Weak Correlations of Stocks Future Returns

Ludovico Latmiral*

Kairos Investment Management Ltd, London W1H 6AZ, United Kingdom

We analyze correlations of stock returns via a series of widely adopted market and stock parameters which we refer to as explanatory variables. We subsequently exploit the results to propose a quantitative adaptive technique to infer predictions on expected relations among future stock returns.

I. INTRODUCTION

The links between financial ratios and assets returns have ever since been a fertile field of discussion in the finance industry [1]. Capital Asset Pricing Model (CAPM) is considered to be the first quantitative result in this direction, stating that expected returns of a given portfolio are linearly correlated with their market of reference. The model implies that regression coefficients of securities returns (in excess of the free-risk rate) with respect to market returns, i.e. the $\beta_i$, are sufficient to describe the cross section of expected returns [2].

Betas can be obtained from the expression $\beta_{i,m} = \rho_{i,m} \frac{\sigma_i}{\sigma_m}$, where $\rho_{i,m}$ is the correlation between the $i^{th}$ asset and the market, $\sigma_i$ is the volatility of the $i^{th}$ asset and $\sigma_m$ is the market volatility. We conclude that behind CAPM theory is the assumption of a linear relation between an asset mean return and its volatility. Furthermore, at the basis of linear regression there is ultimately a least squares minimization procedure, which generally assumes a Gaussian distribution for assets returns. This has eventually induced the identification of risk with volatility.

It is in this context that many studies have been devoted to identifying different kinds of risk premia: i.e. correlations between excess average returns and several possible factors of risk. It is worth mentioning the negative and positive correlations of returns respectively with market capitalization (CAP) [3] and with book-to-market ratio (BtM) [4]. E. Fama and K. French embedded all these considerations in their seminal papers (see Ref. [5, 6]), where they proposed a three factor model based on CAPM and BtM to explain the cross-sectional variation in average stock returns. Interestingly, the Authors also analyzed the relation of BtM and CAP with other widely used estimators. They provided evidence that BtM embeds both effects of book and market leverages, and together with CAP is able to fully explain the correlation of returns with the earnings-price ratio. Also, the small size of a company is shown to be generally linked with negative earnings and higher expected risks-returns.

In this paper we analyze the correlation of returns of stocks with a series of plausible explanatory variables. We adopt parameters which are uniquely dependent on the historical price series such as its volatility, its skewness and its beta with the market, and parameters which are linked to a bottom-up analysis of the company, e.g. book value, dividends, EBITDA, etc.. The strength of our approach lies in finding insightful information whose correlation with future performance of assets lasts on a longer time window than that on which the correlation is measured. In the first part of our study, we regress the returns of each stock over the whole set of explanatory variables computed both over the same, contemporary, time window and over an antecedent time window. This analysis provides useful information on the correlations among different variables, offering possible clues on the time-dependent cause-effect relations between them. Unfortunately, time delayed correlations generally reveal to be weak, actually discouraging from the possibility to fit future returns versus past historical data. Notably, however, we are able to identify an optimal subset of explanatory variables which we prove to have a significant connection with future returns. Our main result is precisely to shed light on these weak time correlations, study how they have reacted to different market environments in the last fifteen years and finally exploit them to make meaningful forecasts.

II. EXPLANATORY FIELDS

We have based our analysis on Bloomberg historical daily data for a set of 100 stocks picked from Standard and Poor’s 500 index (SPX). The selection process was almost random: we sorted companies ensuring to have sufficiently long historical track records as long as to obtain a trustful representative sample of SPX in terms of the GICS sectors (see Suppl. Mat. for more details). Together with prices, downloaded data included the aforementioned market capitalization and book-to-market value, as well as the enterprise value (EV), the dividend yield (DivYield) and the EBITDA. We also considered a set of four benchmark indexes: SPX, Brent Crude Oil, VIX index (VIX) and the generic US ten years government bond yield (US10Y), with respect to which we have computed correlations and $\beta_s$ for each stock under scrutiny. For the sake of a better readability, we briefly summarize below a list of the explanatory parameters we considered in our analysis [11].

