To what extent can we predict the return of a portfolio based on changes in the operating profitability and investment of the bond issuers?

Group 23, Predictability
Nikodem Baehr (2076515)
Boris Baneke (2076430)
Samuel Friedlaender (2070091)
Maximillian Malinov (2078507)
Iris Zweerman (2082926)

May 2022

Word count: 3859

Introduction

Investing is an action that usually carries significant risks with it. Participants in the stock market across the globe share one common goal - put their money into the correct asset so as to generate profits over time. However, the decision where to invest is tied to numerous factors constantly altering the price of stocks. Predicting one's returns based on preliminary research and data analysis is a complicated task which doesn't guarantee accurate outcomes-sudden shocks to the markets are always a present threat. There are many distinct factors to take into consideration when trying to work out the future change in the prices of assets. This paper will focus on the effects of two of them: a company's operational profitability (OP) and realized investments (INV). Utilising the mathematical principles of regression analysis, we aim to illustrate the absence of enough certainty to forecast the predicted gains or losses from a portfolio based on the two factors mentioned above. The results of our research indicate that the operational profitability and investments related to a company, recorded over long time intervals, are not determinant for the outcome of a portfolio.

Abstract

To investigate the potential existence of a correlation between the investments in a company and its operational profitability on one side and the stock value of the corporation on the other, we chose a set of data covering the last 20 years (25 Portfolios Formed on Operating Profitability and Investment (5 x 5)).The exact period we are investigating is April 2002-March 2022. We will be using the section 'Average Value Weighted Returns - Monthly'). The set includes 25 different portfolios from the Kenneth French website. They all describe the average value-weighted returns of portfolios from companies characterized by varying operational probability and investment levels. For some the former is low, while the latter is high, or vice versa. Our aim is to see whether there is

a link between how much companies make in terms of operational profitability (total income after paying taxes, salaries, and other expenses) and the increase/decrease in its presence on the stock market (by investment we refer to the percentage increase in the number of stocks the company has issued between consecutive periods). To perform our research, we make use of the linear regression principle and the STATA software which allows us to make the computations with the data. We run 3 different types of regressions: with data for the past 12 months to see whether we can predict based only on the past year; with data for the past 20 years to investigate whether larger data span generates higher predictability; with data for the past 20 years but with the data for the major financial crises excluded so that we can observe if we can talk about long term predictability (long term investments are less prone to temporary financial shocks since many stocks seem to constantly increase or decrease over long periods of time).

Literature review

Investment

In an article from the Federal Reserve Bank Philadelphia, they explain that the stock prices are based on two components. Firstly, the fundamental value which is the true value of the company. And secondly, the nonfundamental value, which has to do with the expected cash flow and thus can fluctuate based on private information. Private information is information which some investors, mostly hedge funds have, but the firm doesn't have. This information can influence the decisions these funds make. All of that is then reflected in the expected cash flow. Firms which depend on issuing equity for financing new investments can learn from the price of their stocks and thus be influenced by the private information. The firm can then decide whether to issue equity or not. This is still risky because as evidence shows if a stock is undervalued then extra investment means losing money but if stock overvalued then winning money by increasing investment. (Baker, Wurgler; 2003). So stock prices may affect investment decisions because these prices provide new information to firms about the profitability of their investment opportunities. That means that it would make sense that an equity-dependent firm is likely to adjust their prices to the information their stock price gives. That result is also found in the book Review of Financial Studies (2007), in which they also write that the effect of the change in investment may make price manipulation possible.

Profitability

According to empirical evidence (Sholichah, Fatmawati; 2021) profitability has a direct significant positive relationship with stock prices, which means that higher profitability will drive the stock prices up. That is because higher profitability means a higher expected cash flow. Since expected cash flow also has to do with for example dividend payments and other factors, which we aren't considering in our paper, we can't be sure that our results will show the same.

