	62.332
	PF_LS_25_25	10.949	94.57	0.8849	0.0012	14.913	PF_LO_25	16.085	47.27	0.6699	0.0034	55.889
	RF_LS_25_25	1.890	117.97	−0.0515	0.0002	38.959	RF_LO_25	12.258	59.12	0.5087	0.0021	60.526
	PF_LS_50_50	7.063	79.14	0.6179	0.0009	6.459	PF_LO_50	14.480	39.64	0.6372	0.0037	52.333
	RF_LS_50_50	2.004	99.18	−0.1401	0.0002	29.579	RF_LO_50	12.002	49.64	0.5255	0.0024	55.356
	LS_REV	5.273	114.65	0.1653	0.0005	38.490	LO_EW	10.701	25.37	0.4630	0.0042	54.977

PF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with price-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
RF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with return-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
PF_LO_*k*: Long-only strategy based on forecasts obtained with price-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
RF_LO_*k*: Long-only strategy based on forecasts obtained with return-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
LS_REV: Long-short reversal strategy. Weights are proportional to the opposite of the deviation of the corresponding returns from the mean in the previous period.
LO_EW: Long-only strategy. Portfolio is rebalanced at every trading date so that weights are all equal and sum to 1.

Table 4
Out-of-sample results for long-short and long-only naïve strategies using 15 factors and a 1250-day estimation window: Results for performance tests, returns, turnover, annualized Sharpe ratio, breakeven transaction costs and maximum drawdown for k = 5%, 10%, 15%, 20%, 25%, and 50% and fh = 1, 5, and 10 days.

	ew = 1250											
	LS strategies						LO strategies					
	Strategy name	Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)	Strategy name	Annualized return (%)	Annual turnover	Annualized sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
fh = 1	PF_LS_05_05	61.659	700.14	2.2404	0.0009	24.981	PF_LO_05	37.445	346.90	1.0292	0.0011	59.629
	RF_LS_05_05	27.245	1302.80	0.7938	0.0002	78.766	RF_LO_05	23.240	653.86	0.6942	0.0004	81.880
	PF_LS_10_10	48.385	677.61	2.6179	0.0007	20.229	PF_LO_10	34.221	337.33	1.0841	0.0010	52.674
	RF_LS_10_10	27.806	1261.38	0.9494	0.0002	73.606	RF_LO_10	26.016	633.75	0.8394	0.0004	78.578
	PF_LS_15_15	41.234	660.73	2.7330	0.0006	17.303	PF_LO_15	31.268	329.49	1.0632	0.0009	50.094
	RF_LS_15_15	27.415	1232.84	1.0811	0.0002	66.721	RF_LO_15	25.532	618.42	0.8799	0.0004	75.391
	PF_LS_20_20	33.334	646.76	2.5843	0.0005	13.910	PF_LO_20	26.940	322.45	0.9678	0.0008	50.072
	RF_LS_20_20	24.812	1206.81	1.1106	0.0002	59.239	RF_LO_20	24.795	604.65	0.8936	0.0004	71.143
	PF_LS_25_25	29.088	634.83	2.5516	0.0005	13.196	PF_LO_25	24.844	316.69	0.9219	0.0008	50.192
	RF_LS_25_25	22.662	1178.55	1.1254	0.0002	56.476	RF_LO_25	23.281	589.96	0.8705	0.0004	69.561
	PF_LS_50_50	17.228	587.55	2.2412	0.0003	10.088	PF_LO_50	19.948	294.19	0.7990	0.0007	51.441
	RF_LS_50_50	14.005	988.89	0.9918	0.0001	40.585	RF_LO_50	18.599	494.87	0.7580	0.0004	61.736
	LS_REV	14.987	1147.41	0.5202	0.0001	80.218	LO_EW	10.874	252.23	0.4308	0.0004	58.198
fh = 5	PF_LS_05_05	26.973	184.43	1.1805	0.0015	19.236	PF_LO_05	22.367	91.48	0.6991	0.0024	59.147
	RF_LS_05_05	3.540	262.70	0.1423	0.0001	74.116	RF_LO_05	13.555	132.20	0.4509	0.0010	77.219
	PF_LS_10_10	23.861	175.79	1.4183	0.0014	21.634	PF_LO_10	23.444	87.73	0.8101	0.0027	55.559
	RF_LS_10_10	4.949	255.21	0.1675	0.0002	71.765	RF_LO_10	14.162	128.09	0.5006	0.0011	73.884
	PF_LS_15_15	19.791	168.86	1.3747	0.0012	17.831	PF_LO_15	20.667	84.32	0.7523	0.0025	56.054
	RF_LS_15_15	3.489	249.88	0.0866	0.0001	65.941	RF_LO_15	14.371	125.31	0.5325	0.0011	69.625
	PF_LS_20_20	17.417	163.04	1.3223	0.0011	14.202	PF_LO_20	19.110	81.33	0.7182	0.0023	56.103
	RF_LS_20_20	3.636	244.92	0.0799	0.0001	62.678	RF_LO_20	13.262	122.77	0.5071	0.0011	68.846
	PF_LS_25_25	15.525	158.32	1.3089	0.0010	12.205	PF_LO_25	18.387	79.01	0.7109	0.0023	54.490
	RF_LS_25_25	2.851	239.60	0.0220	0.0001	62.631	RF_LO_25	12.793	119.99	0.4999	0.0011	67.268
	PF_LS_50_50	10.109	137.76	1.1214	0.0007	7.092	PF_LO_50	15.914	68.99	0.6607	0.0023	51.931
	RF_LS_50_50	1.824	201.84	−0.1455	0.0001	47.259	RF_LO_50	11.796	100.99	0.4873	0.0012	59.799
	LS_REV	20.217	231.09	0.6904	0.0009	35.479	LO_EW	10.607	50.56	0.4350	0.0021	55.598
fh = 10	PF_LS_05_05	16.592	107.38	0.7324	0.0015	27.002	PF_LO_05	17.265	53.79	0.5923	0.0032	63.040
	RF_LS_05_05	−3.684	130.58	−0.0855	−0.0003	68.245	RF_LO_05	9.878	65.75	0.3464	0.0015	76.061
	PF_LS_10_10	13.715	101.77	0.8015	0.0013	26.881	PF_LO_10	15.892	50.85	0.5987	0.0031	60.509
	RF_LS_10_10	−0.779	126.69	−0.0607	−0.0001	59.480	RF_LO_10	11.493	63.66	0.4238	0.0018	68.839
	PF_LS_15_15	13.109	97.34	0.8892	0.0013	18.146	PF_LO_15	16.326	48.65	0.6342	0.0034	59.003
	RF_LS_15_15	−0.971	123.62	−0.1272	−0.0001	61.783	RF_LO_15	12.333	62.14	0.4765	0.0020	66.098
	PF_LS_20_20	11.865	93.86	0.8795	0.0013	17.583	PF_LO_20	15.674	46.83	0.6225	0.0033	59.620
	RF_LS_20_20	−0.162	120.99	−0.1333	0.0000	57.787	RF_LO_20	11.750	60.72	0.4650	0.0019	65.149
	PF_LS_25_25	10.814	90.42	0.8485	0.0012	15.777	PF_LO_25	15.558	45.22	0.6315	0.0034	58.561
	RF_LS_25_25	−0.373	118.13	−0.1821	0.0000	57.732	RF_LO_25	10.671	59.26	0.4282	0.0018	64.787
	PF_LS_50_50	8.287	76.79	0.8455	0.0011	5.883	PF_LO_50	14.580	38.46	0.6307	0.0038	52.419
	RF_LS_50_50	0.694	99.54	−0.2581	0.0001	43.628	RF_LO_50	10.771	49.82	0.4540	0.0022	59.861
	LS_REV	5.465	114.63	0.1734	0.0005	38.089	LO_EW	10.223	25.36	0.4321	0.0040	55.256

PF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with price-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
RF_LS_*k*_*k*: Long-short strategy based on forecasts obtained with return-based models. Long in those stocks corresponding to the k% highest forecasted returns, short in those stocks corresponding to the k% lowest forecasted returns.
PF_LO_*k*: Long-only strategy based on forecasts obtained with price-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
RF_LO_*k*: Long-only strategy based on forecasts obtained with return-based models. Invests in those stocks corresponding to the k% highest forecasted returns.
LS_REV: Long-short reversal strategy. Weights are proportional to the opposite of the deviation of the corresponding returns from the mean in the previous period.
LO_EW: Long-only strategy. Portfolio is rebalanced at every trading date so that weights are all equal and sum to 1.

