Empirical Asset Pricing : HW2*

Fama-MacBeth Regressions

Maxime Borel

March 5, 2023

Introduction

In this report, I explore the sensitivity of the Fama-MacBeth regression of stock returns on 6 firm-level characteristics that are considered anomalies in the literature: gross profitability (GP), asset growth (AG), book-to-market (BE/ME), market capitalization (ME or size), momentum ($r_{12,2}$), and previous month's return ($r_{1,1}$). I begin by exploring the descriptive statistics and correlation of these variables. To assess the credibility of this report, I partially replicate the Fama-MacBeth regression in Ball et al. (2015). To further evaluate the robustness of the Fama-MacBeth regression, I run multiple regressions on different samples and model specifications. The main specification is very common in the literature, namely using annual accounting data and excess returns as the dependent variable over the entire sample.

First, I found that all of these variables capture the anomalies described in the literature in all specifications. The estimated coefficients are consistent with the economic rationale for these anomalies. On the one hand, gross profitability, asset growth, and last month's returns have the greatest explanatory power among all the variables used. On the other hand, momentum and book-to-market have comparatively low explanatory power. By adjusting the sample for size, the former become larger and more significant for large and small firms compare to micro caps. The second shows stronger anomalies in the micro-cap sample.

Second, I compare the main specification with : the abnormal returns of the Fama and French (2015) + model of the momentum factor as the dependent variable, accounting variables constructed with quarterly data and removing the financial and utilities industries from the sample. Overall, there is little change when moving to the other specifications. Therefore, I consider only the main specification for the following regressions. As mentioned earlier, I also explore the effect on coefficients of using sample composed by large, small, and micro-caps only following the definitions in Fama and French (2008). Micro-caps account for more than 50% of the sample and thus drive the point estimates in the main specification. Therefore, it is reasonable to distinguish in the following regressions between large and small firms and micro stocks.

*The code is available on github here

Finally, in the last section, I explore the impact of including interaction such as reversal with momentum, momentum with book-to-market and market capitalization with book-to-market. When only the first interaction is added, the momentum parameter decreases, meaning that the interaction better explains stock returns. When the book-to-market interactions are added to the regression, the momentum coefficient only slightly decreases. The book-to-market coefficient and t-values increase sharply. The slope of the size follows the same pattern but in the opposite direction. This shows that it is important to control for certain interactions.

Data

To construct my sample, I mainly follow Ball et al. (2015). I retrieved the monthly market values of equity and the stock returns from the Center for Research in Security Prices (CRSP) and annual/quarterly accounting data from Compustat. The sample is composed exclusively by firms quoted on NYSE, AMEX and Nasdaq. I keep only ordinary common share. When firms are delisted, I adapt the return using the delisted table from CRSP. The annual and quarterly accounting variables are lag by 6 and 4 months, respectively. The sample starts in July 1963 and ends in December 2021. At every date, I remove each observation that has missing values. The industry classification is given by the Standard Industrial Classification (SIC) codes. In some specifications, I remove financial and utilities. They are firms with SIC code starting by 6 or starting with 9 and ranging from 4900 to 4939, respectively. I obtain the risk-free rate from French's website[1]. I get the Fama-French 5-factor + momentum model from WRDS database.

In the Fama and MacBeth (1973) regressions, I use 6 anomaly variables. Namely, gross profitability (GP), asset growth (AG), book-to-market (log(BE/ME)), market capitalisation (log(ME)), the last month return/reversal ($r_{1,1}$) and the last 12 months' return skipping the first one/momentum ($r_{12,2}$). I follow Novy-Marx (2013) and trim all independent variable to the 1st and 99th percentiles.

I follow Fama and French (1993) to construct the book-to-market ratio. I compute the book value of equity as the shareholder's equity plus balance sheet deferred taxes and investment tax credit minus preferred stock. The latter is the redemption value if available, or the liquidation value or the carrying value. As mention previously, the book value of equity is lag by 6 months for annual data and by 4 for months for quarterly data. The market value of equity is lag in the same fashion as the book value following Novy-Marx (2013). The market capitalisation is the lag by one month. I construct the gross profitability following Novy-Marx (2013) as the revenue minus the cost of goods sold scaled by the total asset. I compute the asset growth following Cooper et al. (2008) which is the annual growth of total assets. Finally, the reversal is the last month return and the momentum is the last twelve months' return excluding the first one.

