Are Extreme Returns Priced in the Stock Market? European Evidence

— **Source link** ↗

Jan Annaert, Marc J.K. De Ceuster, Kurt Verstegen

Institutions: University of Antwerp

Published on: 01 Feb 2013 - Social Science Research Network

Topics: Returns to scale, Stock market, Systematic risk, Stock (geology) and Capital asset pricing model

Related papers:

- Stock Option Returns, a Puzzle

- Idiosyncratic skewness and cross-section of stock returns: Evidence from Taiwan

- Autocorrelated Order-Imbalance and Price Momentum in the Stock Market

- Carry and Trend Following Returns in the Foreign Exchange Market

- Learning, Asset-Pricing Tests, and Market Efficiency

DEPARTMENT OF ACCOUNTING AND FINANCE

Are extreme returns priced in the stock market?
European evidence

Jan Annaert, Marc De Ceuster & Kurt Verstegen

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ACCOUNTING AND FINANCE

Are extreme returns priced in the stock market?
European evidence

Jan Annaert, Marc De Ceuster & Kurt Verstegen

Are extreme returns priced in the stock market?
European evidence[*]

Jan Annaert

(Corresponding author)

University of Antwerp & Antwerp Management School

Prinsstraat 13 – 2000 Antwerp – Belgium

Tel.: +32 (0) 3 265 41 63 – E-mail: jan.annaert@ua.ac.be

Marc De Ceuster

University of Antwerp & Antwerp Management School

Prinsstraat 13 – 2000 Antwerp – Belgium

Tel.: +32 (0) 3 265 41 24 – E-mail: marc.deceuster@ua.ac.be

Kurt Verstegen

University of Antwerp

Prinsstraat 13 – 2000 Antwerp – Belgium

Tel.: +32 (0) 3 265 41 20 – E-mail: kurt.verstegen@ua.ac.be

This draft: September 2012

Abstract

This paper revisits some recently found evidence in the literature on the cross-section of stock returns for a carefully constructed dataset of euro area stocks. First, we find evidence of a negative cross-sectional relation between extreme positive returns and average returns after controlling for characteristics such as momentum, book-to-market, size, liquidity and return reversal. We argue that this is the case because these stocks have lottery-like characteristics. Second, when we control for this relation, the idiosyncratic volatility puzzle seems to disappear. When extreme positive returns are included in the regression, we find a weak but positive relation between idiosyncratic volatility and returns. Lastly, the maximum return effect holds when we control for skewness. Moreover, skewness is on its own negatively related to returns in our sample, as several asset pricing models predict.

---

1   Introduction

Despite decades of research, it is still not completely clear what determines cross-sectional variation in expected stock returns. It is well accepted that the three factor model of Fama and French (1993) goes a long way in capturing this return variation. Efforts have been made to interpret these factors as fundamental risk factors, making the model consistent with rational asset pricing, see e.g., Aretz, Bartram, and Pope (2010) or Guo, Savickas, Wang, and Yang (2009). Nevertheless, empirical evidence keeps discovering other characteristics that are related to average stock returns (see Fama and French (2008) for a recent overview). As most of these characteristics do not follow directly from theory, they are subject to the data-mining critique implying the patterns are mere statistical flukes (Lo & MacKinlay, 1990). In any dataset, some significant statistical results are bound to be found just by chance. One way to address this critique is to look for corroborating evidence on other markets or from other periods.

In this paper, we verify the U.S. results of Bali, Cakici, and Whitelaw (2011) on a carefully constructed euro area stock market database covering more than thirty years. More specifically, Bali et al. (2011) find a statistically and economically significantly negative relation between the maximum daily return over the past one month and expected stock returns. They argue that this captures individual investors' preference for lottery-like stocks, i.e. stocks that have a low probability of a huge profit and a large probability of a small loss as shown by Kumar (2009). Although this is an idiosyncratic characteristic, demand by individual investors may lead to higher prices (and lower expected returns) for these stocks, given that these investors typically hold under-diversified portfolios, see e.g., Odean (1999) and Goetzmann and Kumar (2008). We show that this effect also exists in the euro area. Moreover, it is unlikely to be arbitraged away by large investors as the typical stock with extreme positive returns is relatively small, illiquid and has relatively high idiosyncratic volatility. Even when short selling these stocks were possible, it would expose arbitrageurs to considerable risk. Hence, it is plausible that individual investors drive the pricing of such stocks. We do not find evidence for the inverse relationship: extreme negative returns are not positively related to expected returns. As only long positions are needed to trade on this pattern, it may be that demand from institutional investors weakens the effect on extreme negative returns.

We also disentangle the comovement between extreme returns and idiosyncratic volatility. Taken on its own, idiosyncratic volatility is negatively related to expected returns. The relation is statistically significant and goes in the direction of the puzzling results of Ang, Hodrick, Xing, and Zhang (2006) for the U.S. and Ang, Hodrick, Xing, and Zhang (2009) for an international sample of stocks. However, when we control for extreme returns, the relation between idiosyncratic risk and expected returns becomes positive, albeit insignificant. This insignificance could be the result of either a true parameter that is (very close to) zero, or multicollinearity in the regression due to high correlation between maximum returns and idiosyncratic volatility, causing monthly Fama-MacBeth estimates to fluctuate more.

Finally, the influence of skewness is investigated. Skewness is positively correlated with positive extreme returns. Three moment asset pricing models as advanced by Kraus and Litzenberger (1976) or Harvey and Siddique (2000) imply that investors have a preference for assets that increase portfolio return skewness. In equilibrium such assets would have lower expected returns. Return skewness may therefore cause the negative relation we find between extreme positive returns and expected returns. As in Bali et al. (2011), we find that the extreme positive return effect is robust to including total skewness, idiosyncratic skewness or coskewness. But unlike their results, we also find that skewness is significantly negatively related to expected returns, no matter which skewness measure is used.

The remainder of the paper is structured as follows. In section 2, we discuss sample selection, construction of variables and filters. Next, we discuss the main results in section 3. Section 4 provides the results of a battery of robustness checks and finally, we conclude.

2 Data

2.1 Sample selection

Our sample comprises thirteen European countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Slovakia and Spain. One might notice that the other four countries of the euro area (Malta, Estonia, Cyprus and Slovenia) are not included. This is inspired by Schmidt, von Arx, Schrimpf, Wagner, and Ziegler (2011), who also exclude these countries. This is motivated by the fact that their financial impact on the euro area is negligible. Their relative share, measured in 2009 nominal GDP, is only 0.80% of the entire euro area. We are therefore convinced that the current sample is representative for the entire euro area. All data comes from Thomson DataStream (TDS) as in Ang, Hodrick, Xing and Zhang (2009). We use several lists covering both active shares and delisted shares[1] thereby minimizing survival bias in our sample. The resulting dataset is then subjected to extensive filtering as described in appendix A in order to select only common stock issues. This results in a sample of 7,861 European companies. For these companies, we download the end-of-month return index (including dividends), unadjusted stock price, market capitalization (MC) and book-to-market ratio (B/M), from 31 December 1979 to 30 June 2011. Additionally, we download the daily total return index and MC over the same period. TDS automatically calculates B/M by dividing the book value per share by market value per share at time t, where book value per share is the company's book value at the company's last fiscal year end (Worldscope item 05476). For the pre-1999 period all data are converted by TDS into synthetic euro. As the risk-free rate, R_f, the monthly money market rate as reported by Frankfurt banks[2] is used.

