Idiosyncratic Volatility and Nominal Stock Prices: Evidence from Approximate Factor Structures

Patrick Roger [1], Tristan Roger [2], Alain Schatt [3]

[1] LaRGE Research Center, EM Strasbourg Business School, University of Strasbourg
[2] DRM Finance, Université Paris-Dauphine, PSL Research University
[3] HEC Lausanne, University of Lausanne.

Corresponding author: Patrick Roger, proger@unistra.fr
Date of submission: 01/12/2016
Date of publication: 15/03/2017

Abstract
Approximate factor structures defined by Chamberlain and Rotschild (1983) allow to test whether a given quantitative firm characteristic (the nominal stock price in this paper) is a determinant of the idiosyncratic volatility of stock returns. Our study of 8,000 U.S stocks over the period 1980-2014 shows that small price stocks exhibit a higher idiosyncratic volatility than large price stocks. This relationship is persistent over time and robust to variations in the number of common factors of the approximate factor structure. Moreover, this small price effect does not hide a small-firm effect because it is still valid when we analyze the tercile of large firms. Our result confirms that small price stocks have lottery-type characteristics and, therefore, it is not in line with the efficient market hypothesis.

1. Introduction

In efficient markets, nominal prices of stocks play no role. Only market values and returns matter. In a recent paper, however, Birru and Wang (2016) show that small price stocks exhibit a peculiar behavior, when compared to large price stocks. In particular, the authors conclude that investors overestimate the room to grow for small price stocks, leading them to overestimate the skewness of future returns. Moreover, Roger et al. (2016) find that financial analysts are much more optimistic when they issue one-year ahead target prices on small price stocks, compared to their price forecasts on large price stocks. In this paper, we show that a small stock price is also associated with a high idiosyncratic volatility of the stock return.

Whichever asset pricing model used, idiosyncratic variance is the difference between total variance and variance coming from the common factors. As a result, the measure of idiosyncratic variance crucially depends on the number and the choice of common factors in the asset pricing model. In a recent paper, Harvey et al. (2015) mention that more than 300 significant factors have been uncovered in hundreds of papers, but they highlight that most claimed research findings in financial economics are likely false. Hence, we do not really know what the common factors are, what is their number, and what are their economic contents.

In this paper, we refer to the general approximate factor structures introduced by Chamberlain and Rotschild (1983) to obtain an estimate of the idiosyncratic volatility (henceforth IVOL) of stock returns. The fact that approximate factor structures take into account correlated residuals is a significant advantage of this approach. We then show that IVOL is strongly linked to the nominal price of stocks. For example, Campbell et al. (2001) use two types of common factors: the usual market factor, measured by the excess return on a broad market index like the S&P500, and industry factors which influence only subsets of stock returns. Considering only the market factor implies that residuals are correlated because of industry factors.

Our empirical study deals with approximately 8,000 stocks traded on the U.S market over a 35-year period, starting in January 1980 and ending in December 2014. When performing principal component analyses (PCA) of daily returns on the 35 non-overlapping periods of one year, we find that the average IVOL (expressed as a percentage of the average total variance) of large price stocks is lower than the corresponding IVOL of small price stocks. The result holds when we consider one factor. The reason to primarily focus our attention on the first eigenvalue comes from the numerous studies showing that a one-factor model explains an important part of the variance of returns (Trzcinka, 1986; Connor and Korajczyk, 1993; Geweke and Zhou, 1996; Jones, 2001). Our result, however, also holds when we consider three common factors.

This result is remarkable because small price stocks are not always the same over time, because of price variations and stock splits. Indeed, Weld et al. (2009) show that nominal prices have remained approximately constant in the U.S since the Great Depression. After a detailed analysis of this phenomenon through the lens of standard theories, they come to the conclusion that firms simply follow market norms, in the sense given by Akerlof (2007) in his Presidential Address to the American Economic Association, even if maintaining the price level stable comes at a cost. In particular, bid-ask spreads increase after stock splits (Copeland, 1979; Conroy et al., 1990).

Overall, we highlight that IVOL is strongly associated with the nominal price of stocks. The percentage of variance explained by common factors is much higher for large price stocks than for small price stocks. Our result is persistent over time (our analysis covers the period 1980-2014) and does not hide a small size effect because it remains valid on the tercile of large caps. Since the standard theory of finance considers that stock prices are irrelevant to describe returns, we therefore argue that the nominal stock price is a supposedly irrelevant factor (SIF), as defined by Thaler (2015). Hence, our result confirms that most small price stocks have lottery-type characteristics, which is not consistent with the efficient market hypothesis.

The paper is structured as follows. Section 2 recalls the basic results related to approximate factor structures (Chamberlain and Rotschild, 1983). Section 3 presents the data and descriptive statistics. Section 4 develops the empirical results and section 5 concludes.

2. Approximate factor structures

Approximate factor structures are an extension of the multifactor model of returns proposed by (Ross, 1976) in his seminal paper « The Arbitrage Theory of Capital Asset Pricing » (hereafter APT). This alternative factor model was first proposed by Chamberlain and Rotschild (1983) and characterized as follows. Suppose that n financial securities are traded on the financial market. The returns on

these securities (denoted $R_i, i=1,..,n$) are square-integrable random variables defined on a probability space (Ω, A, P) . The first moments of the returns are denoted:

$$E(R_i)=m_i ; Var(R_i)=\sigma_i^2 ; cov(R_i, R_j)=\sigma_{ij} \tag{1}$$

$$V_n=(\sigma_{ij}, i=1,...,n ; j=1,...,n) \tag{2}$$

For the consistency of notations, $\sigma_{ii}=\sigma_i^2$.