Therefore, as a result of the above observations we believe $ew = 1000$, $fh = 5$, and $k = 10\%$ is an optimal choice; that is, this is a reasonable trade-off between the model forecasting power, the frequency of the model re-calibration, and the number of stocks entering the strategy at each step.

Further we focus on this combination of parameters and analyze the long-short and long-only "*mean-ETL*" strategies with upper bounds $w^* = 5\%, 10\%, 100\%$. In this way, we analyze six additional long-short strategies, and six additional long-only strategies. Table 5 reports the results obtained for the two naïve strategies along with the six long-short "*mean-ETL*" strategies and six long-only "*mean-ETL*" strategies.

The unconstrained optimizations based on factor models of prices show poor performance. This is due to the fact that the expected return forecasts of the top winners and top losers are significantly higher than the rest. In this way the optimization process constantly allocates funds to a few stocks over time. This led to a significant drop during the financial meltdown in 2008, and therefore poor overall performance. Considering 5% versus 10% constraints, the performance of the 5_PF_LS_O1_10_10 is very similar to 10_PF_LS_O1_10_10 in terms of annualized return, annual turnover, and BETC but the maximum drawdown and the Sharpe ratio of the 5_PF_LS_O1_10_10 strategy are much better. Given those results, we investigate further only $w^* = 5\%$ (i.e., the "*mean-ETL*" strategies 5_PF_LS_O1_10_10 and 5_RF_LS_O1_10_10 along with the respective "max-Sharpe" strategies 5_PF_LS_O2_10_10 and 5_RF_LS_O2_10_10).

We also run the same "*mean-ETL*" optimization problems with $\alpha = 0.01, 0.05, 0.1$ and $\lambda = 0.01, 5, 100$. A summary of the results is reported in Table 6. It is clear that the choice of the parameters does not influence results qualitatively. Given that the results are similar, we stick to those two parameters.

Below we discuss the results for the long-short strategies and long-only strategies based on the selected parameters, $ew = 1000, fh = 5, ti = 5, k = 10\%, \alpha = 0.05, \lambda = 5, w^* = 5\%$. Along with the performance of the strategies during the entire backtest period, we also analyze their performance in consecutive sub-periods during the bull market and bear market. The total out-of-sample backtest time window covers December 14, 1992–December 23, 2011 and it is split into five sub-periods (December 14, 1992–September 1, 2000, September 1, 2000–March 12, 2003, March 12, 2003–October 18, 2007, October 18, 2007–March 11, 2009, and March 11, 2009–December 23, 2011. The periods are determined based on the S&P 500 index performance during the backtest period. The respective S&P 500 returns for each sub-period are:

Period	Dec 14th, 1992–Sep 1st, 2000	Sep 1st, 2000–Mar 12th, 2003	Mar 12th, 2003–Oct 18th, 2007	Oct 18th, 2007–Mar 11th, 2009	Mar 11th, 2009–Dec 23rd, 2011
S&P 500 Return	251.35%	−47.12%	91.51%	−53.16%	75.41%

10.2. Long-short strategies

The upper panel of Table 7 reports the results for the six long-short strategies and the benchmark strategy. The first key finding is that *price*-based long-short strategies largely outperform *return*-based long-short strategies. The naïve *price*-based long-short strategy generates a 25.22% annualized return while the "mean-ETL" and "max-Sharpe" *price*-based strategies generate a 27.29% and 29.04% annualized return, respectively. In contrast, the three *return*-based strategies generate 11.37%, 12.77% and 14.11%.

Note that these comparisons are obtained modellng price and returns with dynamic factor models. Of course, we cannot claim that models based on dynamic factors of prices outperform any possible model based on returns. We compared similar classes of models, that is, dynamic factor models. Adding modeling features such as Markov switching behavior or ARCH/GARCH residuals would complicate the models by increasing the parameter count and ultimately decreasing performance.

Focusing on the Sharpe ratio, we obtain a similar picture. The naïve *price*-based long-short strategy produce a Sharpe ratio of 1.5, the "mean-ETL" and "max-Sharpe" *price*-based strategies produce Sharpe ratios equal to 1.24 and 1.51 respectively. The three *return*-based strategies produce Sharpe ratios equal to 0.42, 0.52, and 0.53. The benchmark strategy exhibits an annualized return of 20.46% and a Sharpe ratio of 0.71.

The turnover of the *price*-based strategies is lower than the turnover of the *return*-based strategies. This leads to respectively higher break-even transaction costs. The turnover of the benchmark strategy is comparable to the turnover of the *return*-based naïve strategy. Finally, the maximum drawdown is significantly smaller among the *price*-based strategies. The benchmark strategy is more similar to the return-based strategies in terms of maximum drawdown. The performance of each of the LS strategies over the backtest period is reported in Panel A of Fig. 2.

To check the robustness of the performance of the strategies, we calculated the statistics for each of the five non-overlapping periods defined at the beginning of this section. The results for the long-short strategies, reported in the left panel of Table 8, confirm that *price*-based long-short strategies obtain positive returns for each sub-period. Those strategies generate positive alpha independently of the market trend. The performance of the three *price*-based long-short strategies is similar with all of those strategies exhibiting relatively uniform performance among the consecutive periods of bull and bear markets. The annualized return and Sharpe ratio are better for each of the sub-periods compared to the respective *return*-based long-short strategies. One interesting observation is that the *return*-based strategies exhibited very good performance during the bear market period October 18, 2007–March 11, 2009 which covers the subprime mortgage crisis and the October 2009 bankruptcy of Lehman Brother Holdings. In contrast, most of those return-based strategies exhibit negative returns during the four-year bull market in the period from March 12, 2003 until October 18, 2007. The benchmark performance is also irregular, achieving a very high return during the financial markets meltdown period and a much lower return during the other bear market period from September 1, 2000 until March 12, 2003. It should be emphasized that all backtesting is rigorously out-of-sample and that there is no anticipation of information. The models are estimated at time t using only information known at time t and forecasts are made utilizing only information from models and therefore without any anticipation of information. In particular, the composition of the estimation window at any time t is based on the constituents of the S&P 500 known at time t.

Table 5
Results for the naïve, LS and LO "*mean-ETL*" strategies for $w^* = 5\%, 10\%, 100\%$ ($\alpha = 0.05$, $\lambda = 5$).

LS strategies	Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)	LO strategies	Annualized return (%)	Annual turnover	Annualized sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
PF_LS_10_10	25.221	183.20	1.4994	0.0014	21.006	PF_LO_10	22.637	91.56	0.7883	0.0025	57.797
RF_LS_10_10	11.365	255.26	0.4197	0.0004	41.515	RF_LO_10	19.577	127.46	0.6801	0.0015	64.498
PF_5_LS_O1_10_10	27.288	100.41	1.2407	0.0027	27.094	PF_5_LO_O1_10	22.983	49.82	0.7849	0.0046	59.025
RF_5_LS_O1_10_10	12.774	179.45	0.5211	0.0007	35.894	RF_5_LO_O1_10	19.555	87.36	0.7528	0.0022	65.446
PF_10_LS_O1_10_10	28.793	108.62	0.9441	0.0027	42.863	PF_10_LO_O1_10	21.262	53.55	0.6550	0.0040	58.878
RF_10_LS_O1_10_10	9.389	190.77	0.3666	0.0005	52.254	RF_10_LO_O1_10	19.225	91.55	0.7777	0.0021	53.614
PF_100_LS_O1_10_10	−2.543	132.16	0.5117	−0.0002	99.914	PF_100_LO_O1_10	−26.459	66.01	0.3201	−0.0040	99.997
RF_100_LS_O1_10_10	13.418%	198.33	0.4631	0.0007	47.261	RF_100_LO_O1_10	17.490	95.98	0.6297	0.0018	55.123

The observations so far for the significant difference between price-based and return-based models are further confirmed by the Diebold–Mariano test described in Section 6.1. The value of the statistic S for the differential between *price*-based MSE and *return*-based MSE is equal to 32.24. The test statistic S is distributed as a standard normal variable. Its critical value at 5% significance level is 1.96; we can therefore reject the null of no difference in forecasting accuracy between the two dynamic factor models. The same holds for the *price*-based model and the model based on reversals used for the benchmark strategy. The value of the statistic S for the differential between *price*-based MSE and reversals MSE is equal to 2.23; that is, we can again reject the null hypothesis of no difference in the forecasting accuracy between the *price*-based dynamic factor model and the model based on reversals of the previous period returns. In addition, all strategies pass the stringent statistical-arbitrage test of Hogan et al. (2004).