2.2 Construction of variables

All returns are calculated using the TDS total return indices. Two corrections are applied to correct for errors that occasionally occur in the TDS database, inspired by Ince and Porter (2006) and Schmidt et al. (2011). First, we need to think about decimal errors. Suppose the return index is 101.52 on a particular day and does not change the next. Also suppose TDS correctly stores 101.52 the first day, but erroneously stores it as 1015.20 the second day. An observed return of 900% when the true return is zero would obviously distort results. Let's call this a right-decimal error, because the decimal moved erroneously to the right. Alternatively, suppose TDS erroneously stores the return index on the second day as 10.152, which would result in a -90% return. Let's call this a left-decimal error. These examples all show nonzero returns while the true return is zero. Additionally, we could have decimal errors when the true return is nonzero. For example, when the return index decreases from 101.52 to 96.44 (-5% true return) but is stored as 964.40, resulting in an observed return of 849.96%. This example shows that there is a need to account for decimal errors in both directions, whether the true return is zero or nonzero. We therefore set to missing any returns that are above 400% (a -50% true return accompanied by right-decimal error) or returns that are below -85% (a 50% true return accompanied by a left-decimal error). A second correction is to set R_t and R_{t-1} to missing if R_t or R_{t-1} is greater

---

[1] The lists are: WSCOPEOE, ALLAS, DEADOE (Austria); WSCOPEBG, FBDO, DEADBG (Belgium); WSCOPEFN, FFIN, DEADFN (Finland); WSCOPEFR, FFRA, ALLFF, DEADFR (France); WSCOPEBD, FGER1, FGER2, DEADBD1, DEADBD2 (Germany); WSCOPEIR, FIRL, DEADIR (Ireland); WSCOPEIT, FITA, DEADIT (Italy); WSCOPENL, FHOL, ALLFL, DEADNL (Netherlands); WSCOPEPT, FPOM, FPOR, FPSM, DEADPT (Portugal); WSCOPEES, FSPN, DEADES (Spain); WSCOPELX, FLUX, DEADLX (Luxembourg); WSCOPEGR, FGREE, FGRPM, FGRMM, FNEXA, DEADGR (Greece); FSLOVAK, FSLOVALL, DEADSLO (Slovakia).

[2] This rate can be found on the website of the Deutsche Bundesbank in the time series database .

than 300% and $(1 + R_{t-1})(1 + R_t) - 1$ is less than 50% (indicating extreme reversal), both for monthly and daily returns. Excess returns are calculated using the risk-free rate. *MC* and *B/M* are not corrected as no obvious errors are detected. In the analyses that follow, we use twelve lags for *B/M* to ensure that accounting data is always available to investors at the time.[3] Since Jegadeesh and Titman (1993), the momentum effect is widely recognized and accepted in the literature on cross-sectional return predictability. Therefore, we also include it in our regressions and sorting algorithms. As in Fama and French (2008), momentum (*Mom*) in month t equals the buy-and-hold return of a particular stock over period t-12 to t-2. We use the return of month t-1 (*LagRet*) separately to control for the short-term reversal effect documented by Jegadeesh (1990). We compute a very simple measure of expected illiquidity (*Illiq*) inspired by Bekaert, Harvey, and Lundblad (2007). *Illiq* in month t is the proportion of zero returns observed over the last 260 trading days prior to month t. We could not calculate the measure of Amihud (2002), since for the 2,971,458 firm-months, only 649,433 observations (or 22%) on volume are available. Using this measure would result in a major loss of data. Lastly, we use the skewness coefficient (*Skew*) of the last 260 daily returns prior to month t as a proxy for expected skewness. If not all 260 daily returns are available, we ignore the missing values, but a minimum of 65 daily observations is required. If not, *Skew* is set to missing. We also calculate the four factors (market, SMB, HML and WML) advanced by Fama and French (1993) and Carhart (1997) to calculate alphas of formed portfolios. The calculation of the factors is based on our own dataset. A detailed description is available in appendix B.

As admitted by Bali et al. (2011), who estimate beta over a month using daily data, this estimate is subject to a significant amount of measurement error. We therefore decide to follow another, more robust approach. Market beta, SMB beta and HML beta in month t are calculated using a minimum of 24 and a maximum of 60 monthly excess stock returns from months prior to month t. If less than 24 returns are available, market beta is set to missing. The following time-series regression model is estimated:[4]

$$R_{i,t} - R_{f,t} = \alpha_i + b_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}.$$

If the beta estimate is smaller than -1 or larger than 5, we set it to missing. These boundaries are quite arbitrary. However, the estimated betas are actually proxies of expected future beta and it would be hard to argue that any investor would expect more extreme betas at any point in time. Not using these boundaries would result in betas as high as 50. We also need a proxy for expected idiosyncratic volatility (*IVol*) in month t. To account for the time-varying volatility of stock returns, *IVol* for month t is calculated as the root mean square error from a daily three factor model (where days are indexed by d) based on the 20 days prior to month t:

$$R_{i,d} - R_{f,d} = \alpha_i + b_i(R_{m,d} - R_{f,d}) + s_i SMB_d + h_i HML_d + \varepsilon_{i,d}.$$

This is similar to the methodology of Ang et al. (2009).[5] Lastly, we use minimum and maximum returns to measure extreme returns. More specifically, we use *MIN(N)* or *MAX(N)* to denote the average of

---

[3] Using a constant lag of 12 months does not always reflect the most recent information. We would sometimes use the book-to-market ratio of 12 months ago when in fact new accounting data would have recently reached the market. We argue that the book value of equity does not change that much in 12 months and that short term variation in book-to-market ratios will mainly be caused by variation in stock prices. The fact that the value effect will still turn out to be very strong (see later) confirms this intuition.

[4] We refrain from adding the WML factor as exposures with respect to it are much less persistent due to the relatively short-lived nature of momentum. Nevertheless, when we do use the WML exposure, differences in b, s and h from both factor models are minimal: correlations are respectively 0.93, 0.97 and 0.89. Moreover, the correlation between market beta from the three factor model and traditional CAPM beta is 0.88, or 0.80 in case we use Dimson (1979) beta with one lag. The usage of other estimators in our analysis does not impact the results in any significant way.