Definition 1 *The market is driven by an approximate K-factor structure if*

$$R_i=E(R_i)+\sum_{k=1}^{K} \beta_{ik} F_k+\widetilde{\varepsilon}_i \tag{3}$$

for $i=1,..,n$ *where the common factors* $F_k, k=1,...,K$ *are uncorrelated zero-mean random variables with unit variance. The factors* F_k *are uncorrelated with the residuals* $\widetilde{\varepsilon}_i$

$$V_n=B_n B_n^{'}+ D_n \tag{4}$$

where the (i,k) *element of* B_n *is* β_{ik} *, and* $(D_n, n=1,...)$ *is a sequence of positive semi-definite matrices satisfying:*

$$\widetilde{\lambda}=\sup_n g_1(D_n)<+\infty \tag{5}$$

where $g_1(D_n)$ *is the largest eigenvalue of* D_n *.*

The variance of $\widetilde{\varepsilon}_i$ is denoted s_i^2 . In the original APT the $\widetilde{\varepsilon}_i$ are assumed i.i.d. This assumption is relaxed in an approximate factor structure. The matrices D_n need not be diagonal.

Equation (3) implies that the variance of returns of stock i , denoted σ_i^2 , is equal to:

$$\sigma_i^2=\sum_{k=1}^{K} \beta_{ik}^2 +s_i^2 \tag{6}$$

where $s_i^2=V(\widetilde{\varepsilon}_i)$ is the idiosyncratic variance of returns. It turns out that s_i^2 is the difference between the total variance of returns, and the systematic variance $\sum_{k=1}^{K} \beta_{ik}^2.$ As a consequence, s_i^2 crucially depends on the number of common factors.

In empirical studies, K is not known with certainty and no universal technique exists to determine $K.$ Prudence is then needed when dealing with idiosyncratic variances in such models. Summing equation (6) over all stocks and rearranging terms leads to:

$$\sum_{i=1}^{n} \sigma_i^2=\sum_{k=1}^{K}\sum_{i=1}^{n} \beta_{ik}^2+\sum_{i=1}^{n} s_i^2 \tag{7}$$

The left-hand side of equation (7) is the trace of the covariance matrix of returns V_n which is also the sum of the eigenvalues of the covariance matrix V_n.

We get a CAPM-like factor model in the case $K=k=1$. We then expect $\sum_{i=1}^{n} \beta_{i1}^2$ to be the first eigenvalue of V_n for a sufficiently large number n of stocks. The reason is simple. When a *n*-th stock is introduced, the covariance matrix V_{n-1} becomes V_n and the trace increases by $\sigma_n^2 = \beta_{n,1}^2 + s_n^2$. But $\beta_{n,1}^2$ is added to the first eigenvalue due to the one-factor assumption and s_n^2 feeds higher-order eigenvalues, including the new *n*-th eigenvalue. It turns out that the first eigenvalue increases without limits when n increases. Nothing equivalent can be expected for the higher-order eigenvalues in a one-factor model. These eigenvalues are bounded when the number of stocks tends to infinity.

The residual terms in equation (3) may be correlated in an approximate factor structure. One economic interpretation of such a structure is that common factors influence the returns of all stocks but some other factors, such as industry factors, influence only a subset of returns. In this case, the share of variance attributed to these factors is, in some sense, residual, and appears in the variables $\tilde{\varepsilon}_i = 1, ..., n$, . This results in non-zero correlations between these variables $\tilde{\varepsilon}_i$ s.

Campbell et al. (2001) use a two-step decomposition of returns to introduce industry factors (this two-step decomposition has also been used in studies focusing on the role of idiosyncratic volatility in Europe, see Kearney and Poti (2008). In the first step, the return of industry j is written:

$$R_{jt} = \beta_j R_{Mt} + \tilde{\zeta}_{jt} \tag{8}$$

where R_{jt} is the date-*t* return of industry *j* and R_{Mt} is the market return.
In the second step, the return of a given firm i belonging to industry j is decomposed as:

$$R_{ijt} = \beta_{ij} R_{jt} + \tilde{\xi}_{ijt} \tag{9}$$

Replacing R_{jt} in equation (9) by the expression in equation (8) gives:

$$R_{ijt} = \beta_{ij} \left(\beta_j R_{Mt} + \tilde{\zeta}_{jt} \right) + \tilde{\xi}_{ijt} \tag{10}$$

$$R_{ijt} = \beta_{ij} \beta_j R_{Mt} + \beta_j \tilde{\zeta}_{jt} + \tilde{\xi}_{ijt} \tag{11}$$

It turns out that the residuals $\beta_j \tilde{\zeta}_{jt} + \tilde{\xi}_{ijt}$ are correlated when two firms belong to the same industry.

This example shows that approximate factor structures allowing correlated residuals are an efficient tool to test the assumption of a negative link between nominal stock prices and idiosyncratic volatility of returns. Moreover, this approach avoids the economic characterization of the common factors. Finally, Chamberlain and Rotschild (1983) provide a useful method to determine the number of common factors, summarized in the following proposition.

Proposition 2 *If the market has an approximate* K *-factor structure, the* $K+1$ *eigenvalue of* V_n *remains bounded when* n *tends to infinity:*

$$\sup_n g_{K+1}(V_n) < +\infty \tag{12}$$

The interpretation of this proposition is intuitive. By the invariance of the trace of V_n, we also know that

$$\sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} g_i(V_n) \tag{13}$$

When a new asset (numbered n + 1) is added, its variance σ_{n+1}^2 can be related to the preceding equation as follows:

$$\sigma_{n+1}^2 + \sum_{i=1}^{n} \sigma_i^2 = g_{n+1}(V_{n+1}) + \sum_{i=1}^{n} g_i(V_{n+1}) \tag{14}$$

Equations (13) and (14) imply

$$\sigma_{n+1}^2 = g_{n+1}(V_{n+1}) + \sum_{i=1}^{n} \left(g_i(V_{n+1}) - g_i(V_n) \right) \tag{15}$$

We cannot expect that the idiosyncratic volatility of stock $n+1$ simply feeds the last eigenvalue $g_{n+1}(V_{n+1})$ because eigenvalues are usually given in decreasing order. This reasoning generalizes the example of a one-factor CAPM-like model given before. If, however, there are K common factors driving the economy, we know that the systematic risk of stock $n+1$ feeds the first K eigenvalues thanks to proposition 2. Only the idiosyncratic variance is disseminated in eigenvalues of an order greater than K and the number of these eigenvalues increases when the number of stocks increases. It explains why, when the number of stocks under consideration increases, the K first eigenvalues increase without bounds and the eigenvalues of larger orders remain bounded.