- The correlation $(\rho_{s,b})$ between the stock under scrutiny and each of the benchmarks. This provides an estimate of the strength of the linear relationship between the two return time series.
- The beta $(\beta_{s,b} = \rho_{s,b} \frac{\sigma_b}{\sigma_b})$ is equal to the correlation
times the ratio between the standard deviations. It corresponds to the angular coefficient obtained with a least square linear regression of the asset returns versus the benchmark of reference.

- The past mean is computed as the algebraic mean of an asset return series \( x_a \) over a defined time interval in the past, i.e. \( \bar{x}_a[0:T] \). Depending on whether the market is following a momentum or a mean reversal trend, we expect it to be respectively positively or negatively correlated with the mean of future returns \( \bar{x}_a[0,t] \), for some time \( t \).

- The volatility (\( \sigma_a = \text{std}(x_a) \)) is commonly considered the best estimator for risk and a proxy for expected returns, accordingly with CAPM theory predictions. Our analysis will fairly confirm this assumption, even though we will stress that, similarly to the past mean, it is a momentum biased predictor.

- The sharpe ratio (Sharpe\(_a = \bar{x}_a / \sigma_a \)) characterizes how well the return of an asset compensates the investor for the risk taken. It is usually adopted as a benchmark to compare different assets.

Further to the above mentioned well known figures of merit, we devoted special attention to the skewness of the returns distribution and to correlations with correlations and betas with respect to meaningful benchmarks, (e.g. Fama and French factors or cross-correlations).

Let us briefly discuss these quantities in the next subsections.

### A. The skewness

In probability theory, skewness of a distribution is defined as the third standardized moment

\[
\zeta_a = \frac{\text{E}[(x_a - \bar{x}_a)^3]}{\sigma_a^3}, \tag{1}
\]

and it measures the asymmetry of the returns distribution about the mean. From a qualitative perspective, negative skew indicates that the tail on the left side is longer or fatter than the tail on the right side and vice versa. The quantity was proposed in Ref.\[7\] as a better proxy for risk than the standard deviation. In that paper Lemetrier et al. argue that the contribution of negative returns is mainly amenable to few severe drops, rather than to many small contributions. The authors further suggest to adopt a revised version of the skewness (which we will indicate with \( \zeta_a^* \)) that better fits with their intent to discard small returns fluctuations around the mean and gauge the contribution of the worst event scenarios. They first normalize the return series with zero mean and unitary variance, then rank them by their absolute value and finally consider the area underneath the compounded curve. Intuitively, what happens is that a large area is obtained when the average return is lowered by few very steep drawdowns, while there are many contributions slightly above the average which make the compounded graph grow fast since from the beginning.

### B. Cross Correlations

A meaningful quantity that has been raising additional interest, especially since the large spread of Exchanged-Traded Funds (ETFs), is the cross correlation among various traded stocks \[8\]. There is a vast amount of literature that has been dedicated to the study of asymmetric correlations, showing that stocks returns are more correlated in bear markets \[9\], and generally the exploitation of spurious anomalous correlations is a field of great interest for many quantitative funds. Here, we look at a synthetic index

\[
P(t) = \frac{1}{N^2} \sum_{\text{pairs}(i,j)} \rho_{i,j}(t), \tag{2}
\]

corresponding to the equally weighted average of all cross pair correlations among the set of stocks under scrutiny. Following the approach proposed by Fama and French \[5\], we then exploit it as a benchmark to evaluate correlations and betas with the excess return time series for each stock.