[5] The correlation between our *IVol* measure and the regular standard deviation of the excess returns used to estimate *IVol* is 0.99. Monthly cross-sectional regressions of *IVol* on the standard deviation yield an average slope coefficient of 1.00. This shows that in the short run it does not really matter whether we use *IVol* or regular standard deviation.

respectively the *N* lowest or highest returns within a given estimation window. For example, *MIN(1)* would be the minimum return, while *MAX(3)* would be the average of the three highest returns. However, we always use the absolute value of *MIN(N)* for easier interpretation. The estimation window always contains 20 trading days prior month *t*. This is done to ensure that an investor trying to make a decision at the start of month *t* has all measures at her disposal. Sometimes we refer to *MAX* or *MIN* instead of *MAX(N)* or *MIN(N)* to avoid copious notation.

3   Results

In this section we discuss the results or our main analyses and robustness checks. We start with extensive sorts, then move on to the persistence of *MAX* and end with cross-sectional regressions at the firm level.

3.1   Sorts

In panel A of table 1, each month *t*, we sort stocks based on *MAX(N)*, with *N* ranging from 1 to 5. We then compute returns for the ten decile portfolios. Portfolio returns are either equally-weighted (EW) or value-weighted (VW). Each month, before sorting, the cross-section of returns is winsorised at the 0.5% and 99.5% levels to ensure that results are not driven by extremes, although the effect of winsorising returns for portfolio sorts or regressions is found to be negligibly small. The table reports the time-series averages of EW and VW portfolio returns along with Newey and West (1987) adjusted *t*-statistics for the hedge portfolio, which is long in the high portfolio and short in the low portfolio. Four-factor alphas of the hedge portfolio and their Newey-West adjusted *t*-statistics are reported as *hedge 4F alpha*. We do not find significant indications of a negative *MAX* effect. Hedge returns are all negative, but show little statistical significance and are unimpressive, while alpha is only statistically significant for sorts on *MAX(1)*, with a *t*-statistic of -2.06. All other EW or VW hedge returns or four-factor alphas are negative but cannot be significantly distinguished from zero. The effect weakens as the maxima are calculated using larger *N*. In any case, these results are much less impressive than those of Bali et al. (2011), who find strong negative hedge returns and alphas for these sorts. These weak results can potentially be explained by looking at panel B of table 1. Here, each month *t* we compute the median of the cross-section of firm characteristics for each decile portfolios. The time-series averages of these medians are reported. We find that high *MAX* stocks have larger market betas, smaller *MC* and higher *B/M*: all effects that have been found to increase average stock returns and therefore could distort or weaken the hedge returns found in panel A. On the other hand, high *MAX* stocks also tend to have lower momentum, higher lagged returns and higher skewness. These characteristics would bias the results towards lower returns. Also, high *MAX* stocks tend to be more volatile. In order to get a clearer picture, we first resort to bivariate sorts that control for other effects. Afterwards, we turn to multivariate regressions. Also note that high *MAX* stocks can indeed be considered as lottery-like stocks, since they have lower prices, higher idiosyncratic volatility and higher skewness. All these are characteristics of lotteries.

For table 2, we first sort stocks each month *t* on a control variable: market beta, *MC*, *B/M*, *Mom* or *LagRet*. We form 10 portfolios. Then we sort all stocks within each decile portfolio on *MAX(1)*. We form 10 portfolios within each control group so that 100 (10x10) portfolios are obtained, each with roughly an equal number of stocks. However, for brevity, we only report portfolio returns that are averaged across the spectrum of the control variable to produce decile portfolios with variation in *MAX* but with similar levels of the control variable, as do Bali et al. (2011). As such, we report the "average" *MAX* effect over all control deciles. Correspondingly, hedge returns are also averaged over the 10 portfolios, so that the hedge return would actually be the return we would earn if we would invest one tenth of our money in each hedge portfolio. We find that the hedge returns are always

negative. Moreover, they are statistically and economically significant for EW sorts on *Beta* (t = -1.96), on *B/M* (t = -2.54) and on *LagRet* (t = -2.74). For VW sorts, the hedge returns are not statistically significant. The hedge alphas are more convincing. They are all statistically and economically significant for all EW sorts and for all VW sorts but those on *B/M* and *LagRet*. In panel B we can see that using *MAX(5)* decreases the effect, since all hedge returns are now insignificant and a some alphas are no longer significant. The results of the bivariate sorts are more convincing than those of the univariate sorts. It is clear that controlling for other variables is very important to get a clear picture of what is going on.

3.2   Persistence of MAX

If past lottery-like behavior does not lead to future lottery-like behavior, i.e. if *MAX* is not persistent, then there is no reason for investors to prefer high *MAX* stocks, especially given that they tend to have lower expected returns. Also, it would be incorrect to use past *MAX* as a proxy for expected future *MAX*. Therefore, we investigate the persistence of *MAX* by forming a month-to-month portfolio transition matrix, which shows the probability $P_{i,j}$ that a stock in decile *i* will be in decile *j* the next month. If *MAX* is not persistent (or random) we can expect all probabilities to be approximately 10%. Table 3 shows the transition probabilities for *MAX(1)*. We find that the diagonal transition probabilities are always above 10%, especially for the outer deciles (low to low and high to high). Stocks in decile 10 (high *MAX*) have a 26.1% probability of remaining there next month. They even have a 52.9% probability of being in decile 8, 9 or 10 in the next month. Results are robust to other levels of *N*. In fact, the probability of high *MAX(N)* stocks to be in decile 8, 9 or 10 the next month increases to 55.5%, 57.0%, 57.6% and 57.4% for *N* set to respectively 2, 3, 4 and 5. Lastly, we also calculate transition probabilities for longer in-between periods. Stocks in the highest *MAX* decile have a 24.2% probability of being there in two months, 23.0% in three months, 21.0% in six months and 18.9% in twelve months. Similarly, they have a 49.6% probability of being in decile 8, 9 or 10 in two months, 47.9% in three months, 44.9% in six months and 41.6% in twelve months.

We also assess persistence by cross-sectionally regressing *MAX* onto its lag (results not reported). We find a coefficient for lagged *MAX* of 0.38 (t = 30.46) with an average R^2 of 14%. Lastly, we examine the total distribution of stock returns for the high and low *MAX* deciles. Each month we place stock returns into one of 10 groups, based on their *MAX(1)*. As we move through the sample, we keep adding returns to the 10 groups. At the end of this process we have 10 groups that each contain about 65,000 returns. Then we calculate descriptive statistics for the outer two groups, which contain next-month returns of stocks that were in the lowest or highest *MAX* at a given point in time. However, we do trim each group of returns at the 0.5% and 99.5% level to be sure that results are not driven by the most extreme observations. The equally-weighted average return of the low *MAX* group is 0.47%, while that of the high *MAX* group is -0.25%. The medians are respectively 0.12% and -1.20%. This again shows that high *MAX* stocks underperform low *MAX* stocks. The high *MAX* group has a much higher standard deviation (17.72%) than the low *MAX* group (6.56%). Also, we find that the skewness of the high *MAX* group is 1.17, compared to 0.51 for the low *MAX* group. While the median is lower, all other percentiles above and including the 75th percentile lie much higher for the high *MAX* group. For example, the 75th percentile is 6.36% for the high *MAX* group, while 1.61% for the other. For the 90th percentile we find respectively 18.89% versus 6.69%, for the 99th percentile we find 63.22% versus 23.48% and lastly, for the maxima, we find 103.70% versus 41.70%. These results indicate that although high *MAX* returns underperform on average, their lottery-like characteristics (high skewness, high idiosyncratic volatility and high maximum returns) are persistent.