In the empirical study of the next section, we cannot know for sure the number of common factors because the number of assets under consideration is bounded. We present the results obtained with 1 and 3 factors, which are the most likely numbers of factors over the period under consideration.

3. Data and descriptive statistics

Daily data on returns, nominal prices and market capitalizations come from the Center for Research in Security Prices (CRSP). We include in our analysis all ordinary common shares (codes 10 and 11) listed on NYSE, Amex and Nasdaq over the 1980-2014 period.

3.1 Descriptive statistics

Descriptive statistics are presented in Table 1 (page 36) on a yearly basis. The eight columns of the table contain: 1) the year under consideration, 2) the number of stocks in the sample, 3) the number of trading days in the year, 4) the adjusted Frobenius norm of the correlation matrix of daily returns (see

subsection 3.2), and 5) to 8) the cross-section yearly average of the four first moments of the distribution of returns.

The number of trading days in a year varies between 248 in 2001 and 254 in 1992 (and 1996). The time-series of the number of stocks is hump-shaped. It is increasing between 1980 and 1998, from 3,853 stocks in 1980 to 6,455 in 1998, then decreasing during the second half of the period, reaching approximately 3,400 stocks in the last years. This column clearly shows the euphoria of the nineties during the rise of the Internet bubble. The number of stocks starts to decline in 1999 due to mergers and acquisitions, and also to bankruptcies after the dotcom bubble burst in 2000.

Regarding the stock returns, the largest median gains occurred in 2003, 2013, 2009 and 1991. These three years follow large market drops in the preceding years, which is in line with the results obtained by DeBondt and Thaler (1985). The largest median losses appear respectively in 2008, 1990, 1998 and 2002. Not surprisingly, the total volatility peaked in 2000, 2008 and 2009. In subsection 3.2, however, we will see that the share of idiosyncratic volatility in total volatility is very different in 2000, compared to 2008 and 2009.

As already mentioned by Albuquerque (2012), the skewness of individual stock returns is positive on average. The only exception found by the author is the second semester of 1987. As we analyze complete years of daily returns, we find a positive skewness in 1987 but it is by far the lowest level of skewness (0.13) over the entire period.

Finally, the kurtosis ranges from 4.73 (in 1994) to 10.48 (in 1987). The very high kurtosis level in 1987 has the same explanation as the skewness level, namely Black Monday during which the market dropped by more than 20% in a single day.

3.2 The dynamics of correlation

To get some synthetic information about correlations, we use the Frobenius norm of the correlation matrix of returns.

The Frobenius norm of a square matrix $\Gamma(n,n)$ is defined as

$$|\Gamma| = \frac{\left[\left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_{ij}^2\right]^{\frac{1}{2}}}{2} \tag{16}$$

In equation (16), $|\Gamma|$ is the usual Euclidean norm of a matrix Γ in a n^2 -dimensional space (apart from the normalizing factor $1/n$). When Γ is a $(n \times n)$ correlation matrix, $|\Gamma|$ varies between $1/\sqrt{n}$ and 1 . To see why it is the case, consider a one-factor model without idiosyncratic risk. In this situation, all stocks are perfectly correlated and the correlation matrix contains only ones. Denote M this matrix. On the contrary, if stocks only carry uncorrelated idiosyncratic risks, the matrix, denoted S , is diagonal. It turns out that $|M|=1$ and $|S|=1/\sqrt{n}$.

Year	Number of stocks	Number of days	Frobenius norm	Return	Volatility	Skewness	Kurtosis
1980	3,853	253	0.1372	0.20	0.41	0.61	7.72
1981	3,946	253	0.1105	0.01	0.37	0.62	7.62
1982	4,048	253	0.1140	0.20	0.41	0.69	8.07
1983	4,481	253	0.0916	0.26	0.40	0.76	8.09
1984	4,849	253	0.0905	-0.12	0.37	0.46	7.99
1985	4,870	252	0.0765	0.22	0.39	0.49	6.52
1986	4,921	253	0.0887	0.04	0.43	0.40	6.18
1987	5,392	253	0.1964	-0.14	0.58	0.13	10.48
1988	5,535	253	0.0890	0.13	0.46	0.47	6.00
1989	5,373	252	0.0795	0.07	0.43	0.37	6.45
1990	5,228	253	0.0943	-0.25	0.55	0.28	6.18
1991	5,139	253	0.0835	0.28	0.57	0.50	5.80
1992	5,144	254	0.0748	0.13	0.58	0.41	5.34
1993	5,395	253	0.0722	0.12	0.54	0.29	4.84
1994	5,891	252	0.0781	-0.06	0.51	0.28	4.73
1995	6,074	252	0.0708	0.24	0.50	0.37	5.13
1996	6,372	254	0.0795	0.12	0.53	0.40	5.37
1997	6,692	253	0.0892	0.20	0.52	0.41	5.72
1998	6,455	252	0.1162	-0.13	0.61	0.43	6.88
1999	5,988	252	0.0747	-0.04	0.62	0.59	6.83
2000	5,760	252	0.0997	-0.15	0.75	0.53	6.57
2001	5,432	248	0.1203	0.07	0.66	0.44	6.58
2002	5,031	252	0.1541	-0.12	0.58	0.34	6.26
2003	4,540	252	0.1508	0.44	0.44	0.39	6.26
2004	4,496	252	0.1535	0.13	0.39	0.30	6.40
2005	4,430	252	0.1483	0.00	0.36	0.34	6.93
2006	4,385	251	0.1586	0.10	0.36	0.32	6.74
2007	4,201	251	0.1911	-0.11	0.40	0.25	7.09
2008	4,109	253	0.2965	-0.48	0.77	0.37	7.29
2009	3,942	252	0.2754	0.30	0.69	0.46	6.23
2010	3,748	252	0.2886	0.19	0.44	0.24	5.50
2011	3,637	252	0.3771	-0.08	0.49	0.20	6.04
2012	3,499	250	0.2131	0.13	0.37	0.26	6.42
2013	3,421	252	0.2010	0.35	0.32	0.23	6.90
2014	3,455	252	0.2104	0.03	0.33	0.21	7.10