Definitions

- OLS estimates- Ordinary Least Squares estimates of the coefficients of the terms in the linear function $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
- x_i for i $\in$ 1, 2, ..., k - regressors (independent variables); in our case these are the portfolio returns which we take from the data set.
- β_i is a coefficient, indicating a change in the dependent variable given a one unit increase in a regressor x_i, ceteris paribus
- Standard error - indicative of the spread of the data; same concept as standard deviation but refers to the examined sample rather than the entire population; if there was a large population with identical estimates, the standard error would constitute its standard deviation.
- p-value - the probability that a given statistic is greater than or equal to the results we observe given the distribution the statistic follows.
- Significant estimators - an estimate obtained via regression is considered to be significant if its p-value is less than or equal to 0.05 since we are working using 95% confidence intervals.
- Confidence interval- A P% confidence interval indicates the interval in which the value of the investigated statistic belongs with P% certainty. The higher P is, the narrower the interval becomes (where P is a value between 0 and 100).
- R^2 and Adjusted R^2- the indicator R^2 constitutes a percentage which illustrates the precision of the results obtained in the regression model. The value of this variable is indicative of the fraction of the variance of the dependent variable (the future returns of portfolios in our study) that is explained by the conducted regression. The Adjusted R^2 is an indicator with identical properties and interpretations, the only difference being that it is adjusted for the number of independent variables involved in the model.

20 years with crises

To check whether the average value weighted return of a portfolio is predictable, we decided to use an OLS regression model. In the regression the return of some asset or portfolio is the dependent variable, while other variables such as past returns can serve as independent variables. If the regression coefficient of one or more of the independent variables turns out to be significantly different from zero, then we say that there is predictability. We will be running a regression of a given portfolio's average value weighted return (dependent variable) against the past 12 pieces of data (12 lags) regarding the given portfolio. Our confidence level is 95% and significance level is 5%. First to check, whether the variance of the error term is constant we must check for heteroskedasticity. Heteroskedasticity implies that the variance of the error term depends upon the particular value of x_i (Wooldridge, 2012) instead of being constant. The Breusch-Pagan test for heteroskedasticity results in not rejecting the initial hypothesis (homoskedasticity) only for OP2INV4 portfolio. Therefore regressing all the portfolio returns, we will be using a robust t-statistics for all the portfolio returns, apart from the aforementioned one, where we will conduct a normal OLS regression.

After running the regressions, we can conclude that 3 out of 25 (12%) portfolios can be said

to be predictable. These are OP3INV1 OP4INV4 HiOPHiINV. In each of the portfolios, there was always only one lag showing the significance. R^2, regarding all portfolios, was never bigger than 0.105 indicating that only 10.5% or less (depending on the portfolio) of the variance of dependent variable was explained by variables in the regression model. We decided to describe in details 5 of 25 portfolios.
These portfolios are: LoOPLoINV, OP2INV2, OP3INV3, OP4INV5 and HiOPHiINV.

Starting with LoOPLoINV, the regression against 12 lags, shows that the none of the previous 12 data points can define the current return of the portfolio. R^2 is slightly more than 0.04 showing that vast majority of the variance of the portfolio's return isn't explained by regressors. Looking at the p-values, the average value weighted return of the portfolio could be said to be predictable at 85% confidence, but this is too uncertain, again combining that fact with a small R^2. Also, the F test for the model shows that all of the coefficients in the regression have no influence on the current average value weighted return. The lower value of Prob>F, the better the model as we can use it with higher confidence. If the value of Prob>F is less than 0.05, we can conclude that the model is significant at 95% confidence level.

Linear regression

Number of obs	=	228
F(12, 215)	=	0.57
Prob > F	=	0.8653
R-squared	=	0.0416
Root MSE	=	6.7006