10.3. Long-only strategies

Let's now consider long-only strategies. Strategies based on factor models of prices continue to outperform strategies based on factor models of returns but the pattern of results is different. In fact, factor models of returns produce good end-of-period returns and the global patterns of cumulative profits over time are similar for both factor models of prices and returns. All the *price*-based long-only strategies exhibit slightly better performance in terms of Sharpe ratio and slightly lower maximum drawdowns can be seen from the bottom panel of Table 7.

We immediately see an important difference compared to long-short strategies. Though final returns are numerically different, all long-only strategies follow similar paths. We observe negative returns during the two bear-market periods and positive returns during the three bull-market periods as shown in the right panel of Table 8. A possible explanation is the existence of one single integrated factor that drives the entire market as discussed in Section 7.2. In contrast, price-based long-short strategies were able to avoid losses in the 2007–2008 period, showing the fundamental importance of the information contained in levels in a period of crisis.

Although the equally weighted benchmark portfolio exhibits poorer performance than the other two strategies, its turnover is much lower than that of the other long-only strategies. In this long-only case, optimization improves results for return-based strategies but fails to improve returns for price-based strategies. The performance of each of the LO strategies over the backtest period is shown in Panel B of Fig. 2.

Once again, we note that all results obtained with the long-only strategies are also strictly out-of-sample and without any anticipation of information.

10.4. Discussion of results

The importance of the results reported in Sections 10.2 and 10.3 is twofold. First, these results have an important bearing on asset pricing theories. Profit obtained with our price-based strategies shows that prices effectively carry more information than returns. Using this information, we can see that there is a considerable level of forecastability in both stock prices and returns. This forecastability is out-of-sample and free from anticipation biases due to the selection of stocks. At the same time, forecastability associated with these strategies is subject to very high turnover and it is therefore sensitive to the volumes eventually traded.

Second, the results we report can be employed in practice in what we might call high-frequency or medium-frequency trading. It seems clear, however, that an optimization of the trade-offs between transaction costs and expected returns has to be

Table 6
Sensitivity of the "mean-ETL" optimization with respect to α and λ for $\alpha = 0.01, 0.05, 0.1$ and $\lambda = 0.01, 5, 100$.

		LS strategies						LO strategies					
		PF_5_LS_O1_10_10			RF_5_LS_O1_10_10			PF_5_LO_O1_10			RF_5_LO_O1_10		
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Annualized return (%)	$\lambda = 0.01$	27.341%	27.293%	27.878%	13.516%	13.725%	13.207%	27.341%	27.293%	27.878%	13.516%	13.725%	13.207%
	$\lambda = 5$	26.627%	27.288%	28.186%	12.978%	12.774%	12.278%	26.627%	27.288%	28.186%	12.978%	12.774%	12.278%
	$\lambda = 100$	27.961%	27.832%	28.482%	12.634%	12.911%	12.546%	27.961%	27.832%	28.482%	12.634%	12.911%	12.546%
Annual turnover	$\lambda = 0.01$	133.90	133.96	133.93	179.55	179.22	179.20	133.90	133.96	133.93	179.55	179.22	179.20
	$\lambda = 5$	101.12	100.41	100.10	179.79	179.45	179.37	101.12	100.41	100.10	179.79	179.45	179.37
	$\lambda = 100$	97.99	96.87	96.36	179.83	179.38	179.37	97.99	96.87	96.36	179.83	179.38	179.37
Annualized Sharpe ratio	$\lambda = 0.01$	1.3697	1.3638	1.3892	0.5405	0.5480	0.5238	1.3697	1.3638	1.3892	0.5405	0.5480	0.5238
	$\lambda = 5$	1.2114	1.2407	1.2749	0.5294	0.5211	0.5005	1.2114	1.2407	1.2749	0.5294	0.5211	0.5005
	$\lambda = 100$	1.2673	1.2594	1.2788	0.5101	0.5267	0.5076	1.2673	1.2594	1.2788	0.5101	0.5267	0.5076
Break-even transaction costs	$\lambda = 0.01$	0.0020	0.0020	0.0021	0.0008	0.0008	0.0007	0.0020	0.0020	0.0021	0.0008	0.0008	0.0007
	$\lambda = 5$	0.0026	0.0027	0.0028	0.0007	0.0007	0.0007	0.0026	0.0027	0.0028	0.0007	0.0007	0.0007
	$\lambda = 100$	0.0029	0.0029	0.0030	0.0007	0.0007	0.0007	0.0029	0.0029	0.0030	0.0007	0.0007	0.0007
Maximum drawdown	$\lambda = 0.01$	21.861%	21.766%	21.538%	31.917%	31.137%	30.614%	21.861%	21.766%	21.538%	31.917%	31.137%	30.614%
	$\lambda = 5$	27.353%	27.094%	27.318%	32.201%	35.894%	36.714%	27.353%	27.094%	27.318%	32.201%	35.894%	36.714%
	$\lambda = 100$	27.848%	28.601%	28.268%	30.575%	34.187%	38.035%	27.848%	28.601%	28.268%	30.575%	34.187%	38.035%

Table 7
Results for the naïve and optimized strategies for the total backtest period.

Strategies/statistics		Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
LS strategies	PF_LS_10_10	25.221	183.20	1.4994	0.0014	21.006
	RF_LS_10_10	11.365	255.26	0.4197	0.0004	41.515
	PF_5_LS_O1_10_10	27.288	100.41	1.2407	0.0027	27.094
	RF_5_LS_O1_10_10	12.774	179.45	0.5211	0.0007	35.894
	PF_5_LS_O2_10_10	29.037	146.37	1.5090	0.0020	18.776
	RF_5_LS_O2_10_10	14.114	179.03	0.5286	0.0008	29.967
	LS_REV	20.460	231.17	0.7142	0.0009	37.516
LO strategies	PF_LO_10	22.637	91.56	0.7883	0.0025	57.797
	RF_LO_10	19.577	127.46	0.6801	0.0015	64.498
	PF_5_LO_O1_10	22.983	49.82	0.7849	0.0046	59.025
	RF_5_LO_O1_10	19.555	87.36	0.7528	0.0022	65.446
	PF_5_LO_O2_10	20.084	65.78	0.7394	0.0031	49.350
	RF_5_LO_O2_10	18.979	89.00	0.6691	0.0021	68.184
	LO_EW	11.078	50.57	0.4631	0.0022	55.262

implemented. Return forecasts obtained with *price*-based factor models have a very robust ordering in the sense that higher expected returns correspond to higher realized returns with a high probability as shown by the profits attained by the strategy. However, any practical implementation should reduce the turnover of the strategy through appropriate optimizations. These practical issues are beyond the scope of this paper whose objective is to show how profitable statistical arbitrage strategies can be built using dynamic factor models of prices.

10.5. Intuition of the main result

The key result of this paper is the following: models based on prices allow for better forecasting of *relative* returns than models based solely on returns. We emphasize that our results are relative to the construction of statistical arbitrage strategies based on the differences of returns between pairs of stocks. Forecasting the absolute returns of stocks or even the direction of markets might be more difficult to achieve and require other considerations. However, in this paper we have concentrated primarily on models widely used in the asset management community: factor models.

The reason why it is possible to forecast the relative performance of stock returns is that there is cointegration in the market. Actually, we obtain a very strong result: after performing PCA of logprices, we find that we cannot reject the null of only one integrated factor. All other principal components are stationary. We obtain exactly the same result (i.e., one common trend) testing logprice processes directly and not through principal components.

Because there is only one common stochastic trend, all logprice processes are mean-reverting around that common trend. It is this fact that is responsible for the forecasting results that we obtain. Why does cointegration allow better relative forecasting? Intuitively, it is because it includes information on the level of a series and not only on past returns. By taking the first differences, this information is lost.

We can now propose a tentative interpretation of our results. First, the ups and downs of the market depend on the total demand and supply of investments; that is, they depend on the total amount of investment money available or, alternatively, on the total need for money. Financial demand and supply are determined by the inflow/outflow of money from/to abroad, changes in domestic money supply, and profits generated by the real economy. It is questionable whether profits generated by the real economy become financial demand through changes in the money supply. We will not addressing these issues here. And we will not address here the question of if and how changes in aggregate are predictable or not.