Table 1: Descriptive statistics
The eight columns of the table contain respectively 1) the year under consideration, 2) the number of stocks in the sample, 3) the number of trading days in the year, 4) the adjusted Frobenius norm of the correlation matrix of daily returns (see subsection 3.2), 5) to 8) the cross-section average of the four first moments of the distribution of returns, expectation, standard deviation, skewness and kurtosis

To normalize the results and obtain a measure between 0 and 1, we perform the following transformation denoted as:

$$\xi(\Gamma)=\frac{\left[\dfrac{n|\Gamma|^2-1}{n-1}\right]^{1}}{2}=\frac{\left[\dfrac{1}{n(n-1)}\displaystyle\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\Gamma_{ij}^2\right]^{\frac{1}{2}}}{2} \tag{17}$$

We call $\xi(\Gamma)$ the adjusted Frobenius norm of the correlation matrix Γ. This figure is always comprised between 0 (only idiosyncratic risk) and 1 (only market risk). To give an intuitive idea of what $\xi(\Gamma)$ is, assume that all correlations between stock returns are equal to a constant c. In this case, the adjusted Frobenius norm is equal to $|c|$.

Figure 1 shows the evolution of $\xi(\Gamma_t)$, the adjusted Frobenius norm of the correlation matrix of daily returns, over time. The main peaks of correlation (as mentioned above, $\xi(\Gamma_t)$ is a proxy of the average correlation of pairs of stocks) occur in years 1987, 1998, 2002, 2008 and 2011. These years correspond to severe market drops and largely negative returns. Figure 1 illustrates the stylized facts of increasing correlations in bear markets, and of a lower share of idiosyncratic volatility (with respect to total volatility) in downward markets.

4. Idiosyncratic volatility and nominal prices

In general, the estimation of factor structures through nested principal component analyses (PCA) is performed by diagonalizing the covariance matrix of returns. Small firms, however, are more volatile than large firms and there is a positive correlation between the nominal price of a stock and the capitalization of the firm. As a consequence, using the covariance matrix to estimate the relative importance of idiosyncratic volatility among small price stocks and large price stocks would bias the results. For that reason, we chose to diagonalize correlation matrices of returns to neutralize the difference in total volatility between small and large price stocks.
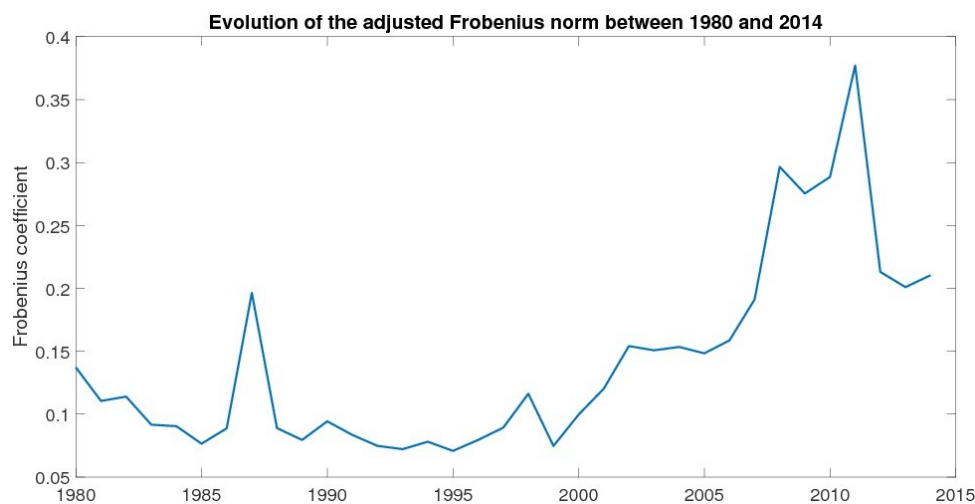

Figure 1: Evolution of the adjusted Frobenius norm of the correlation matrix of daily returns over the 35 years of the 1980-2014 period.
This figure presents the evolution of the adjusted Frobenius norm of the correlation matrix of daily returns over the 1980-2014 period.

The other advantage of this standardization is that the trace of a correlation matrix of returns is equal to the number of stocks under consideration.