looploinv	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
looploinv						
L1.	.1430715	.0945578	1.51	0.132	-.0433075	.3294505
L2.	-.0165632	.07862	-0.21	0.833	-.1715279	.1384015
L3.	.0236794	.0692084	0.34	0.733	-.1127344	.1600933
L4.	.0179979	.0672164	0.27	0.789	-.1144896	.1504855
L5.	.0172475	.0758993	0.23	0.820	-.1323544	.1668495
L6.	-.0977265	.0725652	-1.35	0.179	-.2407568	.0453037
L7.	.0082222	.0763465	0.11	0.914	-.1422613	.1587057
L8.	-.056572	.0810589	-0.70	0.486	-.2163438	.1031999
L9.	-.0367256	.0633663	-0.58	0.563	-.1616243	.088173
L10.	-.0544981	.0705827	-0.77	0.441	-.1936208	.0846245
L11.	-.0034352	.0669384	-0.05	0.959	-.1353747	.1285043
L12.	.0028049	.0544626	0.05	0.959	-.104544	.1101539
_cons	1.147818	.6020313	1.91	0.058	-.0388213	2.334457

Figure 1: Regressing LoOPLoINV over 12 lags (20 years with crises)

Next is portfolio OP2INV2 - the regression results are a bit better, however still not showing predictability. The coefficient of the sixth lag is closest to be accepted at 95% significance level, resulting in its p-value of 0.066. R^2 is slightly better than compared to the previously described portfolio with the value 0.0478. The F test for the model shows more significance, since Prob>F = 0.2461, than the previous portfolio which had Prob>F = 0.8653. The constant coefficient shows significance, but we cannot accept the model to be just the constant ± the robust standard error.

Regression on portfolio OP3INV3 shows worse results than the previous one. Lower R^2 and higher Prob>F, resulting in lower confidence of the model. Here again, one coefficient is close to being significant. That is the coefficient of the 9th lag with the p-value of 0.062, making it significant at 90% confidence interval, but not at 95%. The model is not reliable enough to predict the portfolio's returns.

Linear regression

Number of obs	=	228
F(12, 215)	=	1.26
Prob > F	=	0.2461
R-squared	=	0.0478
Root MSE	=	5.3212

op2inv2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
op2inv2						
L1.	-.0138134	.1028276	-0.13	0.893	-.2164926	.1888659
L2.	-.0136372	.0746906	-0.18	0.855	-.1608569	.1335824
L3.	.0090165	.0733176	0.12	0.902	-.1354967	.1535298
L4.	.0515686	.0749832	0.69	0.492	-.0962278	.1993649
L5.	-.0276375	.0642644	-0.43	0.668	-.1543064	.0990314
L6.	-.1159509	.062827	-1.85	0.066	-.2397866	.0078848
L7.	.0915615	.0792575	1.16	0.249	-.0646597	.2477827
L8.	.0401471	.0749947	0.54	0.593	-.1076718	.187966
L9.	-.1060564	.0687238	-1.54	0.124	-.241515	.0294023
L10.	-.031007	.0683144	-0.45	0.650	-.1656588	.1036448
L11.	-.008171	.0663153	-0.12	0.902	-.1388825	.1225405
L12.	.0099728	.0597585	0.17	0.868	-.1078148	.1277604
_cons	1.174291	.5643208	2.08	0.039	.0619813	2.286601

Figure 2: Regressing OP2INV2 over 12 lags (20 years with crises)

Linear regression

Number of obs	=	228
F(12, 215)	=	0.67
Prob > F	=	0.7799
R-squared	=	0.0374
Root MSE	=	4.6727

op3inv3	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
op3inv3						
L1.	.0390037	.1123118	0.35	0.729	-.1823695	.2603769
L2.	-.0449892	.0766813	-0.59	0.558	-.1961326	.1061542
L3.	.0336741	.0785081	0.43	0.668	-.1210701	.1884183
L4.	.0144896	.0813042	0.18	0.859	-.1457658	.1747451
L5.	-.0061578	.0681576	-0.09	0.928	-.1405004	.1281848
L6.	-.0787705	.0714238	-1.10	0.271	-.2195511	.0620101
L7.	.0193323	.0774374	0.25	0.803	-.1333014	.171966
L8.	.0568674	.0798204	0.71	0.477	-.1004632	.2141981
L9.	-.1279687	.0681842	-1.88	0.062	-.2623638	.0064264
L10.	-.0165049	.0686479	-0.24	0.810	-.1518139	.1188041
L11.	-.0501293	.070226	-0.71	0.476	-.1885489	.0882902
L12.	.050065	.0660994	0.76	0.450	-.0802208	.1803508
_cons	1.194508	.5568689	2.15	0.033	.0968867	2.29213