3.3   Cross-sectional Fama-MacBeth regressions

Portfolio sorts are easy to interpret and they do not impose a functional form on the relationship between *MAX* and future returns. However, a large amount of cross-sectional information is lost

exactly because of the process of portfolio formation. Also, as the dependencies between different characteristics become increasingly complex, sorts will fail to provide a clear picture of what is really going on. Given the large number of characteristics that seem to determine the cross-section of returns, we follow Fama and French (2008) and perform Fama and MacBeth (1973) regressions to disentangle the effects. Each month t we regress the cross-section of excess stock returns on k explanatory variables X_j ($j = 1,2,\dots,k$):

$$R_i - R_f = b_0 + \sum_{j=1}^{k} b_j X_{i,j} + \varepsilon_i.$$

All explanatory variables are constructed so that investors know them prior to month t. We then focus on the time-series average of the estimates for b_j ($j = 1,2,\dots,k$) to evaluate which characteristics are significantly related to the cross-section of stock returns. Statistical significance is based on Newey-West (1987) adjusted t-statistics.

Table 4 reports the results of these regressions using *MAX(1)*.[6] We find that *MAX* on its own (model 1) is negatively related to the cross-section of future stock returns with an average slope coefficient of -0.0365 (t = -1.72). If we add the control variables (model 2), the *MAX* effect is stronger and significant with a coefficient of -0.0591 (t = -3.63). The other slopes are consistent with the literature: a positive but very small slope for beta (t = 0.20), a significantly negative size-effect (t = -2.06), a strong positive value effect (t = 6.63), a strong positive momentum effect (t = 4.32) and a significant negative short term reversal effect (t = -3.01). The coefficient of *Illiq* is weakly significant (t = -1.72) but with a sign that contradicts our expectations: more liquid stocks have higher returns. Model 3 shows that dropping *Illiq* from the model does not significantly change the other coefficients much.

A first robustness check investigates the average slopes for two sub-periods. We split the sample at the end of 1996. The results are shown in models 4 and 5. The *MAX* effect is negative in both sub-periods although it is only significant at the 10% level (t = -1.88) in the first. It the second sub-period the effect is stronger and more significant with an average slope of -0.0728 (t = -3.51). Also note that the size effect disappeared after the late 1990s. As a second robustness check, we add dummies for each country to the regression (model 6), with France as the reference category. This is to see whether the *MAX* effect is not just a proxy for a certain country effect. We find that the *MAX* effect remains of the same order and significance with a coefficient of -0.0683 (t = -5.06). Lastly, the coefficient for *MAX(N)* in model 2 is negative and strongly significant when *N* is set to 2, 3, 4 or 5 (results not shown). Thus far we have found that there is a significant *MAX* effect and that it is robust to controls for intermediate term momentum, short term reversal, and other control variables. Moreover, the *MAX* effect is economically important. For instance, moving from the 5th percentile to the 95th percentile of the *MAX(1)* distribution (where percentiles are averaged over all months) implies a reduction in average return of 0.73% per month according to model 2, keeping all other variables constant.

4   Robustness

It is possible that another explanation accounts for the *MAX* effect. In this section, we discuss other potential explanations and conduct a battery of robustness checks.

---

[6] Returns and variables based on returns (e.g. *Mom*) are multiplied by 100 to express them as percentages. For example, the *MAX(1)* coefficient in model 1 is -0.0365. The interpretation is that an increase of 1 percentage point in *MAX(1)* results in a -0.0365 percentage point decrease in excess returns, *ceteris paribus*.

4.1 Controlling for the MIN effect

The *MAX* variable is highly correlated with the *MIN* variable. Stocks that have higher *MAX* in a month also tend to have higher *MIN*. If we correlate[7] *MIN(N)* with *MAX(N)* for *N* equal to 1, 2, 3, 4 and 5, we find that all correlations are above 0.55 with *p*-values very close to zero. Therefore, it could be possible that *MAX* is actually capturing a *MIN* effect. We investigate this by calculating table 1 using *MIN* instead of *MAX* (results not reported for brevity). We find little proof of a *MIN* effect. For the univariate sorts sorting on *MIN(N)* for *N* equal to 1, 2, 3, 4 and 5, we find no significant hedge returns or alphas. The bivariate sorts generally tell the same story. Hedge portfolio returns and their four-factor alphas are usually small and insignificant. For *LagRet*, we find an insignificant average hedge return of -0.46% (t = -1.73), but an alpha of -0.36% (t = -2.00). For VW returns, we never find significant hedge returns and only one significant alpha for sorts on *LagRet* (t = -2.15). These sorts show little evidence for a *MIN* effect. This is corroborated by the regression analysis. Model 7 in table 4 adds *MIN(1)* to regression model 2. Again, the coefficient of *MAX* remains negative and significant: -0.0545 (t = -3.29), while the average slope for *MIN* is negative but insignificant (t = -0.55). In any case, we find that the *MAX* effect is robust to controls for *MIN*.

4.2 Controlling for idiosyncratic volatility

Table 1 shows that high *MAX* stocks have higher *MIN* and higher idiosyncratic volatility (*IVol*). Stocks with high idiosyncratic volatility should, by definition, exhibit higher minimum and maximum returns. The average cross-sectional correlation coefficient between *MAX* and *IVol* is very high, it is never lower than 0.80. The same holds for *MIN* and *IVol*. Theory suggests that idiosyncratic volatility should either not be priced or be positively priced. However, theory cannot explain the observed idiosyncratic volatility puzzle of Ang et al. (2006) and Ang et al. (2009), where it seems that idiosyncratic volatility is negatively priced. Given the high correlations, a real concern is that *MAX* is a proxy for *IVol*, leading to a negative *MAX* effect. Bivariate sorts are not a good methodology here, since the high correlation between the two variables will result in almost no variation in *MAX* within each decile of *IVol*. Sorting independently does not solve this problem, since the outer decile portfolios will contain very little or no stocks. Regression is a better option, but high correlation between explanatory variables might cause multicollinearity problems. However, this problem is largely mitigated by our methodology. Multicollinearity blows up standard errors of coefficients but does not bias the coefficients themselves. Since we test our hypotheses using the time-series of these coefficients, there is no bias. However, it may be harder to reject to null for a certain Fama-MacBeth estimate because of its higher variation over time. To sum up, although estimates remain unbiased, it might be harder to reject the null. This means that if a parameter estimate is insignificant, it could be because the true parameter is (close to) zero, or because of multicollinearity.