4.1 The number of common factors

The number of factors of the approximate factor structure can be determined in a simple way by looking at the evolution of eigenvalues of the correlation matrix, thanks to the theorem of Chamberlain and Rotschild recalled in proposition 2. Figure 2 provides an illustration of this evolution of eigenvalues for 1999, 2008 and 2014. The horizontal axis gives the number of stocks in the PCA and the vertical axis gives the values of the eigenvalues. We represent the curves of the first three eigenvalues, ranked in decreasing order. These curves are obtained by performing nested principal component analyses, starting with 50 stocks and incrementing the number by 50 at each step until all stocks are included. The upper bold curve shows the evolution of the first eigenvalue when the number of stocks increases. The dashed (dotted) curve represents the second (third) eigenvalue.
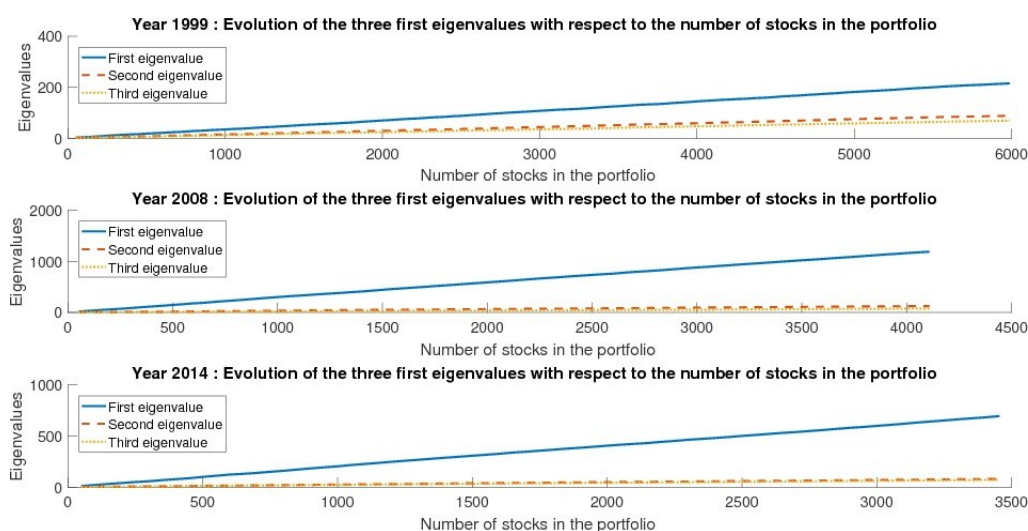

Figure 2: Evolution of the three first eigenvalues of the correlation matrix of returns with respect to the number of stocks in the PCA in 1999, 2008 and 2014

This figure presents the evolution of the first three eigenvalues of the correlation matrix of daily returns when the number of stocks included in the portfolio increases. Panel A corresponds to year 1999, Panel B to year 2008 and Panel C to year 2014. The upper bold curve corresponds to the first eigenvalue, the dashed curve corresponds to the second eigenvalue, and the dotted curve provides the values for the third eigenvalue. These curves are obtained by performing nested principal component analyses, starting with 50 stocks and incrementing the number by 50 at each step until all stocks are included (the curves for the 35 years are available upon request).

The three graphs correspond to very different years in terms of market returns. 1999 was the last year of the dotcom bubble, 2008 the year of Lehman Brothers' bankruptcy, with a yearly market drop close to 50% and an average cross-sectional volatility of 77%. Finally, 2014 was a normal year, with an average return of 3% and a 33% average volatility (see Table 1). The three graphs perfectly illustrate our discussion about the evolution of eigenvalues. As a result, the most likely number of common factors is 1 in a number of periods. Harding (2007), however, showed that considering one factor is often a biased estimate of the true number of factors because of the finite sample properties of the correlation or covariance matrix. We also, therefore, perform a robustness check with three factors to

cover the main cases. Since the number of stocks is much larger than the number of days, three factors is a reasonable choice (Bai and Ng, 2002).

4.2 Nominal stock price as a ranking criterion

Figure 3 shows the evolution of the percentage of the trace of the correlation matrix explained by the first factor in 2014 when nested PCAs are performed. The ranking criterion used to introduce new stocks (50 stocks are added at each step) in the PCAs is the nominal price of stocks. The bold (dashed) curve identifies the evolution of the first eigenvalue as a percentage of the trace when stocks are entered in the matrix in the decreasing (increasing) order of nominal price. Of course, the two curves end at the same point where all stocks are considered. The shape of the curves is qualitatively unchanged when three factors are considered. The only difference is that the upper curve starts at 40% (instead of 33%) and the lower at 15% (instead of 7%). With three factors, the two curves end at 24% (instead of 20% on figure 3). These figures show that in 2014, the first factor accounts for 20% of the trace, and factors 2 and 3 only account for 4% of the trace when all stocks are considered. We highlight that the bold curve is always above the dashed curve with one common factor, which means that the share of idiosyncratic volatility is always larger for small price stocks than for large price stocks.
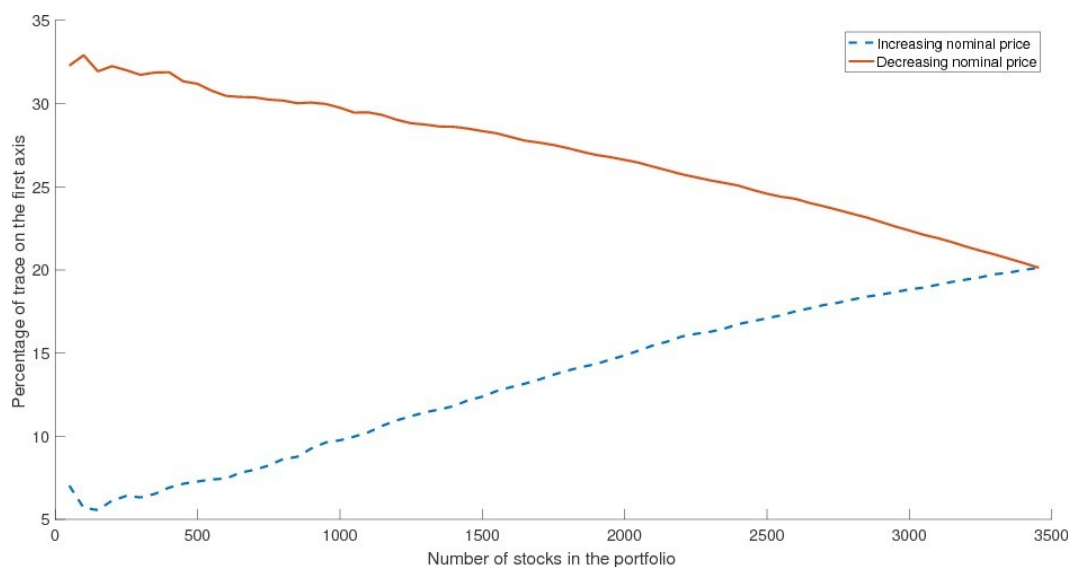

Figure 3: Percentage of trace on the first common factor as a function of the number of stocks (year 2014)
This figure presents the percentage of trace on the first axis as a function of the number of stocks for the year 2014. The nested sequence of Principal Component Analyses (PCA) starts with 50 stocks. 50 stocks are added for the next PCAs. The ranking criterion is increasing (decreasing) nominal stock price for the dashed (bold) curve.