### C. Correlation with \( \rho \) and \( \beta \)

An explanatory field that deserves particular attention is the correlation between the mean of an asset returns and the correlation (and/or the beta) of the return series itself with any relevant benchmark. Let us use the Greek letter \( \eta \) to indicate alternatively \( \rho \) or \( \beta \). We have by definition

\[
\rho_{\eta,X} = \frac{\sum_i (\eta_i - \bar{\eta})(X_i - \bar{X})}{\sigma_X \sigma_{\eta}}, \tag{3}
\]

where the index \( i \) runs over all assets in the sample under analysis (e.g. a portfolio), \( X_t = T^{-1} \sum_{\tau} x_{i,\tau} \) \( (\eta_t = T^{-1} \sum_{\tau} \eta_{i,\tau}) \) is the average return over the time window under consideration \( T \) of the return series \( x_{i,\tau} \) \( (\eta_{i,\tau}) \) (here \( \tau \) indicates the discrete time at which the return is calculated with respect to the asset value at time \( \tau - 1 \)). The great advantage of considering these correlations is that there is in principle no restriction to compute the time averages for \( \eta_{i,\tau} \) and \( x_{i,\tau} \) over the same time window. Besides, while it would be meaningless to regress daily time series with a time lag of several months (i.e. to compute monthly lagged daily betas), we argue that we expect some factors, such as betas and volatility for example, to contain durable peculiar information of the related asset.

The understanding of this expression for \( \eta = \beta \) is pretty straightforward. Indeed, we know that \( \beta_i \) is defined as the best fit slope for the vector equation \( x_{i,\tau} = \beta_i M_{\tau} + q \), where \( M \) is the benchmark used for the regression and the expected value of the intercept \( q \) is zero if the linear fit assumption is satisfied. \( \beta_i \) can be computed via a least
We consider two time windows which were characterized actually, as discussed in Sec. II C, the plot suggests that April 2008 to April 2009, embraced the last financial crisis in the market and was followed by a period of mean reversion. While in Fig. 1(b) the market conditions for returns and with returns obtained on a subsequent annual time series, (see Fig. 1).

Together with SPX, VIX and US10Y, we will follow the approach suggested by Fama and French, adopting a set of explanatory fields as markets of reference for our correlation analysis. The procedure consists in sorting stocks in the market on the basis of a set of factors (e.g., EV/EBITDA, MtB, EV, DivYield) and then construct artificial indexes by subtracting the returns generated by stocks in the high rank of the index from those in the low part. We will show how these indexes play a significant role, especially when considering correlations with future returns.

The correlation of returns with correlations with the volatility, it seems to be anti-correlated with contemporary returns, irrespectively of the market environment. However, it is not a good estimator of future earnings, as it structurally lacks information to identify mean reversions.

The skewness of the returns distribution, reveals to be an excellent indicator of risk reward. When computed over the same time window, it grasps risk-return reward even more effectively than volatility. Most remarkably, it is positively and strongly correlated with contemporary returns also during the drawdown period April 2008 - April 2009, which indicates some bias with the average return. However, skewness seems to perform poorly in terms of future predictability.

While cross pairs correlations (ρp) are generally supposed to rise in crisis environment, they have also risen in the last years of bull market, mostly because of the increasing role played by ETF instruments, pension funds, etc. Nevertheless, they happen to be weakly related with stock returns in the periods under scrutiny and, at least in this format, they would not provide any significant predictive power.

The volatility Sigma (σ) is confirmed to be a highly relevant figure of merit, displaying high correlation with future returns series in a mean reversion environment. As a very important remark, however, it seems to be anti-correlated with contemporary returns in the drawdown period (see Fig. 1(a)), indicating that volatile stocks perform poorly in bear markets, though they outperform in bull markets.

The correlation of returns with correlations with various explanatory factors (ρp,x) does not account any significant relevance in neither the time periods nor for contemporaneous or future returns.

Betas with respect to SPX, VIX (respectively the market to which the stocks under scrutiny belong to and the associated volatility index) and US treasuries show promising correlations with average stock returns.