Figure 3: Regressing OP3INV3 over 12 lags (20 years with crises)

OP4INV5 shows the smallest R^2 out of the 5 described with only 0.0358. Prob>F is very high again showing the model can't be described as a reliable one. All of the coefficients are insignificant (not including the constant one). Therefore, given the data, the portfolio's return is not predictable.

Linear regression						
				Number of obs	=	228
				F(12, 215)	=	0.72
				Prob > F	=	0.7291
				R-squared	=	0.0358
				Root MSE	=	4.9458

op4inv5	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
op4inv5						
L1.	.0530944	.0956405	0.56	0.579	-.1354187	.2416074
L2.	-.0300815	.0712935	-0.42	0.673	-.1706053	.1104423
L3.	.0339389	.0765649	0.44	0.658	-.116975	.1848527
L4.	-.0134893	.074178	-0.18	0.856	-.1596985	.1327199
L5.	-.0298059	.0726045	-0.41	0.682	-.1729136	.1133018
L6.	-.142816	.0734043	-1.95	0.053	-.2875002	.0018682
L7.	.0493945	.0681017	0.73	0.469	-.0848379	.183627
L8.	.0146632	.0714749	0.21	0.838	-.1262181	.1555445
L9.	-.068295	.0647931	-1.05	0.293	-.1960061	.0594161
L10.	-.037812	.0651219	-0.58	0.562	-.166171	.0905471
L11.	-.036509	.0676372	-0.54	0.590	-.1698259	.0968079
L12.	-.0040693	.0637718	-0.06	0.949	-.1297674	.1216287
_cons	1.447091	.5599206	2.58	0.010	.3434546	2.550728

Figure 4: Regressing OP4INV5 over 12 lags (20 years with crises)

Portfolio HiOPHiINV is the only portfolio to be said to be predictable out of the 5 described. The coefficient of the 6th lag is significant, with the value of about -0.12. The interpretation of that is given an increase of the 6th lag of the return of the portfolio by 1 unit of the return, the current return of the portfolio will decrease by 0.12, ceteris paribus. The R^2 is the highest out of all described portfolios (0.0501). In the model, the 9th lag is slightly insignificant with a p-value of 0.051. According to our assumption, the return of a portfolio is predictable if one of the coefficients is significant. However, the F test proves that the model is insignificant (Prob>F=0.2824).

Linear regression						
				Number of obs	=	228
				F(12, 215)	=	1.20
				Prob > F	=	0.2824
				R-squared	=	0.0501
				Root MSE	=	5.1906

hiophiinv	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
hiophiinv						
L1.	.043276	.0946747	0.46	0.648	-.1433333	.2298854
L2.	.0116183	.0815416	0.14	0.887	-.149105	.1723417
L3.	-.0607589	.0860002	-0.71	0.481	-.2302703	.1087525
L4.	.0149617	.077002	0.19	0.846	-.1368138	.1667372
L5.	-.024711	.0715153	-0.35	0.730	-.1656719	.11625
L6.	-.1196993	.0581181	-2.06	0.041	-.2342535	-.0051452
L7.	.0802907	.0665913	1.21	0.229	-.0509648	.2115462
L8.	-.0651523	.0661533	-0.98	0.326	-.1955443	.0652397
L9.	-.1379274	.0703402	-1.96	0.051	-.2765722	.0007173
L10.	.0087487	.0596935	0.15	0.884	-.1089107	.126408
L11.	-.0680773	.066523	-1.02	0.307	-.1991981	.0630434
L12.	.0071891	.0693378	0.10	0.918	-.1294797	.143858
_cons	1.590853	.529128	3.01	0.003	.5479109	2.633796

Figure 5: Regressing HiOPHiINV over 12 lags (20 years with crises)