In an efficient market, nominal prices play no role. Thus no specific shape is expected when such a ranking criterion is used, but we observe the same relationship for every single year of our 1980-2014 period. It means that if there is only one common factor, the idiosyncratic volatility of small price stocks is much higher than the idiosyncratic volatility of large price stocks.

This table shows the percentage of measure points where the idiosyncratic volatility is higher for small price stocks than for high price stocks. The two first columns give the percentages when all

stocks are considered, respectively for a 1-factor and a 3-factor model. The two last columns restrict the sample to the tercile of large capitalizations.

Each year t , denote $N_S(t)$ the number of stocks under consideration. We perform N_t nested PCAs (as in Figure 3), where $N_t = N_S/50$ because we start with 50 stocks in the first PCA and add 50 more stocks at each step. Denote $\lambda_D(n,t)$ ($\lambda_I(n,t)$) the first eigenvalue, in percentage of the trace, obtained in the *n*-th PCA when the decreasing (increasing) order of nominal prices is used to introduce stocks. The measure of the strength of the relationship between price and idiosyncratic volatility is denoted θ and defined in year t by:

$$\theta_t = \frac{1}{N_t} \sum_{n=1}^{N} 1_{\lambda_D(n,t)-\lambda_I(n,t)>0} \tag{18}$$

$$\xi_t = \frac{1}{N_t} \sum_{n=1}^{N} \left(\lambda_D(n,t) - \lambda_I(n,t) \right) \tag{19}$$

where 1_A is the indicator function of the event A worth 1 (0) when A is true(false).
$\lambda_D(n,t)$ and $\lambda_I(n,t)$ are the percentages of variance explained by the common factor. In the three-factor version, we add the three first eigenvalues but the interpretation is unchanged.

Our results over the complete period 1980-2014 are summarized in Table 2 (page 42). The first (second) column contains the indicator θ_t when a one (three)-factor model is used. The third (fourth) column gives the same result as the first (second) column when only the tercile of large caps is used in the PCAs.

We show that, either in the 1-factor model or in the three-factor model, θ_t is always equal to 100%, which means that the idiosyncratic volatility (in percentage of total volatility) is always greater for small price stocks than for large price stocks. This result could be misinterpreted as a small-size effect because of the positive correlation between nominal stock prices and market capitalizations (Baker et al., 2009). In particular, we could argue that retail investors invest more in small firms, which may generate idiosyncratic volatility in returns of small caps. Columns 3 and 4 of Table 2 show it is not the case. For 28 years out of 35, θ_t is equal to 100% in the one and three-factor versions of the model.

Moreover, when θ_t is close to 95%, the two curves of Figure 3 cross at the point located just before the final point. The only cases deserving a special comment are 2009 in the 1-factor model and 2004 in the two models for the tercile of large caps. In 2009, the two curves (respectively decreasing and increasing order; all the graphs are available upon request) are very close. Hence, the nominal stock price is not a strong determinant of the idiosyncratic volatility in this particular year. 2009 is a year of partial recovery from the 2008 crisis with an average yearly return of 30%. This average return is, however, very different for small price stocks and large price stocks. The first column of Table 3 (page 43) shows a difference of 42% between these two average returns in favor of small price stocks. Such a difference has an impact on the ranking of nominal prices that can explain the absence of a significant difference in idiosyncratic volatilities based on the stock price. For 2004, we do not have a satisfying explanation; either in the 1-factor or the 3-factor model the two curves are almost

indistinguishable when the number reaches 500. Each tercile contains approximately 1,500 firms (for a total of 4,496 firms in 2004). As a consequence, the two analyses (decreasing and increasing order of prices) with 500 firms include disjoint sets of firms. It is the reason why we do not have a satisfying interpretation for 2004.

4.3 Are small price stocks lottery-type stocks?

For small price stocks, we argue that their characteristic of lottery-type stocks may explain the results in figure 3 (Mitton and Vorkink, 2007; Kumar, 2009; Doran et al., 2011). Investors buy these stocks because they hope large (upward) price variations. In other words, investors bet on a positively skewed return distribution. The usual description of lottery-type stocks includes high volatility, small price and high skewness. Table 3 confirms this interpretation. The difference between the average moments of order 2, 3 and 4 of small and large price stocks is highly significant. Small price stocks exhibit a higher volatility, a higher skewness and a higher kurtosis than large price stocks. Although we normalized total volatilities by considering correlation matrices in our PCAs, we can nevertheless interpret differences in skewness and kurtosis as illustrations of the fact that small price stocks behave as lottery-type stocks.

The only exception concerns kurtosis in 1987 where the difference is positive, significant at the 10% level, probably due to the black Monday during which the Dow Jones index lost more than 20%. Table 3 is also in line with the conclusions of Birru and Wang (2016), who note that retail investors overestimate the room to grow of small price stocks. In particular, we observe in the first column (difference in average returns) that in 2003 and 2009, the average return of low price stocks is much higher than the average return of large price stocks. These two years correspond to starting recoveries after severe crises, illustrating the interpretation of Birru and Wang (2016). Over the 35-year period, the cumulated return on large price stocks is much higher than the corresponding return for small price stocks. Simultaneously, the volatility of large price stocks is much lower, which confirms the existence of the low volatility anomaly (Baker et al., 2011).

This table shows the yearly result of difference tests between the moments of the small price tercile and the large price tercile. The first column gives the year, the second the value of the difference of mean returns, the third the p-value of the test. The following pairs of columns are defined in the same way for the following moments, standard deviations, skewness and kurtosis.