The dependent variable is, in most empirical studies, the monthly stock return in excess of the risk-free rate. This consists of my baseline specification. For robustness check, I also run Fama-Macbeth regressions using the factor-model adjusted return. Given the independent variable in the regression specification, a natural choice for the factor model is the 5-factor model by Fama and French (2015) + Momentum. To construct the unexplained monthly returns, I run for each stock a time-series regression and keep the residuals. A priory, the effect of using such a model for the dependent variable should decrease the explanatory power of the set of variables used and shrink the coefficients

[1]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

toward zero.

Table 1 shows the descriptive statistics of the explanatory variables. I compute the statistics for the entire panel. Table 2 presents the correlation across these variables.

Table 1: Descriptive Statistics, 1963-2021. This Table presents the descriptive statistics for the variables used in the analysis. The accounting variables are annual with 6 month lag. The value are trimmed at 1% and 99%. GP stands for Gross profitability over total asset. AG stands for total asset growth. $r_{1,1}$ 1 month return(reversal). $r_{12,2}$ is the traditional momentum.

	Mean	standard deviation	1%	25%	50%	75%	99%
GP	0.33	0.26	-0.01	0.12	0.29	0.48	1.07
AG	0.17	0.37	-0.34	-0.01	0.08	0.21	1.70
log(BE/ME)	-0.53	0.88	-2.92	-1.06	-0.45	0.07	1.30
log(ME)	5.13	2.07	1.86	3.47	4.82	6.51	10.59
$r_{1,1}$	0.01	0.13	-0.32	-0.06	0.00	0.07	0.40
$r_{12,2}$	0.12	0.51	-0.75	-0.20	0.06	0.33	1.82

Table 2: Pearson correlations, 1963-2021. This Table presents the correlations for the variables used in the analysis. The accounting variables are annual with 6 month lag. The value are trimmed at 1% and 99%. GP stands for Gross profitability over total asset. AG stands for total asset growth. $r_{1,1}$ 1 month return(reversal). $r_{12,2}$ is the traditional momentum.

	GP	AG	log(BE/ME)	log(ME)	$r_{1,1}$	$r_{12,2}$
GP	1					
AG	-0.04	1				
log(BE/ME)	-0.16	-0.20	1			
log(ME)	-0.10	0.04	-0.33	1		
$r_{1,1}$	0.02	-0.02	0.04	0.05	1	
$r_{12,2}$	0.05	-0.05	-0.14	0.14	0.01	1

Replication of Ball et al. (2015)

First, I replicate the first two columns of table 3 in Ball et al. (2015) in order to assess the quality of the construction of the variables. The goal is to assess the credibility of the sample construction. The sample starts in July 1963 and ends in December 2013. The micro capitalisation as define in Fama and French (2008) are excluded. The results are shown in table 3. The point estimates and the t-values are very close for both specifications to the one in Ball et al. (2015). In the second specification, the coefficients are almost identical except the reversal where I obtain -3.84 and they obtain -3.22. The R^2 are of the same magnitude. This suggests that the preparation of the data and the computation of the different metrics are satisfying. Therefore, the results shown in this report are most likely reliable.

Table 3: Replication of Ball et al. (2015) Panel A of Table 3 for benchmark
The sample period is 1963-2013. The table presents the average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict monthly returns. The sample is only constitute by large and small firms as define in Fama and French (2008)

	Dependent variable:	
	ER	
	(1)	(2)
Gross Profit		0.84***
		(4.38)
log(BE/ME)	0.24***	0.35***
	(2.97)	(3.89)
log(ME)	−0.07*	−0.06
	(−1.59)	(−1.5)
$r_{1,1}$	−3.63***	−3.84***
	(−6.98)	(−7.11)
$r_{12,2}$	0.99***	0.97***
	(4.07)	(3.95)
Observations	772,450	772,450
R^2	5.77%	6.34%
Note:	*p<0.1; **p<0.05; ***p<0.01	

Fama-MacBeth Regressions

The goal of this report is to assess the sensitivity of the Fama-MacBeth regression to different assumptions that can be made to construct the sample. In this study, I explore the effect of the type of dependent variable use in the regression, the frequency of the accounting data, removing certain industries and filtering the sample based on size. I start by briefly describe the results of the main specification in Table 4, column 1. This table shows coefficients estimated on the entire dataset without distinguishing between large, small and micro firms.

The sign of the point estimates are in line with the different anomaly in the literature supporting the economic reasoning of these variables. On the one hand, gross profit, book-to-market and momentum have positive coefficients. More profitable, value and past winner firms generate higher returns, respectively. On the other hand, asset growth, market capitalisation and past month return have a negative sign. Small stock, low investment firms and last month looser produce higher returns. All the coefficients are strongly significant using the Newey and West (1986) correction except the size characteristic. Nevertheless, when the standard errors are not corrected, the size parameter is significant. This is in line with previous studies (Novy-Marx (2013); Ball et al. (2015, 2016)). Momentum and book-to-market have the lowest t-values among the significant ones. The low explanatory power of momentum might be induced by the tremendous number of micro-cap firms. Moreover, the R^2 is in the same magnitude as in other studies.