20 years without crises

As already stated, to explore whether the exclusion of major economical and political crises that could have caused the prices of stocks to go up or down would make an investment more predictable or not, we will run a time series regressions of the same 25 portfolios used before but this time with the data for the periods 01.12.2007-31.05.2009 (the Financial Crisis of 2007-2008) and 20.02.2020-07.04.2020 (the first and most severe hit of the Covid pandemic) removed from our data set. The exact dates of the crises have been taken from two articles, published by Forbes and CNBC, respectively (). Our aim here was to see whether we could distinguish some trends that are persistent over time irrespective of the emergence of crises, which normally cause extraordinary volatility on the market and temporarily inflated/deflated prices of stocks. The methodology of our study remains the same as described before, so the confidence interval we use is again 95%, significance level is 5%, and we regress each dependent variable (present value of a given portfolio) against 1, 2, up to 12 lags, so we trace whether the stock price in a given month can be predicted based on the stock's value 1 month before (lag 1), 2 months before (lag 2), etc. up until 12 months before (lag 12). In a similar way as before, we run the Brausch-Pagan test for each regression to establish homo- or heteroskedasticity. Here arises the first distinction between the analysis of the stock prices over the past 20 years with and without crises-with crises all but one of the regressions were heteroskedastic, while here we most of the results of the Brausch-Pagan test (20 out of 25) point at homoscedasticity. Thus, only 5 of the portfolios require the robust t-statistic regression, and for the remaining 20 we can use ordinary OLS-regressions.

After running the regressions, we can conclude that in this adjusted model it makes more sense to talk about some predictability of portfolio returns over time-12 portfolios out of the 25 pointed at predictability for at least 1 lag (almost 5%). These portfolios are: OP1INV4, OP2INV2, OP3INV2, OP3INV3, OP3INV4, OP4INV1, OP4INV3, OP4INV5, HIOPLOINV, OP5INV2, OP5INV3, and HIOPHIINV. However, considering that only two of the aforementioned portfolios, OP4INV1 and OP4INV3 in particular, yielded a Prob>F value of less than 0.05, we can't assume the regressions to provide reliable results. In addition, the value for the R^2 never exceeded 10.99%, so we can again deduce the imprecision of the model as it manages to capture only a small fraction of the variance in the dependent variables.To illustrate the results of our research, we will now describe what we obtained regarding the portfolios LOOPLOINV, OP2INV2, OP3INV3, OP4INV5 and HIOPHIINV.

Starting with the first one of them, LOOPLOINV, none of the lags seem to be able to predict the current value of the portfolio with 95% confidence. Only the lags for 7 and 12 months are significant at the 80%-confidence level, but the negligibly small R^2 value, estimated to be 3.36% indicates that a tiny fraction of the variance in the dependent variable has been explained by the model we constructed. In addition, the Prob>F value is very high (0.8729), so there is a high chance that all coefficients in the regressions are actually 0. Thus, we conclude that the given portfolio's returns are unpredictable given the data set we constructed.

Source	SS	df	MS			
				Number of obs	=	206
				F(12, 193)	=	0.56
Model	227.631286	12	18.9692738	Prob > F	=	0.8729
Residual	6550.33988	193	33.9395849	R-squared	=	0.0336
				Adj R-squared	=	-0.0265
Total	6777.97117	205	33.063274	Root MSE	=	5.8258

looploinv	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
looploinv						
L1.	.0262524	.0716744	0.37	0.715	-.1151133	.1676181
L2.	.0310463	.0715554	0.43	0.665	-.1100847	.1721773
L3.	-.0060953	.072356	-0.08	0.933	-.1488053	.1366147
L4.	-.0653509	.0709234	-0.92	0.358	-.2052353	.0745334
L5.	.0446246	.069415	0.64	0.521	-.0922849	.181534
L6.	-.009008	.068907	-0.13	0.896	-.1449154	.1268995
L7.	-.0900996	.0675155	-1.33	0.184	-.2232626	.0430634
L8.	-.0284087	.0675389	-0.42	0.674	-.1616178	.1048004
L9.	.0091017	.0671083	0.14	0.892	-.1232581	.1414615
L10.	-.0129417	.0665374	-0.19	0.846	-.1441755	.1182921
L11.	-.0618213	.0661242	-0.93	0.351	-.1922401	.0685975
L12.	-.0900544	.065962	-1.37	0.174	-.2201533	.0400445
_cons	1.789611	.5243608	3.41	0.001	.755398	2.823825