There are some differences concerning the skewness. Indeed, the differences are not significant over the ten first years of the period but they become highly significant after 2001. More generally, when looking at the first four moments in table 3, we show a much more unstable market after 2001. More precisely, the behaviors of large price and small price stocks become significantly different in the second half of our total period under scrutiny.

	Percentages of measure points where the large price curve is above the small price curve			
Year	All stocks-1 factor	All stocks-3 factors	Large caps-1 factor	Large caps-3 factors stocks
1980	100 %	100 %	100 %	100 %
1981	100 %	100 %	100 %	100 %
1982	100 %	100 %	100 %	100 %
1983	100 %	100 %	100 %	100 %
1984	100 %	100 %	100 %	100 %
1985	100 %	100 %	100 %	100 %
1986	100 %	100 %	100 %	100 %
1987	100 %	100 %	100 %	100 %
1988	100 %	100 %	100 %	100 %
1989	100 %	100 %	100 %	100 %
1990	100 %	100 %	100 %	100 %
1991	100 %	100 %	100 %	100 %
1992	100 %	100 %	100 %	100 %
1993	100 %	100 %	100 %	100 %
1994	100 %	100 %	100 %	100 %
1995	100 %	100 %	100 %	100 %
1996	100 %	100 %	100 %	100 %
1997	100 %	100 %	100 %	100 %
1998	100 %	100 %	100 %	100 %
1999	100 %	100 %	100 %	100 %
2000	100 %	100 %	100 %	100 %
2001	100 %	100 %	94 %	97 %
2002	100 %	100 %	100 %	100 %
2003	100 %	100 %	96 %	96 %
2004	100 %	100 %	46 %	17 %
2005	100 %	100 %	100 %	100 %
2006	100 %	100 %	96 %	96 %
2007	100 %	100 %	96 %	96 %
2008	100 %	100 %	100 %	100 %
2009	100 %	100 %	8 %	100 %
2010	100 %	100 %	100 %	100 %
2011	100 %	100 %	100 %	100 %
2012	100 %	100 %	100 %	100 %
2013	100 %	100 %	100 %	100 %
2014	100 %	100 %	95 %	95 %

Table 2: Comparison of differences between one and three factors for all stocks and for large capitalizations

This table shows the percentage of measure points where the idiosyncratic volatility is higher for small price stocks than for high price stocks. The two first columns give the percentages when all stocks are considered, respectively for a 1-factor and a 3-factor model. The two last columns restrict the sample to the tercile of large capitalizations.

Year	Difference in returns	P-values	Difference in volatility	P-values	Differences in skewness	P-values	Differences in kurtosis	P-values
1980	0.0265 ***	0.0238	-0.2463 ***	0.0000	-0.6413 ***	0.0000	-4.9544 ***	0.0000
1981	0.0956 ***	0.0000	-0.2128 ***	0.0000	-0.5833 ***	0.0000	-5.2212 ***	0.0000
1982	0.1852 ***	0.0000	-0.2870 ***	0.0000	-0.6428 ***	0.0000	-6.4749 ***	0.0000
1983	0.0723 ***	0.0002	-0.3238 ***	0.0000	-0.9203 ***	0.0000	-5.8832 ***	0.0000
1984	0.3736 ***	0.0000	-0.3022 ***	0.0000	-0.3957 ***	0.0000	-5.8070 ***	0.0000
1985	0.4046 ***	0.0000	-0.3813 ***	0.0000	-0.4586 ***	0.0000	-5.5047 ***	0.0000
1986	0.2587 ***	0.0000	-0.3757 ***	0.0000	-0.5443 ***	0.0000	-3.8548 ***	0.0000
1987	0.1976 ***	0.0000	-0.4122 ***	0.0000	-0.8720 ***	0.0000	0.6039 ***	0.0755
1988	0.1662 ***	0.0000	-0.4582 ***	0.0000	-0.4505 ***	0.0000	-4.1851 ***	0.0000
1989	0.3019 ***	0.0000	-0.4598 ***	0.0000	-0.5477 ***	0.0000	-4.2677 ***	0.0000
1990	0.2692 ***	0.0000	-0.5198 ***	0.0000	-0.4745 ***	0.0000	-5.5577 ***	0.0000
1991	0.1033 ***	0.0001	-0.6229 ***	0.0000	-0.6394 ***	0.0000	-5.3492 ***	0.0000
1992	0.0134 ***	0.4581	-0.7021 ***	0.0000	-0.4417 ***	0.0000	-1.1979 ***	0.0000
1993	0.0123 ***	0.5316	-0.6254 ***	0.0000	-0.4053 ***	0.0000	-0.0871 ***	0.2519
1994	0.1359 ***	0.0000	-0.5744 ***	0.0000	-0.2794 ***	0.0000	-0.0787 ***	0.6294
1995	0.1509 ***	0.0000	-0.5999 ***	0.0000	-0.2986 ***	0.0000	-0.3176 ***	0.0022
1996	0.1638 ***	0.0000	-0.5606 ***	0.0000	-0.3545 ***	0.0000	-0.2817 ***	0.0024
1997	0.3639 ***	0.0000	-0.5205 ***	0.0000	-0.4466 ***	0.0000	-0.3913 ***	0.0000
1998	0.2487 ***	0.0000	-0.5485 ***	0.0000	-0.6428 ***	0.0000	-1.4120 ***	0.0000
1999	-0.1919 ***	0.0000	-0.5908 ***	0.0000	-0.7029 ***	0.0000	-1.9317 ***	0.0000
2000	0.2003 ***	0.0000	-0.4340 ***	0.0000	-0.4783 ***	0.0000	-1.4411 ***	0.0000
2001	-0.0167 ***	0.0007	-0.6125 ***	0.0000	-0.8066 ***	0.0000	-1.7869 ***	0.0000
2002	0.1722 ***	0.0000	-0.5360 ***	0.0000	-0.5587 ***	0.0000	-1.8708 ***	0.0000
2003	-0.5369 ***	0.0000	-0.4728 ***	0.0000	-0.7492 ***	0.0000	-2.2603 ***	0.0000
2004	0.0981 ***	0.0000	-0.3553 ***	0.0000	-0.5755 ***	0.0000	-1.7351 ***	0.0000
2005	0.1478 ***	0.0000	-0.2822 ***	0.0000	-0.4068 ***	0.0000	-1.3449 ***	0.0000
2006	0.0629 ***	0.0000	-0.2461 ***	0.0000	-0.3369 ***	0.0000	-1.4938 ***	0.0000
2007	0.1677 ***	0.0000	-0.2310 ***	0.0000	-0.3348 ***	0.0000	-1.8052 ***	0.0000
2008	0.2567 ***	0.0000	-0.3638 ***	0.0000	-0.4427 ***	0.0000	-2.0290 ***	0.0000
2009	-0.4208 ***	0.0000	-0.5560 ***	0.0000	-0.7704 ***	0.0000	-2.4335 ***	0.0000
2010	0.0598 ***	0.0012	-0.3126 ***	0.0000	-0.4924 ***	0.0000	-1.7292 ***	0.0000
2011	0.2181 ***	0.0000	-0.2638 ***	0.0000	-0.5062 ***	0.0000	-1.3395 ***	0.0000
2012	0.0664 ***	0.0003	-0.3084 ***	0.0000	-0.4507 ***	0.0000	-1.1897 ***	0.0000
2013	0.0289 ***	0.2361	-0.2828 ***	0.0000	-0.6206 ***	0.0000	-2.5452 ***	0.0000
2014	0.1306 ***	0.0000	-0.2686 ***	0.0000	-0.5366 ***	0.0000	-1.6511 ***	0.0000