Models 8 to 10 examine idiosyncratic volatility more closely. In model 8, we add *IVol* as an explanatory variable but leave *MIN* and *MAX* out. We can confirm the idiosyncratic volatility puzzle, since the average slope for *IVol* is -0.1182 (t = -1.93). But if we add *MAX* to that model (model 9), we find that the puzzle disappears and the coefficient of *IVol* actually becomes positive, albeit insignificant. The average slope is 0.0221 (t = 0.30). Note that the coefficient of *MAX* remains stable: -0.0637 (t = -2.61). Adding *MIN* to this model (model 10) does not change much: the average slope for *MAX* is -0.0710 (t = -2.71). The coefficient of *IVol* increases quite a lot to 0.0890 (t = 0.92). This insignificance could be due to multicollinearity problems. The average slope of *MIN* is -0.0285 (t = -1.15). These results show that the idiosyncratic volatility puzzle disappears after controlling for *MAX* and that the *MAX* effect remains after controlling for *IVol*.

---

[7] Each month the cross-sectional correlation between two variables is computed. These coefficients are then averaged over time to get the average cross-sectional correlation.

4.3 Controlling for skewness

Stocks with higher *MAX* have higher skewness, as can be seen in panel B of table 1. Several asset pricing models that include a preference for skewness show that stocks with higher coskewness should have lower expected returns, see e.g. Kraus and Litzenberger (1976) or Harvey and Siddique (2000). The average cross-sectional correlation between *MAX(1)* and *Skew* is 0.19 ($p < 0.001$) and it only drops for higher *N*. Despite the low correlation, *MAX* could still be a proxy for skewness. If investors have a preference for positive skewness, but we do not include skewness in our model, then *MAX* might be capturing the negative effect associated with skewness. Model 11 in table 4 shows what happens when we add *Skew* to model 2 but leave *MAX* out. We find that skewness is indeed negatively priced with an average slope of -0.0524 (t = -3.44). In model 12, we add *MAX(1)* to the model and find that little has changed. *MAX* and *Skew* are still negatively priced with an average slope of respectively -0.0569 (t = -3.47) and -0.0618 (t = -2.89). In model 13 we combine all variables and we find that the *MAX* effect is robust to all controls with an average slope of -0.0665 (t = -2.54). *MIN* and *IVol* are not priced but *Skew* is negatively priced with an average slope of -0.0686 (t = -3.35). Finally, we also test some other measures for skewness, namely systematic skewness and idiosyncratic skewness. To calculate systematic and idiosyncratic skewness, we follow Harvey and Siddique (2000) and estimate the following model for each stock in each month, based on 65 to 260 prior daily returns:

$$R_{i,d} - R_{f,d} = \alpha_i + \beta_i\left(R_{m,d} - R_{f,d}\right) + \gamma_i\left(R_{m,d} - R_{f,d}\right)^2 + \varepsilon_{i,d}.$$

Coskewness is defined as the estimate of γ_i in the above model, while idiosyncratic skewness is defined as the skewness coefficient of the residuals $\varepsilon_{i,d}$ in the above model. Results from these regressions are not reported separately. If we re-estimate model 13 but with coskewness instead of *Skew*, we find an average slope of -1.0781 (t = -2.62) for coskewness and of -0.0688 (t = -2.64) for *MAX*. If we use idiosyncratic skewness instead of *Skew*, we find an average slope of -0.0644 (t = -3.35) for idiosyncratic skewness and of -0.0666 (t = -2.55) for *MAX*. We conclude that the *MAX* effect is robust to a wide range of controls, including *MIN*, skewness and idiosyncratic volatility.

5   Conclusions

Using an extensive sample of euro area stocks we are able to corroborate the results of Bali et al. (2011) that stocks with extreme positive returns have systematically lower average returns. The result holds out after wide-ranging robustness checks. Although extreme positive return stocks also display low momentum and have high lagged monthly returns, these characteristics do not explain their underperformance. This finding is consistent with individual investors preferring stocks with lottery characteristics that imply a low probability of earning a very high return. This preference would increase demand for these stocks, thereby lowering their expected returns. The fact that extreme positive return stocks are generally small, relatively illiquid high *B/M* stocks with high idiosyncratic return volatility is consistent with this interpretation. These are precisely the stocks where the influence of individual investors on pricing is arguably larger. On the other hand, we find little evidence of a reverse effect for extreme negative return stocks. We conjecture that it is easier to trade against the extreme negative stock effect as this involves only long positions. Trading against the extreme positive stock effect implies taking short positions which is much more difficult for the typical stock in this category.

We also find that the extreme positive stock effect is not due to idiosyncratic volatility or skewness, which are both correlated to our measures of extreme positive returns. However, these robustness checks also shed light on two related issues that are interesting in their own respect. First, since Ang et al. (2006) reported a negative relation between idiosyncratic volatility and average returns this puzzle

has been central to empirical asset pricing. Initially we find corroborating evidence for this negative relation using Fama and MacBeth (1973) regressions. However, when we combine idiosyncratic volatility and a measure of extreme positive return in the regression, the relation between return and idiosyncratic volatility turns positive, although not significantly. Second, we find univocal evidence that skewness is negatively priced in the cross-section. This is consistent with a preference for skewness and theoretical three-moment asset pricing models such as Kraus and Litzenberger (1976). However, neither idiosyncratic volatility nor the different skewness measures weaken the extreme positive return effect.

Tables

Table 1 – Univariate *MAX* sorts

In panel A, each month *t* we sort stocks based on *MAX(N)*. We then form 10 portfolios so that each portfolio contains an equal number of stocks. Subsequent portfolio returns are either equally-weighted (EW) or value-weighted (VW). Each month, the cross-section of returns is winsorised at the 0.5% and 99.5% levels. The table reports the time-series averages of EW and VW portfolio returns, along with Newey-West (1987) adjusted *t*-statistics for the hedge portfolio, which is long in the high portfolio and short in the low portfolio. Four factor alphas of the hedge portfolio and their Newey-West (1987) adjusted *t*-statistics are reported as *Hedge 4F alpha*. In panel B, each month *t* we compute the median of the cross-section of firm characteristics for each decile portfolios. The time-series averages of these medians are reported. *MAX(N)* is calculated by taking the average of the *N* highest returns in the estimation window, which contains the 20 daily returns prior to month *t*. *MIN(N)* is calculated by taking the average of the absolute values of the *N* lowest returns in the estimation window. *Beta* is the market beta from the three factor model estimated using 24 to 60 monthly returns prior to month *t*. *MC* and *B/M* are the market capitalization and book-to-market ratio obtained from TDS. *Mom* is the buy-and-hold return from month *t*-12 to month *t*-2 and the *LagRet* is the return of month *t*-1. *IVol* is the RMSE of the three factor model estimated using the 20 days prior to month *t*. *Illiq* is the proportion of zero returns observed in the 260 days prior to month *t*. *Skew* is the skewness of the 260 daily returns prior to month *t*. The sample period is 1979/12 – 2011/06.