Figure 6: Regressing LOOPLOINV over 12 lags (20 years without crises)

Next we have OP2INV2, which was discussed earlier when we investigated the predictability over 20 years with crises. Here we see that excluding the prices changed the results we got-now the returns of the portfolio are predictable based on its value 1 month ago (lag 1). The coefficient of lag1 in the regression has a p-value of 0.022, which indicates that with 95% confidence we can say that there exists predictability based on the value of the portfolio one month back in time. However, the p-value alone isn't a sufficient factor to claim overall success of the model, since both the R^2 is tiny (6.84%) and the Prob>F is quite large (0.2990), so the overall precision of the model used in this regression can't be trusted. Thus, our final inference is that the portfolio is unpredictable.

Source	SS	df	MS			
				Number of obs	=	206
				F(12, 193)	=	1.18
Model	277.922524	12	23.1602103	Prob > F	=	0.2990
Residual	3784.74696	193	19.6100879	R-squared	=	0.0684
				Adj R-squared	=	0.0105
Total	4062.66949	205	19.8178999	Root MSE	=	4.4283

op2inv2	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
op2inv2						
L1.	-.165776	.0716272	-2.31	0.022	-.3070485	-.0245034
L2.	-.0571262	.0716854	-0.80	0.426	-.1985136	.0842612
L3.	-.0476095	.071029	-0.67	0.503	-.1877022	.0924833
L4.	-.1198664	.0703282	-1.70	0.090	-.2585768	.0188441
L5.	-.0356892	.070406	-0.51	0.613	-.1745532	.1031747
L6.	-.1160903	.0701059	-1.66	0.099	-.2543623	.0221817
L7.	-.1019918	.0691467	-1.48	0.142	-.238372	.0343883
L8.	-.0319427	.069283	-0.46	0.645	-.1685918	.1047064
L9.	-.0638497	.0673007	-0.95	0.344	-.196589	.0688896
L10.	-.10329	.0670728	-1.54	0.125	-.2355799	.0289998
L11.	-.0872119	.0672146	-1.30	0.196	-.2197815	.0453576
L12.	-.0771004	.066994	-1.15	0.251	-.2092348	.0550339
_cons	2.813054	.5440673	5.17	0.000	1.739972	3.886135

Figure 7: Regressing OP2INV2 over 12 lags (20 years without crises)

A portfolio which is closer to being predictable is OP3INV3. The lags for 1, 7, and 11 months were all significant. The p-value for each is around 0.03, so they should be statistically significant. The R^2 comes at 9.86%, which indicates that the regression fails to account for most of the variance in the dependent variable. Even though we get a lower Prob>F value (0.0575), which is closer to the desired 0.05 margin, we would need another model if we wanted to talk about the existence of predictability because of the R^2.

Source	SS	df	MS			
				Number of obs	=	206
				F(12, 193)	=	1.76
Model	275.077522	12	22.9231268	Prob > F	=	0.0575
Residual	2515.03343	193	13.0312613	R-squared	=	0.0986
				Adj R-squared	=	0.0425
Total	2790.11096	205	13.6102973	Root MSE	=	3.6099