Table 3: Differences between large price tercile and small price tercile

This table shows the yearly result of difference tests between the moments of the small price tercile and the large price tercile. The first column gives the year, the second the value of the difference of mean returns, the third the p-value of the test. The following pairs of columns are defined in the same way for the following moments, standard deviations, skewness and kurtosis.

5. Conclusion

Using the approximate factor structures defined by Chamberlain and Rotschild (1983), this paper highlights that nominal prices are a strong determinant of the idiosyncratic volatility of stock returns. In a large empirical study including approximately 8,000 stocks over 35 years, we show that small price stocks exhibit a much higher idiosyncratic volatility than high price stocks, even after controlling for the total volatility of returns. This small price effect is not a small size effect because the relationship remains valid for 34 of 35 years when we restrict the analysis to the tercile of large capitalizations. In fact, the higher idiosyncratic volatility is associated with a higher skewness, which allows us to conclude that small price stocks share some features with lotteries characterized by a small (or negative) expected return but a small probability of winning a big amount. Overall, our result is not in line with the efficient market hypothesis, which considers that nominal prices do not influence the distribution of returns.

References

Akerlof GA. 2007. The Missing Motivation in Macroeconomics. *American Economic Review* 97:5-36.

Albuquerque R. 2012. Skewness in Stock Returns: Reconciling the Evidence on Firm Versus Aggregate Returns. *Review of Financial Studies* 25: 1630-1673.

Bai J, Ng S. 2002. Determining the number of factors in approx- imate factor models. *Econometrica* 70:191-221.

Baker M, Bradley B, Wurgler J. 2011. Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal* 67: 40-54.

Baker M, Greenwood R, Wurgler J. 2009. Catering through nominal share prices. *Journal of Finance* 64:2559-2590.

Birru J, Wang B. 2016. Nominal Price Illusion. *Journal of Financial Economics* 119:578-598.

Campbell JY, Lettau M, Malkiel BG, Xu Y. 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *Journal of Finance* 56:1-43.

Chamberlain G, Rotschild M. 1983. Arbitrage, Factor Structure, and MeanVariance Analysis of Large Asset Markets. *Econometrica* 51:1281-1304.

Connor G, Korajczyk RA.1993. A Test for the Number of Factors in an Approximate Factor Model. *The Journal of Finance* 48:1263-1291.

Conroy RM, Harris RS, Benet BA. 1990. The Effects of Stock Splits on Bid-Ask Spreads. *The Journal of Finance* 45:1285-1295.

Copeland TE. 1979. Liquidity Changes Following Stock Splits. *The Journal of Finance* 34:115-141.

DeBondt WFM, Thaler R. 1985. Does the Stock Market Overreact?. *The Journal of Finance* 40:793-805.

Doran JS, Jiang D, Peterson DR. 2011. Gambling Preference and the New Year Effect of Assets with Lottery Features. *Review of Finance* 16: 685–731.

Geweke J, Zhou G. 1996. Measuring the Pricing Error of the Arbitrage Pricing Theory. *Review of Financial Studies* 9:557-587.

Harvey CR, Liu Y, Zhu H. 2015. …and the Cross-Section of Expected Returns. *Review of Financial Studies* 29:5-68.

Jones CS. 2001. Extracting factors from heteroskedastic asset returns. *Journal of Financial Economics* 62:293-325.

Kearney C, Poti V. 2008. Have European Stocks become More Volatile? An Empirical Investigation of Idiosyncratic and Market Risk in the Euro Area. *European Financial Management* 14:419-444.

Kumar A. 2009, Who gambles in the stock market. *Journal of Finance* 64:1889-1933.

Mitton T, Vorkink K. 2007, Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies* 20:1255-1288.

Roger P, Roger T, Schatt A. 2016. Behavioral Biases in Number Processing: the Case of Analysts' Target Prices, *Working paper 6th Helsinki Finance Summit*.

Ross SA. 1976, The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13:341-360.

Trzcinka C. 1986. On the Number of Factors in the Arbitrage Pricing Model. *The Journal of Finance* 41:347-368.

Weld WC, Michaely R, Thaler R, Benartzi S. 2009. The nominal share price puzzle. *Journal of Economic Perspectives* 23 :121-142.