As expected from past literature, betas with respect to EV and MtB were negative in 2008-2009, while they are reverted when computed for lagged returns.

\[ \frac{d}{d\beta_i} \sum_{\tau} [(x_{i,\tau} - X_i) - \beta_i (M_\tau - \bar{M})]^2 = 0, \]

which leads to \( X_i = \beta_i \bar{M} \). With this information at hand and substituting in Eq. (3), we expect \( \rho_{\beta,X} \) to have the same sign of the average \( \bar{M} \). This is a trivial though insightful consideration, meaning that the average return of an asset is positively (negatively) correlated with the \( \beta \) with respect to its reference benchmark depending on the expected value of the benchmark itself.

On the other hand, there is no immediate interpretation of the correlation \( \rho_{p,x} \), whose sign could generally be independent of the correlations \( \rho \). For a qualitative understanding of the differences between \( \rho_{\beta,X} \) and \( \rho_{p,x} \), it is helpful to consider the contribution of the volatility, which is positively correlated with the returns and strongly correlated with \( \beta \), though weakly and oppositely correlated with \( \rho \).

III. CORRELATIONS BETWEEN EXPLANATORY FIELDS AND RETURNS

Now that we have introduced the analytical tools and factors at the basis of our analysis, we present the first part of this paper, consisting in the study of correlations between a set of possible explanatory fields and return time series (see Fig. 1). We consider two time windows which were characterized by almost opposite market conditions. The former, from April 2008 to April 2009, embraced the last financial crisis and was marked by considerable volatility on the market and was followed by a period of mean reversion growth. The latter spans from June 2016 to June 2017 and refers to a period of almost stable growth with low volatility. In both cases we plot yearly correlations of a set of (possible) explanatory fields with contemporary returns and with returns obtained on a subsequent annual time window. While in Fig. 1(b) the market conditions for the two years remained the same and all the correlations maintained their signs, the mean reversion that took place between June 2008 and June 2010 is reflected in a switch of sign in Fig. 1(a) for all the most significant parameters. Actually, as discussed in Sec II C, the plot suggests that contemporary correlations with respect to volatility and to the most relevant betas change sign in drawdown market periods (with the exception of the correlation with the revised skewness \( \zeta^* \)). This is also confirmed by several identical analysis over different time windows embracing the 2008 crisis.

We believe the comparison between the two time periods of both contemporary and time delayed correlations to be very insightful. Hereafter we summarize the most relevant aspects in a point by point discussion.

- As expected by definition, the sharpe ratio is highly correlated with contemporary returns, irrespective of the market environment. However, it is not a good estimator of future earnings, as it structurally lacks information to identify mean reversions.
- The skewness of the returns distribution, reveals to be an excellent indicator of risk reward. When computed over the same time window, it grasps risk-return reward even more effectively than volatility. Most remarkably, it is positively and strongly correlated with contemporary returns also during the drawdown period April 2008 - April 2009, which indicates some bias with the average return. However, skewness seems to perform poorly in terms of future predictability.
- While cross pairs correlations (ρp) are generally supposed to rise in crisis environment, they have also risen in the last years of bull market, mostly because of the increasing role played by ETF instruments, pension funds, etc. Nevertheless, they happen to be weakly related with stock returns in the periods under scrutiny and, at least in this format, they would not provide any significant predictive power.
- The volatility Sigma (σ) is confirmed to be a highly relevant figure of merit, displaying high correlation with future returns series in a mean reversion environment.
- As expected from past literature, betas with respect to EV and MtB were negative in 2008-2009, while they are reverted when computed for lagged returns.
FIG. 1: Measured correlations of a set of explanatory fields with the average returns achieved in the same time window (upper diagrams) and in the following trading year (lower plots). Histograms (a) and (b) are referred, respectively, to a significantly volatile period (including the financial crisis in 2008) and a period of smoother growth and lower volatility. The analysis has been conducted over 100 stocks picked from the S&P500 index. Error bars indicate the 95% confidence interval for each correlation.

We infer that this is because the largest companies that survived the crisis, and that were subsequently able to master financial leverage, were also those which outperformed in the recovery period. A point that should be further investigated is the decreasing in magnitude of the correlations with $\beta_E$ and $\beta_{M/B}$.
over years.