The remaining three specifications measure the sensitivity of the Fama-MacBeth regressions to different sample assumptions. Column 2 of the table 4 shows how the point estimates change when the dependent variable is the abnormal return of the 5-factor Fama and French (2015) + momentum model. As expected, the point estimates approach zero, except for market capitalization and the reversal anomaly. However, the effect on size is marginal. The coefficient on the last month's return is more negative and significant, meaning that it captures an anomaly beyond the factor model. This is expected since it is the only variable not covered by the factor model and therefore removed from the residual. The R^2 decreases slightly, which is consistent with the specification since part of the dependent variable has already been partially compensated for by some factors that capture the same anomalies as those on the right side of the regression. Surprisingly, controlling for the factor model on the dependent variable made the momentum and size parameters more significant. This suggests that the construction of these two factors does not fully capture these anomalies. This logic also applies to all other variables. Overall, the effect of fitting returns to a factor model is marginal. Moreover, this specification is not common in the literature.

Column 3 presents the effect of updating accounting data on a quarterly basis. The sample in this specification is shorter and starts in July 1972 because data are not available before. Although updating accounting more frequently, the statistical importance of book-to-market decreases. The use of quarterly data mainly affects asset growth, which becomes more negative and significant. Surprisingly, the momentum slope does not become more significant with the use of quarterly data in accounting variable. This might be due to the fact that the book-to market is not constructed using the most updated version of market capitalisation. However, this specification has the lowest R^2 among others. The effect of using quarterly accounting data is again marginal and seems to reduce the explanatory power of the main specification. Therefore, it appears to be sufficient to use annual accounting data. This is consistent with the fact that accounting variables are by nature stable and therefore do not provide substantial additional information. Moreover, most studies in the literature use annual accounting variables.

Finally, it is common in the literature to exclude certain industries such as firms in the financial sector. In this report, I measure the effect of excluding firms in the financial and utilities sectors from the sample. These are firms with SIC codes starting with 6 for the former. Firms in the utilities sector are firms with SIC codes starting with 9 and ranging from 4900 to 4939. The reason for these exclusions is that companies operating in these industries have specific capital and leverage rules or different ways of handling data than others by providers. Thus, calculating the characteristic in the same way as other firms could be misleading. The result of this specification is shown in the last column of the table 4. The exclusion only concerns a small fraction of the sample. The slopes are almost not affected by the filter on the industries. Nevertheless, t-values slightly decreases compare to the main specification in column 1. This obviously induces a lower R^2 than in the main specification.

Overall, the reversal, the asset growth and the gross profit are the variables with the higher explanatory power in all specifications. The point estimates and the t-values are relatively stable through the different specification of the sample or the model. The main specification seems robust to the various alternative specifications one might think of. Given the following, the remainder of this report is based solely on the main specification.

Table 5 presents the results of the Fama-MacBeth regression based on different size criteria, namely all firms or large, small, and micro firms only. Firms are classified into size buckets according to Fama and French (2008). Micro firms are those with a market capitalization below the 20th percentile of the logarithm of market capitalization. Small firms are those between the 20th and 50th percentile and large firms are those above the 50th percentile. The column 1 is the same as in table 4. Micro-cap accounts for more than 50% of the observation in the sample, therefore, the coefficients of micro-cap subsample should be close to the one using the entire dataset.

As expected, by comparing the four columns in table 5 one can notice that the point estimates in the last columns are always larger in absolute terms than for large and small firms except for momentum. The momentum effect is weaker for micro-caps and this drive down the parameter and the t-values in the main specification. Indeed, the t-values of large and small firms reach 3.31 and 4.33 for momentum, respectively. Fama and French (2008) finds similar patterns. This is due to the fact that in OLS all the observation are equally weighted and thus micro-caps account for a large portion of the sample. However, they represents only a small portion of the total market capitalisation. It leads to overestimate the coefficients. A possible way to solve this problem is to use WLS weighted by market capitalistion. Small stocks also have larger parameters but they are in the same proportion in the entire sample. One can notice that most of the anomalies are less prominent for large stocks as all the coefficients are smaller in magnitude but also in terms of explanatory power as their t-values are lower at the exception of momentum. This is a well-known phenomenon, anomalies are usually stronger for small stocks.