Panel A: Average portfolio returns for univariate sorts on MAX(N)

	Average EW portfolio returns					Average VW portfolio returns				
Portfolio	MAX(1)	MAX(2)	MAX(3)	MAX(4)	MAX(5)	MAX(1)	MAX(2)	MAX(3)	MAX(4)	MAX(5)
Low	0.69%	0.66%	0.66%	0.63%	0.64%	0.62%	0.63%	0.63%	0.63%	0.65%
2	0.92%	0.96%	0.95%	0.99%	0.97%	0.74%	0.73%	0.72%	0.72%	0.67%
3	1.16%	1.16%	1.19%	1.18%	1.16%	1.02%	1.02%	0.97%	0.94%	0.95%
4	1.32%	1.29%	1.28%	1.24%	1.27%	1.18%	1.16%	1.20%	1.16%	1.15%
5	1.14%	1.18%	1.15%	1.19%	1.19%	1.00%	1.00%	0.97%	1.07%	1.14%
6	1.09%	1.10%	1.14%	1.12%	1.13%	1.05%	1.12%	1.16%	1.07%	1.02%
7	1.07%	1.12%	1.12%	1.14%	1.15%	0.99%	1.09%	1.09%	1.01%	0.95%
8	1.35%	1.06%	1.00%	0.99%	1.01%	0.89%	0.90%	0.91%	0.94%	0.99%
9	1.11%	1.23%	1.07%	1.03%	0.98%	0.92%	0.69%	0.66%	0.66%	0.67%
High	0.22%	0.32%	0.52%	0.55%	0.57%	0.32%	0.36%	0.41%	0.47%	0.45%
Hedge	-0.47%	-0.35%	-0.15%	-0.08%	-0.08%	-0.30%	-0.27%	-0.23%	-0.16%	-0.21%
Hedge *t*	-1.20	-0.82	-0.30	-0.16	-0.15	-0.70	-0.62	-0.48	-0.35	-0.44
Hedge 4F alpha	-0.55%	-0.47%	-0.30%	-0.26%	-0.26%	-0.48%	-0.46%	-0.45%	-0.40%	-0.48%
Hedge 4F alpha *t*	-2.06	-1.68	-0.92	-0.76	-0.75	-1.34	-1.13	-1.06	-0.94	-1.16

Panel B: Averages of cross-sectional median of characteristics of decile portfolios sorted on MAX(1)

Portfolio	MAX(1)	MIN(1)	Beta	MC (€10⁶)	B/M	Mom	LagRet	IVol	Illiq	Skew	Price (€)
Low	0.80%	1.04%	0.40	140.7	0.62	4.63%	-0.57%	0.44%	7.69%	0.45	56.27
2	1.49%	1.82%	0.63	201.3	0.65	8.28%	-0.96%	0.78%	5.93%	0.30	47.11
3	2.17%	2.41%	0.80	202.9	0.65	9.98%	-0.80%	1.06%	5.08%	0.28	41.79
4	2.85%	2.93%	0.94	193.5	0.65	10.96%	-0.45%	1.33%	4.86%	0.28	41.26
5	3.50%	3.33%	1.01	173.1	0.66	10.20%	0.01%	1.55%	4.92%	0.29	38.65
6	4.20%	3.69%	1.06	144.6	0.67	9.34%	0.67%	1.78%	5.15%	0.32	35.94
7	5.05%	4.13%	1.10	113.5	0.69	10.21%	1.26%	2.05%	5.51%	0.33	32.44
8	6.15%	4.69%	1.14	89.3	0.70	9.94%	2.48%	2.39%	5.86%	0.38	30.02
9	7.97%	5.40%	1.22	69.0	0.72	9.39%	4.14%	2.94%	6.65%	0.47	23.26
High	13.08%	7.34%	1.20	38.6	0.78	-3.26%	8.05%	4.57%	8.94%	0.89	16.13

Table 2 – Bivariate *MAX* sorts

Each month *t* we first sort stocks based on a control variable: market beta, size (*MC*), book-to-market ratio (*B/M*), momentum (*Mom*) and lagged return (*LagRet*). We form 10 portfolios. Then we sort all stocks within each portfolio on *MAX(1)*. We form 10 portfolios within each control group so that the result is 100 (10x10) portfolios, each with roughly an equal number of stocks. However, for brevity, we only report portfolio returns that are averaged across the spectrum of the control variable to produce decile portfolios with variation in *MAX* but with similar levels of the control variable. Each month, the cross-section of returns is winsorised at the 0.5% and 99.5% levels. The table reports the time-series averages of EW and VW portfolio returns, along with Newey-West (1987) adjusted *t*-statistics for the hedge portfolio, which is long in the high portfolio and short in the low portfolio. Four factor alphas of the hedge portfolio and their Newey-West (1987) adjusted *t*-statistics are reported as *Hedge 4F alpha*. *MAX(1)* is the highest daily return over the 20 daily returns prior to month *t*. Beta is the market beta from the three factor model estimated using 24 to 60 monthly returns prior to month *t*. *MC* and *B/M* are the market capitalization and book-to-market ratio obtained from TDS. *Mom* is the buy-and-hold return from month *t*-12 to month *t*-2 and the *LagRet* is the return of month *t*-1. The sample period is 1979/12 – 2011/06.

Panel A: bivariate sorts on control variables and MAX(1)

Portfolio	Average EW portfolio returns					Average VW portfolio returns				
	Beta	*Size*	*B/M*	*Mom*	*LagRet*	*Beta*	*Size*	*B/M*	*Mom*	*LagRet*
Low	0.88%	0.76%	0.95%	0.91%	1.09%	0.66%	0.78%	0.73%	0.67%	0.86%
2	1.11%	0.91%	1.12%	1.13%	1.14%	1.00%	0.90%	1.05%	0.96%	0.89%
3	1.17%	1.14%	1.33%	1.18%	1.11%	1.08%	1.13%	1.26%	0.95%	1.05%
4	1.19%	1.21%	1.28%	1.14%	1.27%	0.99%	1.18%	1.28%	1.01%	1.23%
5	1.09%	1.17%	1.14%	1.13%	1.14%	1.02%	1.13%	1.16%	0.96%	0.92%
6	1.13%	1.13%	1.04%	1.11%	1.14%	1.02%	1.10%	1.11%	0.77%	1.06%
7	1.13%	1.15%	1.16%	1.03%	1.10%	0.98%	1.14%	1.15%	0.78%	0.97%
8	1.19%	1.35%	1.22%	1.05%	0.95%	1.10%	1.30%	1.14%	0.68%	0.95%
9	1.26%	0.95%	1.14%	0.91%	0.82%	1.13%	0.90%	1.06%	0.80%	0.78%
High	0.36%	0.29%	0.20%	0.44%	0.28%	0.20%	0.22%	0.47%	0.20%	0.33%
Hedge	-0.52%	-0.47%	-0.74%	-0.47%	-0.81%	-0.47%	-0.56%	-0.26%	-0.46%	-0.53%
Hedge *t*	-1.96	-1.28	-2.54	-1.70	-2.74	-1.63	-1.53	-0.83	-1.78	-1.66
Hedge 4F alpha	-0.79%	-0.60%	-0.75%	-0.86%	-0.71%	-0.64%	-0.69%	-0.33%	-0.82%	-0.45%
Hedge 4F alpha *t*	-4.42	-2.41	-3.16	-4.48	-3.41	-2.68	-2.79	-1.17	-3.16	-1.77