- Eventually, the dividend yield and the ratio $\text{EV/EBITDA}$ seem to have minor correlations with return series, the former acquiring larger significance in scenarios of stable growth (see Fig. 1(b)).

IV. TIME CORRELATIONS WITH EXPLANATORY PARAMETERS

The analysis on the correlations in the previous section reveals that observables measured in a certain time frame generally have little explanatory power on the dynamics in the subsequent time frame. On the other hand, it shows that a significant amount of bias is amenable to the market environment when the study is carried out.

Besides, de-trending the return series does not guarantee a significant improvement in this direction and several attempts have failed to bring any further stability in the correlations, nor any significant insight on their dynamics. As a matter of fact, it has historically been a hard task to foresee drawdowns in the market, and in those rare cases when this was possible, evidence was usually found in macro-economical or geo-political factors, rather than in historical time series \[^{10}\].

It is in this context that we trust an accurate and plain analysis of stock picking techniques, that could be applicable within different market environments, would be of high relevance. As a matter of fact, we believe that a reasonable convenience can derive from a change in perspective: instead of aiming at predicting future asset returns, we focus on forecasting the relationships amongst those returns. We do not attempt to investigate whether investing in a given market could be profitable in a certain macro economical and political environment. Instead, once an investment decision has been made (\textit{e.g.} a long position in the US stock market), we try to address the problem of an efficient allocation of the resources.

In the following we will thus present two possible approaches to deal with asset allocation and backtest their effectiveness from 2003 until today. The first method consists in the well known creation of a portfolio via the efficient frontier, while the second is a new proposal that takes into account the correlations analyzed in the previous section.

We will devote the following two subsections to the presentation of these methods, which are both based on a monthly reallocation of the stocks in the portfolio.

A. Allocating via Efficient Frontier

Given a set of assets, each with an historical average return and volatility, the efficient frontier can be defined as the ensemble of all portfolios that minimize volatility, given an expected return. In our backtested simulation, on each asset allocation date we considered the time series of the previous year and retrieved on this basis the efficient portfolio having the same volatility as the equally weighted portfolio containing all the 100 stocks, but higher expected return. Resources were monthly allocated with the following simple rule: the portfolio was fully invested with equal weights only on those stocks whose proposed optimal weights were larger than the average proposed weight (\textit{i.e.} $w_i > 1/N$, where $N = 100$ is the total number of stocks in the basket under consideration and $w_i$ the proposed weight for the $i^{th}$ asset). In the following, we will refer to this method as $\text{EF}$. 

B. Allocating via Returns Correlations

Returns correlations are the main focus of this paper and we propose a technique to exploit them to construct effective portfolios (we will denote such protocol with the shortcut $\text{RC}$). The idea is to base every allocation on the regression of return time series of the last year with respect to the explanatory fields measured over the previous year, \textit{i.e.} with one year time lag with respect to the returns (overall, this method thus requires two years of backtest data). The multi-regression coefficients are then used, together with explanatory fields measured in the latest year, to infer returns for the incoming year. Resources are then fully and equally allocated in those assets whose predicted returns outperform the average predicted return of the whole set of $N$ stocks.

The effectiveness of this approach crucially depends on the ability to capture future returns through a set of explanatory fields. Thanks to the analysis presented in Sec. \[^{III}\] we are able to select the most significant and stable parameters. We choose eight factors: the revised skewness $\zeta^*$, the volatility and six betas with respect to SPX, US10Y, VIX, DivYield, EV and MtB.

V. DISCUSSION

We plot in Fig. 2 a backtest of the proposed strategies, together with a comparison with the equally weighted, monthly rebalanced portfolio (referred to as $\text{EW}$): both strategies under scrutiny significantly outperform $\text{EW}$. In particular, we show that a mixed strategy ($\text{MIX}$), which allocates by averaging between $\text{EF}$ and $\text{RC}$, provides high returns with limited volatility. A summary of the results with the most relevant figures of merit is presented in Table I: $\text{RC}$ outperforms $\text{EW}$ for 104 months over 176 (the period January 2003 - June 2018), which corresponds to around $\approx 59\%$ of cases ($\text{MIX}$ reveals to be a winning strategy in $63\%$ of cases), and exhibits a volatility and a sharpe ratio which are respectively $6\%$ and $24\%$ larger than $\text{EW}$.