Our strategies work by exploiting mean reversion in relative price changes, that is, cointegration of prices, and therefore we now focus on an economic explanation of cointegration. Here we offer a very simple explanation of financial cointegration. Investors and investment managers apportion the funds in search of investments to different financial assets in function of the relative

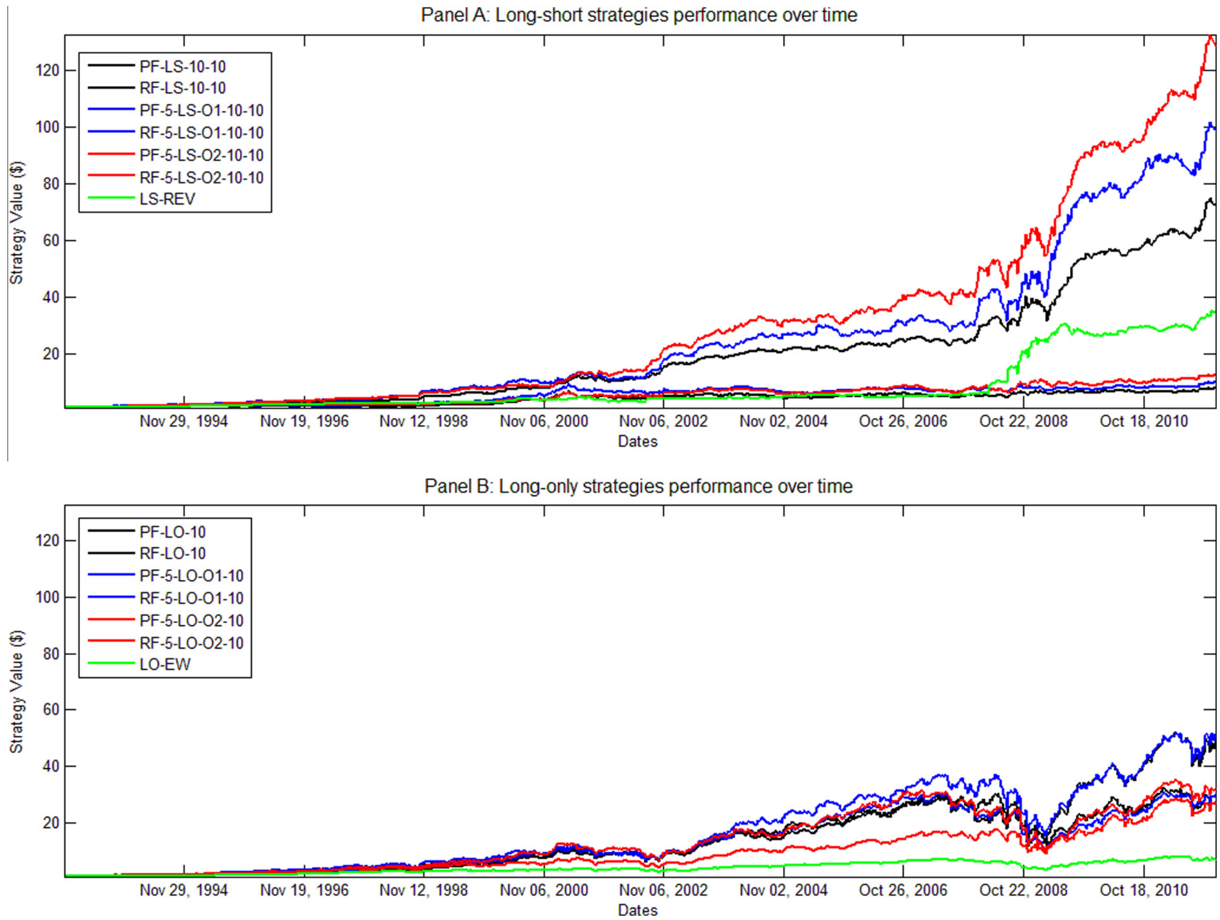


Figure 2. Performance of the strategies over time (December 14 1992–December 23, 2011). Panel A: Long-short strategies. Panel B: Long-only strategies.

attractiveness of investments. Ultimately investments are rated by their fundamentals, in particular their ability to provide a stream of profits.

Attractive investments attract money and therefore their prices increase. But due to price increases, investments become progressively overpriced and therefore less attractive. The opposite is also true. Overpriced investments are sold and therefore their price decreases. Progressively, they become more attractive. Therefore, there is a natural mean reversion in investment fundamentals. This is, we argue, the main mechanism of cointegration.

The above explanation assumes that profits generated by investments be stationary. This is not always true. Some companies, especially those that work in near monopoly situations, might exhibit prolonged periods of growth of profits and earnings. Or other companies might exhibit prolonged periods of decreasing earnings. However, we can assume that, on average, competitive pressure keeps profits and earnings stationary.

There are other phenomena, such as momentum, that seem to invalidate the mean-reversion property of fundamentals. However, the momentum effect is strongly observable only in relatively small fractions of the market. On average, financial markets exhibit relative mean reversion of fundamentals. This is the key phenomenon which prevents markets from behaving like random walks.

10.6. A simulation test

To gain additional intuition for our model, we introduce some heuristic considerations: If price processes were driven by a multitude of random walks, the cross-sectional variance of logprice processes would grow much faster than the growth observed empirically. Though not a formal test, this is a powerful argument that lends support to the key finding of only one integrated factor. Our argument is a multivariate variant of the variance ratio test.

Formal variance ratio tests are based on comparing the variance of an empirical univariate process with the theoretical variance of a random walk. The variance of a random walk $p_t = p_{t-1} + \varepsilon_t$, $\mathrm{var}(\varepsilon_t) = \sigma^2, t = 1, 2, \ldots$ grows linearly over time with rate σ^2. If the variance of an empirical time series does not grow linearly, we have arguments to conclude that the empirical series is not a path of a random walk. Because the variance of price differences over k periods is the product of k and the variance of price differences over one period, it is possible to construct a variance ratio test that is invariant to σ^2. Lo and MacKinlay (1988) gives the formal analysis of the variance ratio test as well as determining the sampling distributions of variance ratios.

Now consider a multivariate time series of prices. The multivariate equivalent of variance is the covariance matrix which cannot, of course, be fully represented by a single number. We gain a better intuition of the random walk behavior of a series by considering the "spread" of its paths as measured by the cross-sectional variance of the paths. As explained in the appendix, the cross-sectional variance of the path of a random walk grows linearly with time. If the growth of the empirical cross-sectional variance of a multivariate series is nonlinear over time, we have a basis for rejecting the assumption that the multivariate empirical time series is a path of a multivariate random walk. More importantly, we can test the null of a single integrated factor by testing the null of constant cross-sectional variance.

Table 8
Results for the naïve and optimized strategies for the total backtest period and selected sub-periods.