Panel B: summary of hedge portfolios for bivariate sorts on control variables and MAX(5)

Hedge	-0.24%	-0.19%	-0.47%	-0.40%	-0.75%	-0.33%	-0.29%	-0.30%	-0.45%	-0.57%
Hedge *t*	-0.58	-0.42	-1.18	-1.28	-2.26	-1.01	-0.66	-0.83	-1.60	-1.67
Hedge 4F alpha	-0.59%	-0.37%	-0.60%	-0.84%	-0.73%	-0.59%	-0.48%	-0.46%	-0.89%	-0.58%
Hedge 4F alpha *t*	-1.88	-1.22	-1.91	-3.94	-3.42	-2.08	-1.55	-1.40	-3.17	-1.98

Table 3 – Transition matrix for *MAX(1)*

Each month *t* we sort stocks based on *MAX(1)*. We then form 10 portfolios so that each portfolio contains an equal number of stocks. In month *t*+1, we do the same. Then we check to which decile the stock has moved. The table shows average transition probabilities, giving us the probability that a stock in decile *i* (table rows) will be in decile *j* (table columns) the next month. *MAX(1)* is the highest daily return over the 20 daily returns prior to month *t*. The sample period is 1979/12 – 2011/06.

	Low	2	3	4	5	6	7	8	9	High
Low	42.4%	17.2%	7.8%	5.3%	4.4%	3.5%	3.1%	2.7%	2.4%	3.7%
2	17.9%	21.5%	13.7%	9.2%	7.6%	6.6%	5.4%	4.6%	3.9%	4.1%
3	8.6%	14.1%	17.0%	12.6%	10.4%	8.8%	7.8%	6.4%	5.4%	4.6%
4	5.4%	9.8%	13.1%	13.9%	12.5%	11.0%	9.8%	8.5%	7.1%	5.3%
5	4.4%	8.0%	10.9%	12.8%	13.2%	12.4%	10.9%	9.6%	8.1%	5.8%
6	3.6%	6.7%	9.3%	11.6%	12.6%	12.8%	12.1%	10.8%	9.6%	6.8%
7	2.9%	5.3%	8.1%	10.3%	11.2%	12.3%	13.5%	12.5%	10.7%	8.5%
8	2.5%	4.5%	6.6%	8.8%	10.1%	11.4%	12.9%	14.1%	13.7%	10.2%
9	2.3%	3.8%	5.5%	7.0%	8.5%	9.9%	11.3%	14.3%	17.1%	14.0%
High	3.4%	3.6%	4.1%	5.1%	5.9%	7.1%	8.6%	11.3%	15.5%	26.1%

Table 4 – Cross-sectional regressions of firm returns on firm characteristics

Each month t we regress the cross-section of excess stock returns onto several explanatory variables. This procedure yields a time-series of slope coefficients. The table reports the time-series averages of these coefficients along with their Newey-West (1987) adjusted t-statistics in parentheses. The control variables are market beta, the log of market capitalization (*logMC*), the log of the book-to-market ratio (*logB/M*), momentum (*Mom*), lagged return (*LagRet*) and illiquidity (*Illiq*). Each month, all variables are winsorised at the 0.5% and 99.5% levels. *MAX(1)* is the highest daily return over the 20 daily returns prior to month t. *MIN(1)* is the absolute value of the lowest daily return over the 20 daily returns prior to month t. *Beta* is the market beta from the three factor model estimated using 24 to 60 monthly returns prior to month t. *MC* and *B/M* are the market capitalization and book-to-market ratio obtained from TDS. *Mom* is the buy-and-hold return from month t-12 to month t-2 and the *LagRet* is the return of month t-1. *Illiq* is the proportion of zero returns observed in the 260 days prior to month t. *Skew* is the skewness coefficient of the 65 to 260 daily returns prior to month t. *IVol* is the RMSE of the three factor model estimated using the 20 days prior to month t. Excess returns and variables based on returns (*MAX*, *MIN*, *Mom*, *LagRet* and *IVol*) are multiplied by 100 to express them as percentages. The sample period is 1979/12 – 2011/06.

Model/sample	MAX(1)	Beta	LogMC	LogB/M	Mom	LagRet	Illiq	MIN(1)	IVol	Skew
Panel A: univariate										
Model 1	-0.0365									
	(-1.72)									
Panel B: multivariate, including sub-periods and country dummies										
Model 2	-0.0591	0.0265	-0.0764	0.4437	0.0106	-0.0197	-1.6653			
	(-3.63)	(0.20)	(-2.06)	(6.63)	(4.32)	(-3.01)	(-1.72)			
Model 3	-0.0596	0.0269	-0.0732	0.4451	0.0107	-0.0191				
	(-3.68)	(0.20)	(-2.28)	(6.62)	(4.29)	(-2.87)				
Model 4	-0.0459	0.1593	-0.1164	0.4026	0.0118	-0.0249	-1.7861			
1982 – 1996	(-1.88)	(0.76)	(-2.51)	(5.08)	(4.71)	(-2.83)	(-1.89)			
Model 5	-0.0728	-0.1117	-0.0347	0.4865	0.0094	-0.0142	-1.5396			
1997 – 2011	(-3.51)	(-0.71)	(-0.60)	(4.47)	(2.19)	(-1.51)	(-0.90)			
Model 6	-0.0683	0.0104	-0.0357	0.3721	0.0091	-0.0323	-0.7365			
Country dum.	(-5.06)	0.09)	(-1.26)	(6.25)	(4.17)	(-6.28)	(-1.25)			
Panel C: including minimum returns										
Model 7	-0.0545	0.0244	-0.0790	0.4453	0.0107	-0.0199	-1.6981	-0.0113		
	(-3.29)	(0.18)	(-2.20)	(6.70)	(4.43)	(-3.00)	(-1.77)	(-0.55)		
Panel D: including idiosyncratic volatility										
Model 8		0.0072	-0.0789	0.4517	0.0108	-0.0237	-1.7298		-0.1182	
		(0.05)	(-2.16)	(6.68)	(4.50)	(-3.62)	(-1.77)		(-1.93)	
Model 9	-0.0637	0.0135	-0.0769	0.4466	0.0107	-0.0188	-1.6386		0.0221	
	(-2.61)	(0.10)	(-2.16)	(6.64)	(4.37)	(-2.94)	(-1.70)		(0.30)	
Model 10	-0.0710	0.0132	-0.0771	0.4457	0.0107	-0.0202	-1.6452	-0.0285	0.0890	
	(-2.71)	(0.10)	(-2.17)	(6.71)	(4.48)	(-3.09)	(-1.74)	(-1.15)	(0.92)	
Panel E: including skewness										
Model 11		-0.0101	-0.0442	0.4394	0.0104	-0.0428	-0.2061			-0.0524
		(-0.08)	(-1.27)	(7.05)	(4.29)	(-6.73)	(-0.59)			(-3.44)
Model 12	-0.0569	0.0257	-0.0805	0.4469	0.0112	-0.0192	-1.5752			-0.0618
	(-3.47)	(0.19)	(-2.18)	(6.66)	(4.54)	(-2.95)	(-1.69)			(-2.89)
Model 13	-0.0665	0.0165	-0.0823	0.4486	0.0114	-0.0206	-1.5444	-0.0356	0.0924	-0.0686
	(-2.54)	(0.13)	(-2.32)	(6.74)	(4.70)	(-3.17)	(-1.69)	(-1.45)	(0.96)	(-3.35)