Therefore, excluding micro-cap stocks is coherent and it is a common practice in the literature. In the next section, I run regression including interaction terms for a sample with large and small companies and another using micro-cap stock only.

Table 4: The table presents the average Fama and MacBeth (1973) regression slopes (multiplied by 100) and their t-values from cross-sectional regressions that predict monthly returns. the sample goes from July 1963 to December 2021 on a monthly basis. The dependent variable in column (1) is the excess return using the treasury bill from Keneth french website. The dependent variable in column (2) is the predicted return from the FF5 + mom model. The dependent variable in column (3) is the same as (1) but with quarterly accounting data. The sample in column (4) excludes financial and utilities firm. The firms with SIC code starting with a 6 or between 9. The dependent variable in column (4) is the same as (1). The t-values are computed using Newey and West (1986).

	Dependent variable:			
	ER	FF6	ER-Q	ER-Ind
	(1)	(2)	(3)	(4)
Gross Profit	0.87***	0.61***	0.81***	0.87***
	(5.44)	(4.78)	(5.51)	(5.05)
Asset Growth	−0.85***	−0.66***	−1.13***	−0.85***
	(−6.99)	(−5.93)	(−7.74)	(−6.94)
log(BE/ME)	0.32***	0.18***	0.35***	0.32***
	(4.24)	(3.59)	(3.85)	(4.12)
log(ME)	−0.06	−0.08***	−0.06	−0.07
	(−1.32)	(−3.22)	(−1.3)	(−1.39)
$r_{1,1}$	−6.07***	−6.67***	−5.11***	−6.09***
	(−9.98)	(−10.51)	(−9.36)	(−9.8)
$r_{12,2}$	0.59***	0.52***	0.50**	0.53**
	(2.79)	(3.24)	(2.1)	(2.41)
Constant	1.01**	0.53**	1.15**	1.03**
	(2.42)	(2.54)	(2.36)	(2.27)
Observations	2,434,644	2,369,538	1,710,509	2,078,197
R^2	4.24%	3.8%	3.43%	4.18%
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 5: The table presents the average Fama and MacBeth (1973) regression slopes (multiplied by 100) and their t-values from cross-sectional regressions that predict monthly returns. the sample goes from July 1963 to December 2021 on a monthly basis. The dependent variable in column (1) is the excess return using the treasury bill from Keneth french website. The size definition are the same as in Fama and French (2008). The t-values are computed using Newey and West (1986).

	Dependent variable:			
	All	Large	Small	Micro
	(1)	(2)	(3)	(4)
Gross Profit	0.87^{***}	0.60^{***}	0.76^{***}	0.79^{***}
	(5.44)	(3.82)	(4.24)	(4.42)
Asset Growth	-0.85^{***}	-0.36^{**}	-0.52^{***}	-1.03^{***}
	(-6.99)	(-2.38)	(-3.82)	(-6.71)
log(BE/ME)	0.32^{***}	0.15^{**}	0.28^{***}	0.40^{***}
	(4.24)	(2.20)	(3.34)	(4.74)
log(ME)	-0.06	-0.05	-0.02	-0.17^{*}
	(-1.32)	(-1.61)	(-0.27)	(-1.96)
$r_{1,1}$	-6.07^{***}	-3.54^{***}	-3.56^{***}	-7.14^{***}
	(-9.98)	(-6.04)	(-6.93)	(-8.76)
$r_{12,2}$	0.59^{***}	0.79^{***}	0.97^{***}	0.40^{*}
	(2.79)	(3.31)	(4.33)	(1.91)
Constant	1.01^{**}	0.83^{**}	0.73	1.45^{**}
	(2.42)	(2.29)	(1.57)	(2.56)
Observations	2,434,644	536,157	516,080	1,380,936
R^2	4.24%	8.01%	5.42%	3.1%
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6: The table presents the average Fama and MacBeth (1973) regression slopes (multiplied by 100) and their t-values from cross-sectional regressions that predict monthly returns including interactions. the sample goes from July 1963 to December 2021 on a monthly basis. The dependent variable in column (1)-(6) is the excess return using the treasury bill from Keneth french website. The columns (4)-(6) are based on all but micro-caps. The size definition are the same as in Fama and French (2008). The t-values are computed using Newey and West (1986).