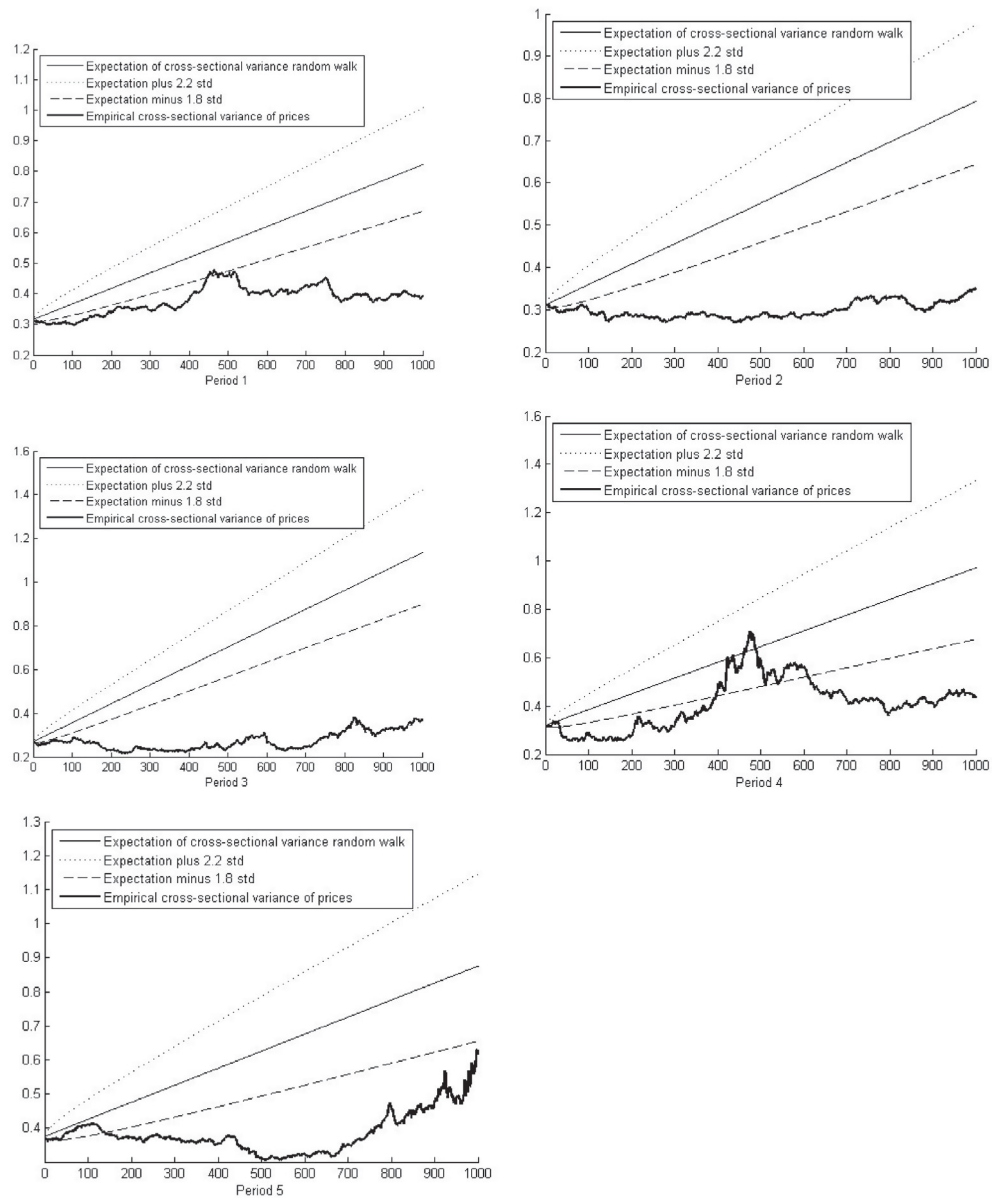
Periods	LS strategies	Annualized return (%)	Annual turnover	Annualized Sharpe ratio	Break-even transaction costs	Maximum drawdown (%)	LO strategies	Annualized return (%)	Annual turnover	Annualized sharpe ratio	Break-even transaction costs	Maximum drawdown (%)
Dec 14th, 1992–Sep 1st, 2000	PF_LS_10_10	30.651	188.3995	2.0151	0.0016	9.561	PF_LO_10	31.272	93.6201	1.3978	0.0033	24.531
Sep 1st, 2000–Mar 12th, 2003		32.778	183.8595	1.5916	0.0018	16.005		0.443	91.9555	0.0948	0.0000	45.154
Mar 12th, 2003–Oct 18th, 2007		9.808	172.3439	0.6786	0.0006	12.421		30.057	86.5536	1.4066	0.0035	13.156
Oct 18th, 2007–Mar 11th, 2009		27.350	204.0900	1.0365	0.0013	15.332		−43.145	103.2953	−0.9887	−0.0042	57.797
Mar 11th, 2009–Dec 23rd, 2011		29.446	175.8767	1.9450	0.0017	10.045		60.263	88.2284	1.7969	0.0068	23.160
Total		**25.221**	**183.1991**	**1.4994**	**0.0014**	**21.006**		**22.637**	**91.5590**	**0.7883**	**0.0025**	**57.797**
Dec 14th, 1992–Sep 1st, 2000	RF_LS_10_10	18.425	256.7578	0.6711	0.0007	21.816	RF_LO_10	30.992	128.7719	1.3669	0.0024	18.822
Sep 1st, 2000–Mar 12th, 2003		11.482	252.8496	0.3890	0.0005	41.515		2.899	126.3389	0.1603	0.0002	37.725
Mar 12th, 2003–Oct 18th, 2007		−1.456	252.4834	−0.2116	−0.0001	28.343		25.574	124.5039	1.1365	0.0021	17.436
Oct 18th, 2007–Mar 11th, 2009		26.488	251.2940	0.7715	0.0011	20.525		−46.465	122.6943	−1.0010	−0.0038	61.001
Mar 11th, 2009–Dec 23rd, 2011		6.819	259.7722	0.4146	0.0003	16.272		46.727	132.1998	1.4876	0.0035	25.511
Total		**11.365**	**255.2569**	**0.4197**	**0.0004**	**41.515**		**19.577**	**127.4611**	**0.6801**	**0.0015**	**64.498**
Dec 14th, 1992–Sep 1st, 2000	PF_LS_O1_10_10	35.169	106.1886	1.6765	0.0033	11.078	PF_LO_O1_10	34.267	52.1903	1.3839	0.0066	23.459
Sep 1st, 2000–Mar 12th, 2003		27.793	99.3680	1.1198	0.0028	20.535		−3.144	49.7819	0.0091	−0.0006	46.928
Mar 12th, 2003–Oct 18th, 2007		10.514	90.1712	0.5740	0.0012	17.813		33.177	45.5931	1.4882	0.0073	15.061
Oct 18th, 2007–Mar 11th, 2009		21.985	121.4049	0.6528	0.0018	27.094		−44.168	61.4074	−0.9653	−0.0072	58.900
Mar 11th, 2009–Dec 23rd, 2011		38.063	91.9684	2.2530	0.0041	8.409		53.921	44.7028	1.9048	0.0121	20.345
Total		**27.288**	**100.4058**	**1.2407**	**0.0027**	**27.094**		**22.983**	**49.8224**	**0.7849**	**0.0046**	**59.025**
Dec 14th, 1992–Sep 1st, 2000	RF_LS_O1_10_10	23.342	182.3225	0.9835	0.0013	21.529	RF_LO_O1_10	33.345	88.7039	1.5867	0.0038	11.386
Sep 1st, 2000–Mar 12th, 2003		9.063	171.1829	0.3295	0.0005	34.820		−0.232	85.7775	0.0203	0.0000	39.866
Mar 12th, 2003–Oct 18th, 2007		0.063	177.6611	−0.1764	0.0000	32.334		23.852	85.5249	1.1646	0.0028	18.232
Oct 18th, 2007–Mar 11th, 2009		4.568	178.1887	0.1949	0.0003	18.841		−47.584	85.1340	−1.3421	−0.0056	60.309
Mar 11th, 2009–Dec 23rd, 2011		11.827	182.2313	0.7524	0.0006	12.354		45.161	89.3043	1.7495	0.0051	18.739
Total		**12.774**	**179.4512**	**0.5211**	**0.0007**	**35.894**		**19.555**	**87.3609**	**0.7528**	**0.0022**	**65.446**
Dec 14th, 1992–Sep 1st, 2000	PF_LS_O2_10_10	32.107	161.3119	1.6988	0.0020	8.760	PF_LO_O2_10	25.691	67.8737	1.1182	0.0038	24.792
Sep 1st, 2000–Mar 12th, 2003		46.172	120.0852	2.0219	0.0038	12.132		−4.213	63.7843	−0.0717	−0.0007	46.993
Mar 12th, 2003–Oct 18th, 2007		14.184	148.3251	0.9202	0.0010	12.100		27.503	63.4721	1.3787	0.0043	13.606
Oct 18th, 2007–Mar 11th, 2009		19.852	149.6657	0.6668	0.0013	18.776		−34.939	74.0882	−0.8288	−0.0047	49.350
Mar 11th, 2009–Dec 23rd, 2011		35.619	123.7062	2.6320	0.0029	4.892		57.431	61.6886	1.9266	0.0093	21.688
Total		**29.037**	**146.3726**	**1.5090**	**0.0020**	**18.776**		**20.084**	**65.7803**	**0.7394**	**0.0031**	**49.350**
Dec 14th, 1992–Sep 1st, 2000	RF_LS_O2_10_10	18.858	181.0287	0.7026	0.0010	17.916	RF_LO_O2_10	33.810	90.3482	1.4905	0.0037	16.398
Sep 1st, 2000–Mar 12th, 2003		17.716	173.5566	0.5389	0.0010	29.099		−1.111	86.0649	0.0347	−0.0001	47.161
Mar 12th, 2003–Oct 18th, 2007		1.820	178.1238	−0.0182	0.0001	29.967		26.372	88.0453	1.1775	0.0030	17.782
Oct 18th, 2007–Mar 11th, 2009		35.592	177.3218	0.9735	0.0020	27.779		−51.619	85.5064	−1.1871	−0.0060	64.919
Mar 11th, 2009–Dec 23rd, 2011		8.429	180.7927	0.5006	0.0005	18.331		43.039	91.2847	1.4873	0.0047	21.983
Total		**14.114**	**179.0298**	**0.5286**	**0.0008**	**29.967**		**18.979**	**88.9986**	**0.6691**	**0.0021**	**68.184**
Dec 14th, 1992–Sep 1st, 2000	LS_REV	19.059	231.3256	1.1820	0.0008	12.859	LO_EW	16.324	50.5012	0.8275	0.0032	19.310
Sep 1st, 2000–Mar 12th, 2003		0.985	228.9880	0.0508	0.0000	37.516		−5.012	50.5362	−0.2391	−0.0010	31.820
Mar 12th, 2003–Oct 18th, 2007		7.565	232.0904	0.3781	0.0003	17.647		20.619	50.7356	1.2472	0.0041	10.660
Oct 18th, 2007–Mar 11th, 2009		190.273	232.1599	1.7371	0.0082	19.077		−41.969	49.9593	−1.5693	−0.0084	54.178
Mar 11th, 2009–Dec 23rd, 2011		13.465	230.4981	1.0730	0.0006	15.015		34.219	50.8086	1.3951	0.0067	20.273
Total		**20.460**	**231.1681**	**0.7142**	**0.0009**	**37.516**		**11.078**	**50.5716**	**0.4631**	**0.0022**	**55.262**