A figure of merit that deserves special attention from a statistical perspective is the confidence level with which
Compounded backtests

FIG. 2: Compounded monthly backtest of the discussed strategies from January 2003 to June 2018: monthly rebalanced equally weighted portfolio with 100 stocks (blue circles), allocation via efficient frontier (EF) (purple crosses), returns correlations (RC) (red triangles), mix strategy (MIX) (green rhombuses) and adaptive returns correlations (yellow squares).

a strategy can be distinguished from a random allocation of weights (we indicate it with Fidelity in Table I). In other words, it is mandatory to demonstrate that our allocation truly contains some extra information on future returns distribution, which confers it effectiveness in the stock picking process. Hence, we compare each monthly asset allocation with EF and RC techniques (they are both supposed to select the outperforming stocks within a basket of \(N\) titles) with the true historical subsets of out- and under-performing stocks. If there was no predictive power, the logical and comparison between forecasts and true results would behave as a random binomial variable with probability \(p\) equal to the probability that a stock randomly picked from a basket has greater returns than the average basket return. We expect \(p \approx 0.5\), with \(p \to 0.5\) for large \(N\). Satisfactorily, both our approaches are able to dismiss the random hypothesis and RC technique is slightly more precise than EF at selecting the outperforming stocks with a confidence level of almost 90%. This is an encouraging result and can be explained by considering how the two procedures are constructed. RC allocation is expected to fulfill this precise scope, i.e. to identify the subset of outperforming stocks, while EF approach does not necessarily select the assets with higher returns, but it accounts also for low volatility, generally resulting in a higher differentiation amongst stocks in the portfolio (see sharpe ratios and means in Table I).

There is another promising remark that Fig.2 brings to our attention. As expected from its back-looking approach, EF is strongly momentum biased and performs particularly well in periods of stable growth, while it deeply underperforms during drawdowns. Instead, this bias is well handled by our proposal that uses returns correlations. On the one hand, RC losses during the drawdowns in 2008, end of 2011 and in 2015 are comparable with the decreasing of the EW basket, significantly reducing the maximum drawdown registered by EF (see Table I). On the other hand, RC is extremely efficient at exploiting market rebounds after drawdowns, so that it largely outperforms EF both in the period 2008-2009 (+7%) and in the subsequent recovery 2009-2010 (+57%). A similar situation is replicated in 2015-2016 and 2016-2017.

This ability can be further enhanced capitalizing what we learned in Sec. III. Indeed, we have observed that corre-

| Return | Sharpe | Max DD | Months+ | Fidelity |
|--------|--------|--------|---------|----------|
| Eff Frontier | 21.7% | 1.08 | -55% | 103 | 0.83 |
| Returns Corr | 19.6% | 0.94 | -48% | 104 | 0.87 |
| Returns Corr* | 20.8% | 0.95 | -47% | 107 | 0.88 |
| MIX | 20.7% | 1.05 | -51% | 110 | - |
| Equal Weights | 15.8% | 0.85 | -46% | - | - |

TABLE I: Comparison of the investment strategies described in the text over the period January 2003 - June 2018. We consider annualized returns, sharpe ratio and the number of months in which the strategy has outperformed the allocation with equal weights (Months+). Max DD is the maximum annual drawdown and Fidelity is the average fidelity (over the entire set of 176 months) with which the strategy can be distinguished from a random allocation.
lations of betas with contemporary returns are inverted in strongly bear markets, correlation with the volatility being the first indicator of such a reversion. Since market drawdowns usually last for a short period of time with respect to the whole economic cycle, when this condition is detected, we could adaptively revert the sign of the coefficients in the multi-regression analysis. This shrewdness allows to exploit the market rebound even better and boosts performance after a negative period (+150% in 2009). The result is plotted in Fig.2 and reported in Table I with the label “Returns Corr*”.