	Dependent variable:					
	ER			ER-All-but micro		
	(1)	(2)	(3)	(4)	(5)	(6)
Gross Profit	0.83^{***}	0.77^{***}	0.76^{***}	0.68^{***}	0.66^{***}	0.65^{***}
	(5.24)	(4.92)	(4.87)	(2.57)	(4.3)	(4.15)
Asset Growth	-0.89^{***}	-0.87^{***}	-0.89^{***}	-0.50^{***}	-0.48^{***}	-0.49^{***}
	(−7.0)	(−7.24)	(−6.97)	(−4.08)	(−3.96)	(−4.18)
log(BE/ME)	0.33^{***}	0.71^{***}	0.70^{***}	0.23^{***}	0.55^{***}	0.55^{***}
	(4.23)	(6.95)	(6.77)	(3.14)	(3.71)	(3.76)
log(ME)	−0.06	-0.12^{***}	-0.12^{**}	-0.06^{*}	-0.11^{***}	-0.11^{***}
	(−1.36)	(−2.59)	(−2.55)	(−1.81)	(−3.14)	(−3.09)
$r_{1,1}$	-5.85^{***}	-6.13^{***}	-5.86^{***}	-4.04^{***}	-3.63^{***}	-4.07^{***}
	(−9.31)	(−10.26)	(−9.34)	(−8.02)	(−7.7)	(−8.55)
$r_{12,2}$	0.45^{**}	0.54^{**}	0.49^{**}	0.65^{**}	0.74^{***}	0.65^{***}
	(2.0)	(2.38)	(2.17)	(2.55)	(2.94)	(2.68)
$r_{12,2}$ X $r_{1,1}$	7.50^{***}		7.45^{***}	3.48^{***}		3.50^{***}
	(6.86)		(4.97)	(4.78)		(4.93)
$r_{12,2}$ X log(BE/ME)	-0.30^{***}	-0.24^{**}	-0.26^{***}	-0.23^{*}	-0.24^{*}	-0.23^{**}
	(−3.24)	(−2.56)	(−2.73)	(−1.92)	(−1.88)	(−1.97)
log(BE/ME) X log(ME)		-0.08^{***}	-0.07^{***}		-0.05^{**}	-0.05^{**}
		(−4.99)	(−4.84)		(−2.45)	(−2.50)
Constant	1.02^{***}	1.32^{***}	1.32^{***}	0.92^{***}	1.23^{***}	1.23^{***}
	(2.5)	(3.22)	(3.23)	(2.57)	(3.51)	(3.44)
Observations	2,434,644	2,434,644	2,434,644	1,053,708	1,053,708	1,053,708
R^2	4.56%	4.47%	4.67%	7.23%	7.08%	7.33%
Note:				$^{*}p < 0.1$;$^{**}p < 0.05$;$^{***}p < 0.01$		

Fama MacBeth Regressions with Interactions

In this section, I explore the effect of including interaction of explanatory variables. There are 15 possible interactions. I decide to examine only a subset of them, namely the interaction between momentum with reversal, momentum with book-to-market and market capitalization with book-to-market. I select these combinations because the first is two return-based metrics. They are constructed in such a way that they are not overlapping. The motivation for the second, momentum and book-to-market, is that these two anomalies are negatively correlated as described in Asness et al. (2013). The reason behind the latter is because two variables control well for each other as mention in Fama and French (1993).

Table 6 presents the results of the Fama-MacBeth regression including interaction variables. The first three columns are based using the entire sample and the last three columns are excluding the micro-cap stock as suggested in the previous section.

The interaction between momentum and reversal has a large and positive coefficient in all specifications. Moreover, when this interaction is included in the regression, the point estimates for momentum decrease significantly. This suggests that this interaction is complementary to momentum. The magnitude of the parameter is similar to that of the reversal.

The interaction between momentum and book-to-market and book-to-market with size have negative and significant coefficients in all specifications. The momentum parameter is less affected by these two interactions, but is still reduced. The book-to-market coefficient increases dramatically and becomes highly significant. The size anomaly shows the same pattern but in the other direction. As a side effect, the inclusion of all these interactions also increases the intercept.

References

C. S. Asness, T. J. Moskowitz, and L. H. Pedersen. Value and momentum everywhere. *The journal of finance*, 68(3):929–985, 2013.

R. Ball, J. Gerakos, J. T. Linnainmaa, and V. V. Nikolaev. Deflating profitability. *Journal of Financial Economics*, 117(2):225–248, 2015.

R. Ball, J. Gerakos, J. T. Linnainmaa, and V. Nikolaev. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, 121(1):28–45, 2016.

M. J. Cooper, H. Gulen, and M. J. Schill. Asset growth and the cross-section of stock returns. *the Journal of Finance*, 63(4):1609–1651, 2008.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

E. F. Fama and K. R. French. Dissecting anomalies. *The journal of finance*, 63(4): 1653–1678, 2008.

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

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. 1986.

R. Novy-Marx. The other side of value: The gross profitability premium. *Journal of financial economics*, 108(1):1–28, 2013.