Note: The bolded values denote the entire period statistics.

Let's compare the empirical cross-sectional variance of the stocks included in the S&P 500 universe with the theoretical cross-sectional variance of a random walk with the same covariance matrix of the S&P 500. In order to perform this comparison, we construct a band $(E(S_t) - l\,\mathrm{var}(S_t), E(S_t) + h\,\mathrm{var}(S_t))$ and we verify if the empirical cross-sectional variance remains outside this band. Given the nature of the distribution of $E(S_t)$, it is impossible to compute the band theoretically. However, our extensive simulations show that for the estimated covariance matrix of the S&P 500, 98% of the theoretical variance falls in the band with $l = 1.8$ and $h = 2.2$.

Empirical results show that the empirical covariance definitely falls outside the theoretical band. For each time window, we consider the logprices and form the covariance of their first differences: $\Omega = \mathrm{cov}(\Delta \log \mathit{prices}) = \mathrm{cov}(\log \mathit{ret})$. For each time window, the plots in Fig. 3 show graphically the expectation of variance $E(S_t)$ (thin line) and the interval $(E(S_t) - l\,\mathrm{var}(S_t), E(S_t) + h\,\mathrm{var}(S_t))$ (dotted and dashed lines), and the empirical cross-sectional variance (thick line).

This is not a formal test because, though we consider the sampling distribution of random walks given a covariance matrix, we use the estimated covariance matrix as if it were the true



Period 1	Period 2	Period 3	Period 4	Period 5
03-Jan-1989-	15-Dec-1992-	27-Nov-1996-	14-Nov-2000-	10-Nov-2004-
14-Dec-1992	26-Nov-1996	13-Nov-2000	09-Nov-2004	29-Oct-2008

Figure 3. Theoretical and empirical cross-sectional variance of logprices of the S&P 500 universe in five time windows. The figure in each of the five panels shows that, at the end of each period, the empirical cross-sectional variance of prices is consistently outside the 98% confidence level obtained by simulating random walks with the empirical covariance matrix of returns in the corresponding period. The empirical cross-sectional variance of prices seems to have an almost flat behavior with some changes in level. Though not a formal test, the evolution of the cross-sectional variance of prices offers a good heuristic to support the conclusion of one integrated factor per period.

Table 9
Long/Short Strategies regression to the Carhart four-factor model. The table reports regression coefficients, the respective *p*-values and the R-squared for each strategy for the entire period and five sub-periods.

	LS strategies	Factor/strategy	PF_LS_10_10	RF_LS_10_10	PF_5_LS_O1_10_10	RF_5_LS_O1_10_10	PF_5_LS_O2_10_10	RF_5_LS_O2_10_10	LS_REV
Total Period	Regression Coefficients	Intercept	0.020	0.009	0.023	0.009	0.023	0.011	0.020
		Rm–Rf	0.080	−0.085	0.194	−0.022	0.078	−0.103	−0.214
		SMB	0.075	0.227	−0.005	0.208	0.106	0.281	0.014
		HML	0.008	0.299	−0.090	0.332	−0.051	0.361	−0.124
		Mom	−0.218	0.148	−0.368	0.207	−0.301	0.208	−0.239
	p-values	Intercept	0.000%	3.024%	0.000%	1.301%	0.000%	1.222%	0.001%
		Rm–Rf	18.437%	39.584%	1.701%	80.332%	24.851%	30.590%	4.058%
		SMB	37.217%	10.147%	96.641%	8.579%	25.658%	4.436%	92.443%
		HML	93.128%	3.992%	44.352%	0.910%	59.896%	1.355%	40.870%
		Mom	0.154%	18.890%	0.007%	3.624%	0.010%	6.678%	4.133%
	R-squared		10.734%	1.156%	14.762%	2.012%	12.974%	2.557%	1.356%
Jan/1993–Aug/2000	Regression Coefficients	Intercept	0.022	0.021	0.025	0.023	0.024	0.019	0.019
		Rm–Rf	0.312	−0.275	0.483	−0.182	0.327	−0.210	0.047
		SMB	0.181	0.033	0.183	0.102	0.198	0.163	0.181
		HML	0.166	−0.256	0.132	−0.176	0.103	−0.054	0.046
		Mom	−0.170	−0.194	−0.299	−0.158	−0.257	−0.044	−0.317
	p-values	Intercept	0.000%	0.176%	0.001%	0.008%	0.000%	0.269%	0.000%
		Rm–Rf	0.210%	11.018%	0.134%	21.663%	0.615%	19.997%	60.297%
		SMB	12.542%	87.272%	29.508%	56.062%	15.866%	40.258%	9.430%
		HML	22.319%	27.781%	51.091%	38.151%	52.282%	80.888%	71.257%
		Mom	13.916%	33.089%	8.086%	35.479%	6.050%	81.641%	0.317%
	R-squared		13.129%	−0.632%	13.850%	−0.970%	12.080%	−1.045%	15.196%
Sep/2000–Feb/2003	Regression Coefficients	Intercept	0.031	−0.014	0.040	−0.011	0.042	−0.011	0.013
		Rm–Rf	0.099	−0.056	−0.006	0.291	−0.121	0.044	−0.472
		SMB	0.018	−0.152	−0.274	−0.180	0.064	−0.043	−0.342
		HML	0.010	1.511	−0.475	1.490	−0.361	1.586	−0.588
		Mom	−0.333	0.423	−0.609	0.617	−0.554	0.423	−0.498
	p-values	Intercept	0.649%	45.499%	0.597%	45.187%	0.152%	54.004%	44.590%
		Rm–Rf	65.758%	88.217%	98.297%	33.585%	63.052%	90.530%	19.446%
		SMB	94.537%	72.827%	40.208%	60.189%	82.614%	91.925%	40.963%
		HML	97.233%	0.421%	19.746%	0.060%	26.982%	0.218%	20.760%
		Mom	8.167%	18.688%	1.477%	1.927%	1.286%	17.328%	10.326%
	R-squared		21.640%	22.088%	21.810%	30.743%	23.583%	23.407%	−1.926%
Mar/2003–Sep/2007	Regression Coefficients	Intercept	0.007	0.000	0.007	−0.002	0.012	−0.001	0.005
		Rm–Rf	0.129	−0.128	0.080	0.049	−0.048	0.046	0.139
		SMB	0.347	0.485	0.514	0.609	0.565	0.556	0.207
		HML	−0.437	−0.021	−0.392	−0.165	−0.617	−0.041	−0.307
		Mom	−0.553	−0.136	−0.634	−0.257	−0.753	−0.232	-0.019
	p-values	Intercept	7.756%	99.720%	19.161%	80.289%	1.230%	93.775%	36.581%
		Rm–Rf	46.407%	69.016%	74.889%	86.414%	82.008%	87.241%	59.343%
		SMB	8.252%	18.170%	7.053%	6.135%	1.980%	8.935%	47.756%
		HML	9.316%	96.367%	28.424%	69.197%	4.888%	92.171%	41.937%
		Mom	0.092%	63.902%	0.644%	31.926%	0.021%	37.194%	93.650%
	R-squared		24.290%	−3.284%	17.319%	5.603%	26.089%	3.982%	−2.261%
Oct/2007–Feb/2009	Regression Coefficients	Intercept	0.032	0.038	0.044	0.014	0.045	0.032	0.113
		Rm–Rf	−0.330	0.542	−0.213	0.088	−0.142	−0.034	−0.470
		SMB	−0.903	1.149	−1.087	0.842	−1.101	1.495	0.912
		HML	0.796	−0.490	0.959	0.607	0.888	0.236	2.322
		Mom	−0.165	−0.291	−0.436	0.645	−0.464	0.264	0.827
	p-values	Intercept	1.720%	17.830%	4.019%	60.734%	0.195%	37.934%	3.223%
		Rm–Rf	10.203%	22.468%	50.545%	83.839%	45.511%	95.281%	54.393%
		SMB	9.812%	33.566%	21.919%	47.255%	4.643%	33.938%	66.176%
		HML	4.933%	56.626%	14.028%	47.453%	2.930%	83.258%	13.995%