Appendix A: filtering procedure

A first filter makes sure that a stock is the "major" stock for a company, as TDS defines it, to eliminate cross-listings. Next, the currency has to be euro. A third filter checks the industry-type for each company. We do not want to include investments such as REITs, investment trusts, venture capital trusts, asset management funds, or other funds. Also, the expanded company name cannot contain "suspicious" words such as *pref, prf, %, duplicate, dupl, afv, vvpr* or *strip*. Next, we only allow the type of equity to be EQ (for equity) or CF (for certificate). Another filter makes sure that the asset's home country matched the chosen list's country to avoid TDS errors. Also, the ISIN code has to start with the correct two letters (e.g. BE for Belgium or FI for Finland), or the ISIN code has to be missing. Lastly, we remove doubles from the list. These occur often since multiple lists per country are used. Doubles are removed based on their *type*, which is a unique identifier code in TDS.

Appendix B: construction of market, SMB, HML and WML factors

The market factor in month t is a value-weighted average constructed from the individual excess stock returns from month t. The *SMB* and *HML* factors are constructed using six value-weighted portfolios formed on *MC* and *B/M*. The portfolios. which are constructed at the end of each June and then held for a year, are the (independent) intersections of two portfolios formed on *MC* and three portfolios formed on *B/M*. The *MC* breakpoint for year t is the 80th percentile at the end of June of year t. Fama and French (1993) actually use the median of NYSE market capitalizations, but Schmidt et al. (2011) have shown for their European dataset (which is very similar to ours) that the 80th quantile is very similar to the size breakpoints used by Fama and French (1993). The *B/M* breakpoints are the 30th and 70th percentiles. The *SMB* factor is then the average return on the three low *MC* portfolios minus the average return on the three high *MC* portfolios. The *HML* factor is the average return on the two high *B/M* portfolios minus the average return on the two low *B/M* portfolios (the middle *B/M* group is not considered). The *WML* factor is constructed in a slightly different way. Six value-weighted portfolios are formed on *MC* and the prior t-12 to t-2 buy-and-hold return (*Mom*). The portfolios are constructed monthly and are the (independent) intersections of two portfolios formed on *MC* and three portfolios formed on momentum. The monthly size breakpoint is again the 80th percentile. The momentum breakpoints are the 30th and 70th percentiles. The *WML* factor is then the average return on the two high momentum portfolios minus the average return on the two low momentum portfolios. Descriptive statistics and correlations are provided in tables B1 and B2.

Table B1 – Descriptive statistics for the four factors

Each month, the four factors are calculated as described in appendix B. The following table provides summary statistics for the four factors.

	Market (excess)	SMB	HML	WML
Average return	0.56%	0.02%	0.76%	0.94%
t-statistic	1.95	0.14	4.38	3.49
Minimum	-21.26%	-11.43%	-7.88%	-29.09%
25th percentile	-1.88%	-1.41%	-0.78%	-0.54%
Median	1.11%	0.17%	0.74%	1.35%
75th percentile	3.40%	1.72%	2.21%	3.39%
Maximum	14.35%	7.23%	12.63%	15.37%
Skewness	-0.69	-0.36	0.57	-1.41
Excess kurtosis	2.30	1.27	2.60	7.83

Table B2 – Correlation matrix for the four factors

Each month, the four factors are calculated as described in appendix B. The following table provides a correlation matrix for the four factors. Numbers above the diagonal are Pearson product moment correlation coefficients, while numbers below the diagonal are Spearman rank correlation coefficients.

	Market (excess)	SMB	HML	WML
Market (excess)	1.00	-0.43	0.16	-0.28
SMB	-0.40	1.00	-0.04	0.18
HML	0.16	0.00	1.00	-0.37
WML	-0.10	0.07	-0.32	1.00

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets, 5*(1), 31-56.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance, 61*(1), 259-299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics, 91*(1), 1-23.

Aretz, K., Bartram, S. M., & Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking & Finance, 34*(6), 1383-1399.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics, 99*(2), 427-446.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and Expected Returns: Lessons from Emerging Markets. *The Review of Financial Studies, 20*(6), 1783-1831.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

Dimson, E. (1979). Risk Measurement When Shares Are Subject to Infrequent Trading. *Journal of Financial Economics, 7*(2), 197-226.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (2008). Dissecting Anomalies. *The Journal of Finance, 63*(4), 1653-1678.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), 607-636.

Goetzmann, W. N., & Kumar, A. (2008). Equity Portfolio Diversification. *Review of Finance, 12*(3), 433-463.

Guo, H., Savickas, R., Wang, Z., & Yang, J. (2009). Is the Value Premium a Proxy for Time-Varying Investment Opportunities? Some Time-Series Evidence. *Journal of Financial and Quantitative Analysis, 44*(01), 133-154.

Harvey, C. R., & Siddique, A. (2000). Conditional Skewness in Asset Pricing Tests. *The Journal of Finance, 55*(3), 1263-1295.

Ince, O. S., & Porter, R. B. (2006). Individual Equity Return Data from Thomson Datastream: Handle with Care! *Journal of Financial Research, 29*(4), 463-479.

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *The Journal of Finance, 45*(3), 881-898.

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance, 48*(1), 65-91.

Kraus, A., & Litzenberger, R. H. (1976). Skewness Preference and the Valuation of Risk Assets. *The Journal of Finance, 31*(4), 1085-1100.

Kumar, A. (2009). Who gambles in the stock market? *Journal of Finance, 64*(4), 1889-1933.

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies, 3*(3), 431-467.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), 703-708.

Odean, T. (1999). Do investors trade too much? *American Economic Review, 89*(5), 1279-1298.

Schmidt, P. S., von Arx, U., Schrimpf, A., Wagner, A. F., & Ziegler, A. (2011). *On the Construction of Common Size, Value and Momentum Factors in International Stock Markets: A Guide with Applications*. Center of Economic Research at ETH Zurich Working Paper.