VI. CONCLUSIONS

Instead of looking at the cross section of stock returns, i.e. trying to explain the dependency of historical time series on a set of contemporary benchmark time series, we have investigated correlations of (average) asset returns with respect to several explanatory fields, i.e. quantities that are supposed to deliver durable information on assets performance. We have specifically looked at the correlations of returns with betas with respect to widely recognized financial indexes (SPX, VIX, Oil and US10Y) and factors (EV, DivYield, MtB, EBITDA). This approach was suggested by what we believe would be the interest of a generic long-short fund manager who is always concerned about the future performance of his/her portfolio with respect to relevant benchmarks. We have chosen not to concentrate on the analysis and the (extremely hard) prediction of daily, contemporary time series, but rather on the relationship among different time-lagged return series. While the former problem is usually tackled looking at different betas, we furthered the analysis studying the relationship between past betas, volatility and skewness, and present average stock returns. We have thus defined a set of explanatory variables which we have adaptively used to understand which assets were likely to outperform their basket of reference. We have provided convincing empirical and statistical analysis on a fifteen years period of backtest data (January 2003 - June 2018) to corroborate our results. On the one hand, our method managed to achieve an overall annual performance 25% higher than that of an equally weighted portfolio, almost comparable with the one obtainable with an efficient frontier technique. On the other hand, it demonstrated a very low momentum bias, being able to deal well with 2008-2009 market crisis, reducing the drawdown registered with EF by 7% and increasing the rebound by 57%.

VII. ACKNOWLEDGEMENTS

The author is very grateful to the Multi-manager team of Kairos Investment Management Ltd for insightful discussions on the subject of this paper. The opinions and views expressed in this paper are uniquely those of the author, and do not necessarily represent those of Kairos Group. The Author also declare no competing financial interest.

Supplementary Material

A. Basket Composition

We report in this section further details on the basket of 100 stocks that was used for the analysis (we would be pleased to provide the full list on demand, we simply refrain here for the sake of compactness). The pie chart in Fig.1 shows the composition of the basket (equally weighted) in terms of GICS sectors. Even though there is no fundamental reason for which our selection of stocks should replicate S&P500 index, a strong correlation with the index would witness the diversification of our basket and the generality of our results. We thus computed the return series of a basket containing our stocks and arranged with equal weights on 1st of January 2002 (with no subsequent rebalancements) and satisfactorily measured a correlation of 0.967 with S&P500 index over the time range until June 2018.

Suppl. Fig. 1: Equally weighted basket composition in terms of GICS factors.

[1] Welch, I. & Goyal, A. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508 (2008).
[2] Sharpe, W. F. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finance 19, 425–442 (1964).
[3] Banz, R. W. The relationship between return and market value of common stocks. Journal of financial economics
9, 3–18 (1981).

[4] Stattman, D. Book values and stock returns. *The Chicago MBA: a journal of selected papers* 4, 25–45 (1980).

[5] Fama, E. F. & French, K. R. The cross-section of expected stock returns. *The Journal of Finance* 2, 427–465 (1992).

[6] Fama, E. F. & French, K. R. A five-factor asset pricing model. *Journal of financial economics* 116, 1–22 (2015).

[7] Lemperiere, Y. *et al.* Risk premia: asymmetric tail risks and excess returns. *Quantitative Finance* 17, 1–14 (2016).

[8] Pollet, J. M. & Wilson, M. Average correlation and stock market returns. *Journal of financial economics* 96, 364–380 (2010).

[9] Hong, Y., Tu, J. & Zhou, G. Asymmetries in stock returns: Statistical tests and economic evaluation. *The Review of Financial Studies* 20, 1547–1581 (2006).

[10] Kroencke, T. A. Recessions and the stock market. *SSRN* 3161979 (2018).

[11] In the following we will always refer to returns after that the risk-free rate has been subtracted.