Table 9 (*continued*)

LS strategies		Factor/strategy	PF_LS_10_10	RF_LS_10_10	PF_5_LS_O1_10_10	RF_5_LS_O1_10_10	PF_5_LS_O2_10_10	RF_5_LS_O2_10_10	LS_REV
		Mom	62.790%	70.558%	44.403%	40.320%	18.157%	79.377%	54.716%
	R-squared		51.013%	4.377%	38.619%	−18.414%	67.498%	−19.961%	0.424%
Mar/2009–Dec/2011	Regression Coefficients	Intercept	0.016	0.011	0.012	0.012	0.018	0.015	0.009
		Rm–Rf	0.133	−0.529	0.481	−0.484	0.177	−0.538	−0.015
		SMB	−0.091	0.134	−0.319	−0.010	−0.108	−0.191	−0.108
		HML	0.347	1.264	0.069	1.513	0.200	1.684	−0.115
		Mom	0.091	0.487	−0.230	0.585	0.023	0.794	−0.176
	p-values	Intercept	3.120%	18.693%	15.536%	14.563%	1.067%	10.971%	18.586%
		Rm–Rf	45.991%	1.577%	2.764%	1.806%	29.307%	2.505%	92.645%
		SMB	75.201%	68.866%	34.727%	97.353%	68.712%	60.695%	67.813%
		HML	38.646%	0.983%	88.278%	0.140%	58.923%	0.232%	74.788%
		Mom	67.541%	6.009%	36.654%	1.783%	90.727%	0.709%	37.043%
	R-squared		2.604%	14.995%	29.625%	23.353%	2.856%	18.685%	−6.794%

covariance matrix and do not consider its sampling distribution. However, it is quite evident that there are large differences in the behavior of the real and the simulated series. The empirical variance falls outside the theoretical confidence interval by a significant amount. This fact lends support to the formal conclusions in Section 7.2.

10.7. Strategies exposures to well-known market factors

We tested the exposure of the proposed long-short strategies to well-known industry factors. To this end, we ran a regression for each of the six long-short strategies along with the benchmark strategy versus the four factors proposed by Carhart (1997). We used monthly factor returns available on Professor Kenneth French's website. The four factors are the market factor ($Rm - Rf$), the size factor (*SMB*), the value factor (*HML*), and the momentum factor (*Mom*). We perform those regressions for the entire backtest period as well as for the same five sub-periods described in the beginning of this section.

Table 9 reports factor exposures, the respective p-values as well as the R-squared for each of the strategies and each sub-period. The alpha (intercept) of our long-short *price*-based strategies is always positive and significant in most cases. The explanatory power (R-squared) of the four-factor model is very low. All the price-based long-short strategies have significant negative exposure to the momentum factor; the momentum factor, however, is not sufficient to explain the behavior of those strategies. The *return*-based strategies have significant exposures to the value factor (HML). Their alpha is not significant during most of the market regimes. The benchmark strategy also exhibits significant alpha and significant negative exposure to the momentum factor.

11. Conclusions

The results we present in this paper provide empirical evidence that profitable statistical arbitrage can be constructed using dynamic factor models of prices. The theoretical result is that dynamic factor models of prices provide a general theoretical framework for statistical arbitrage. We demonstrate empirically that models based on prices lead to better return forecasts than models based only on returns. The level of forecastability obtained is surprisingly high with returns of the order of 27% in long-short strategies even in a highly liquid market for the stock traded, such as component stocks of the S&P 500. This level of profit associated with a medium-frequency strategy (weekly trading) has an important bearing on asset pricing theory.

The theoretical underpinning of these results is the fact that stock markets are mean-reverting around a single integrated factor. In practice, dynamic factor models of prices largely outperform dynamic factor models of returns. The results obtained have practical application after a judicious application of optimization procedures to improve the trade-off between returns and turnover.

Appendix A

In this appendix we compute the expectation and the standard deviation of the cross-sectional variance of a multivariate random walk as functions of time.

Suppose the logarithm of prices follow a multivariate random walk:

$$p_t = p_{t-1} + \varepsilon_t, \mathrm{cov}(\varepsilon_t) = \Omega, \quad i = 1, 2, \ldots, n, \ t = 1, 2, \ldots \qquad (17)$$

The cross-sectional variance of logprices $S_t = \frac{1}{n}\sum_{i=1}^{n}(p_{it} - \bar{p}_t)^2$ depends on the covariance matrix Ω_t of prices at time t. Intuitively, the growth rate of the cross-sectional variance decreases with increasing average positive correlations of logprices and goes to zero in the case of perfect correlation between the time series.

Let's recall a few facts related to variances and to the probability distribution of quadratic forms of Gaussian random vectors $x = (x_1, \dots, x_n)'$. It is well known that the empirical variance V of a set of n observations x is a positive definite quadratic form:

$$V = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = x'C_n x \tag{18}$$

where C_n, called a centering matrix, is defined as follows:

$$C_n = \left(I - \frac{1}{n}J\right)$$

where J is a matrix with all elements equal to 1: $J = 1_n 1_n', 1_n = \underbrace{[1, \dots, 1]_n'}$.

It is well known that the matrix C_n is symmetrical and idempotent, has rank $n-1$, and a trace equal to 1. C_n has one zero eigenvalue and $n-1$ eigenvalues equal to $1/(n-1)$.

The quantity $V = x'C_n x$ is a positive definite quadratic form. The distribution of a quadratic form – even when x is a multivariate normal variable with covariance matrix Σ – cannot be expressed in simple forms in terms of elementary functions. Exact, but not simple, and relatively simple approximate representations of the distribution of V are obtained, among others, by Gurland (1953), Box (1954), Imhof (1961), Khuri and Good (1977) and Tziritas (1987).

For our purposes, however, the following property is sufficient. Following Tziritas (1987), if the variables x are multivariate normal variables with mean $m = (m_1, \dots, m_n)'$ and covariance matrix Σ, then the quadratic form (18) can be diagonalized in the sense that V it is distributed as the sum $\sum_{i=1}^{r}\xi_i Z_i^2$ where $Z_i = V_i \Sigma^{-\frac{1}{2}} x$ and V_i, ξ_i are, respectively, the i^{th} eigenvector and eigenvalue of the matrix $\Sigma^{\frac{1}{2}} C_n \Sigma^{\frac{1}{2}}$. The variables Z_i are independent normal variables with variance 1 and mean $\mu_i = V_i \Sigma^{-\frac{1}{2}} m$ and therefore the Z_i^2 are distributed as independent non-central chi-square variables $\chi^2(1, m_i^2)$ with non-centrality parameter m_i^2.

If the variables x are multivariate normal variables with zero mean and covariance matrix Σ, given that the matrix C_n has rank $r = n-1$, then V is distributed as the sum $\sum_{i=1}^{r}\xi_i \chi_i^2(1)$ where the $\chi_i^2(1)$ are independent central chi-square variables with one degree of freedom and the ξ_i are the eigenvalues of the matrix $\Sigma^{\frac{1}{2}} C_n \Sigma^{\frac{1}{2}}$ (see Box, 1954).

From these two properties, it follows (see Tziritas, 1987) that the expectation and the variance of the quadratic form (18) are:

$$\begin{aligned} E(V) &= trace(\Sigma^{\frac{1}{2}} C_n \Sigma^{\frac{1}{2}}) + m'C_n m \\ \mathrm{var}(V) &= \begin{cases} 2\sum \xi_i^2 & \text{if} \quad m = 0 \\ 2\sum \xi_i^2 (1 + 2m_i^2) & \text{if} \quad m \neq 0 \end{cases} \end{aligned} \tag{19}$$

Let's apply these properties to the random walk (17). By repeating application of that equation, we obtain $p_t = \sum_{s=1}^{t}\varepsilon_s + p_0$. If we assume that the price distribution vector p_0 is non-random, then the random vector p_t is the sum of the initial constant p_0 plus t i.i.d. multivariate normal variables. Hence, given that $m = p_0$, the following properties hold:

$$\begin{aligned} \Sigma_t &= t\Omega \\ \lambda_{it} &= t\upsilon_i \end{aligned} \tag{20}$$

where υ_i are the eigenvalues of the matrix $C_n\Omega$. Using (14), the expectation and the variance of S_t are:

$$\begin{aligned} E(S_t) &= t \times trace(C_n\Omega) + p_0'C_n p \\ \mathrm{var}(S_t) &= t^2 \times \sum \xi_i^2 2\left(1 + \tfrac{2m_i^2}{t}\right) \end{aligned} \tag{21}$$

The relationships given by (21) show that the expectation of the cross-sectional variance of a multivariate random walk grows linearly with a rate that depends only on the covariance matrix of the innovations while the growth of the standard deviation of S_t is nonlinear and depends on both the covariance matrix of innovations and the initial prices. Note that if $\Omega = \sigma^2 I$, then $C_n\Omega = \sigma^2 C_n$ and, given that $trace(C_n) = 1$, we find the familiar result that the expectation of the variance of a standard normal variable is simply equal to $\sigma^2 t$ and the distribution of the variance is

$$\sum_{i=1}^{n-1}\frac{1}{n-1}\chi_i^2(1) = \frac{1}{n-1}\chi^2(n-1).$$

References

Alexander, C., 1999. Optimal hedging using cointegration. Philosophical Transactions of the Royal Society London A 357, 2039–2058.

Alexander, C., Dimitriu, A., 2002. The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies. Working paper, ISMA Centre of the University of Reading.

Alexander, C., Dimitriu, A., 2005a. Indexing and statistical arbitrage. Journal of Portfolio Management 31, 50–63.

Alexander, C., Dimitriu, A., 2005b. Indexing, cointegration and equity market regimes. International Journal of Financial Economics 10, 1–19.

Avellaneda, M., Lee, J.-H., 2010. Statistical arbitrage in the US equities market. Quantitative Finance 10, 761–782.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2008. Large dimensional factor analysis. Foundations and Trends in Econometrics 3, 89–163.

Banbura, M., Giannone, D., Reichlin, L., 2010. Nowcasting. Working paper, ECARES.

Barhoumi, K., Darné, O., Ferrara, L., 2013. Dynamic factor models: A review of the literature. Working paper, Banque de France.

Boguslavsky, M., Boguslavskaya, E., 2003. Optimal arbitrage trading. Working paper, University of Amsterdam.

Bondarenko, O., 2003. Statistical arbitrage and security prices. Review of Financial Studies 16, 875–919.

Bossaerts, P., 1988. Common nonstationary components of asset prices. Journal of Economic Dynamics and Control 12, 347–364.

Box, G.E.P., 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Annals Mathematical Statistics 25, 290–302.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Deistler, M., Hamann, E., 2005. Identification of factor models for forecasting returns. Journal of Financial Econometrics 3, 256–281.

Diebold, F.X., Mariano. R.S., 1994. Comparing pedictive accuracy. Working paper, National Bureau of Economic Research.

Do, B., Faff, R., Hamza, K., 2006. A new approach to modeling and estimation for pairs trading. In: Proceedings of 2006 Financial Management Association European Conference, Stockholm, June.

Eklund, J., Kapetanios, G., 2008. A review of forecasting techniques for large data sets. National Institute Economic Review 203, 109–115.

Elliot, R., Van Der Hoek, J., Malcolm, W., 2005. Pairs trading. Quantitative Finance 5, 271–276.

Engle, R., Vahid, F., 1993. Common trends and common cycles. Journal of Applied Econometrics 8, 341–360.

Escribano, A., Peña, D., 1994. Cointegration and common factors. Journal of Time Series Analysis 15, 577–586.

Fernholz, R., Maguire, C., 2007. The statistics of statistical arbitrage. Financial Analysts Journal 63, 46–52.

Forni, M., Hallin, M., Reichlin, M.L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: Structural factor models with large cross-sections. Econometric Theory 25, 1319–1347.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies 19, 797–827.

Geweke, J., 1997. The dynamic factor analysis of economic time series. In: Aigner, D. J., Golderdberger, A.S. (Eds.), Latent Variables in Socio-Economic Models. North Holland Press, Amsterdam.

Gonzalo, J., Granger, C., 1995. Estimation of common long-memory components in cointegrating systems. Journal of Business & Economic Statistics 13, 27–35.

Gurland, J., 1953. Distribution of quadratic forms and ratios of quadratic forms. Annals of Mathematical Statistics 24, 416–427.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Harris, D., 1997. Principal components analysis of cointegrated time series. Econometric Theory 13, 529–557.

Hogan, S., Jarrow, R., Teo, M., Warachka, M., 2004. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial Economics 73, 525–565.

Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48, 419–426.
Jarrow, R., M. Teo, T. Y. Tse, M. Warachka, 2005. Statistical arbitrage and market efficiency: enhanced theory, robust tests and further applications. Working paper, Cornell University.
Jarrow, R., Teo, M., Tse, T.Y., Warachka, M., 2012. An improved test for statistical arbitrage. Journal of Financial Markets 15, 47–80.
Kapetanios, G., 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business & Economic Statistics 28, 397–409.
Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54, 159–178.
Khuri, A.I., Good, I.J., 1977. The distribution of quadratic forms in non-normal variables and an application to the variance ratio. Journal of the Royal Statistical Society B39, 217–221.
Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447–1479.
Lehmann, B.N., 1990. Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1–28.
Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.
Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–206.
Lopes, H.F., Carvalho, C.M., 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference 137, 3082–3091.
Marcet, A., 2005. Overdifferencing VAR's is ok. Working Paper, Institut d'Anàlisi Econòmica (CSIC), UPF and CEPR.
Montana, G., Triantafyllopoulos, K., Tsagaris, T., 2009. Flexible least squares for temporal data mining and statistical arbitrage. Expert Systems with Applications 36, 2819–2830.
Mudchanatongsuk, S., Primbs, J.A., Wong, W., 2008. Optimal pairs trading: a stochastic control approach. American Control Conference, 1035–1039.
Ng, V., Engle, R., Rothschild, M., 1992. A multi-dynamic-factor model for stock return. Journal of Econometrics 52, 245–256.
Nyblom, J., Harvey, A., 2000. Tests of common stochastic trends. Econometric Theory 16, 176–199.
Peña, D., Box, G.E.P., 1987. Identifying a simplifying structure in time series. Journal of the American Statistical Association 82, 836–843.
Peña, D., Poncela, P., 2004. Nonstationary dynamic factor analysis. Journal of Statistical Planning and Inference 136, 1237–1257.
Pole, A., 2007. Statistical Arbitrage. John Wiley & Sons, Hoboken, NJ.
Sargent, T. J., Sims, C., 1977. Business cycle modelling without pretending to have too much a priori economic theory. Working paper, Federal Reserve Bank of Minneapolis.
So, M., Lam, W., Li, K.W., 1998. A stochastic volatility model with Markov switching. Journal of Business & Economic Statistics 16, 244–253.
Stock, J.H., Watson, M.W., 1998. Testing for common trends. Journal of the American Statistical Association 83, 1097–1107.
Stock, J.H., Watson, M.W., 2004. Forecasting with many predictors. In: Elliot, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam.
Stock, J.H. Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.
Triantafyllopoulos, K., Montana, G., 2011. Dynamic modeling of mean-reverting spreads for statistical arbitrage. Computational Management Science 8, 23–49.
Tziritas, G., 1987. On the distribution of positive-definite Gaussian quadratic forms. IEEE Transactions on Information Theory IT-33, 895–906.
Vidyamurthy, G., 2004. Pairs Trading: Quantitative Methods and Analysis. John Wiley & Sons, Canada.