op3inv3	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
op3inv3						
L1.	-.1585404	.0722024	-2.20	0.029	-.3009474	-.0161334
L2.	-.0072634	.0718047	-0.10	0.920	-.148886	.1343592
L3.	-.0815931	.0706693	-1.15	0.250	-.2209764	.0577903
L4.	-.1301079	.0710003	-1.83	0.068	-.2701441	.0099283
L5.	.0214315	.0712897	0.30	0.764	-.1191753	.1620384
L6.	-.0640977	.0709159	-0.90	0.367	-.2039674	.0757719
L7.	-.1508448	.0699807	-2.16	0.032	-.28887	-.0128197
L8.	.0621007	.0706302	0.88	0.380	-.0772055	.2014069
L9.	-.053044	.0686771	-0.77	0.441	-.1884979	.08241
L10.	-.0569203	.0677887	-0.84	0.402	-.1906221	.0767815
L11.	-.1477576	.0678939	-2.18	0.031	-.2816669	-.0138482
L12.	-.0202221	.0686441	-0.29	0.769	-.155611	.1151668
_cons	2.611901	.4972517	5.25	0.000	1.631155	3.592646

Figure 8: Regressing OP3INV3 over 12 lags (20 years without crises)

The next portfolio to be described is OP4INV5. The regression model indicates that the value of the portfolio in the previous month (lag 1) is significant on the 95%-confidence level, but all others aren't. The Prob>F value also is large, coming at 0.3866, so we aren't certain that the coefficients obtained in our regression are not all 0. Only the R^2 is low, 6.24% to be precise, so we conclude that this portfolio still doesn't allow us to talk about predictability in gerenal.

Source	SS	df	MS			
				Number of obs	=	206
				F(12, 193)	=	1.07
Model	238.973144	12	19.9144287	Prob > F	=	0.3866
Residual	3588.47842	193	18.5931525	R-squared	=	0.0624
				Adj R-squared	=	0.0041
Total	3827.45157	205	18.6704955	Root MSE	=	4.312

op4inv5	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
op4inv5						
L1.	-.1434321	.0718531	-2.00	0.047	-.2851503	-.0017138
L2.	-.1089841	.0726792	-1.50	0.135	-.2523315	.0343634
L3.	-.0617209	.074325	-0.83	0.407	-.2083145	.0848726
L4.	-.1031148	.0740922	-1.39	0.166	-.2492492	.0430196
L5.	.0145994	.0742189	0.20	0.844	-.131785	.1609837
L6.	-.13961	.074456	-1.88	0.062	-.286462	.0072419
L7.	-.0334634	.0738528	-0.45	0.651	-.1791256	.1121988
L8.	-.0008865	.0735891	-0.01	0.990	-.1460286	.1442555
L9.	-.0460126	.0721383	-0.64	0.524	-.1882932	.0962681
L10.	-.0714529	.0712459	-1.00	0.317	-.2119735	.0690676
L11.	-.0754686	.0709496	-1.06	0.289	-.2154048	.0644676
L12.	-.0567035	.0703604	-0.81	0.421	-.1954776	.0820706
_cons	2.706732	.572216	4.73	0.000	1.578132	3.835332

Figure 9: Regressing OP4INV5 over 12 lags (20 years without crises)

The last portfolio to be discussed, HIOPHIINV, showed significance for a lag of 3 months. However, as observed before, the rest of the lags don't show a significant relationship with the current value of the portfolio. In addition, the R^2 is yet again very low (5.4%) and the Prob>F value is high (0.5295), so in a similar way as in our preceding discussion we conclude that the model doesn't provide reliable results that can serve as a reference point for predictability assumptions.

Source	SS	df	MS			
				Number of obs	=	206
				F(12, 193)	=	0.92
Model	224.786103	12	18.7321753	Prob > F	=	0.5295
Residual	3936.30435	193	20.3953593	R-squared	=	0.0540
				Adj R-squared	=	-0.0048
Total	4161.09045	205	20.2980022	Root MSE	=	4.5161

hiophiinv	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
hiophiinv						
L1.	-.0259085	.0718141	-0.36	0.719	-.1675497	.1157328
L2.	-.0089215	.0717265	-0.12	0.901	-.15039	.132547
L3.	-.1536528	.0728733	-2.11	0.036	-.2973831	-.0099225
L4.	-.0396918	.0731031	-0.54	0.588	-.1838753	.1044917
L5.	.0490386	.0730998	0.67	0.503	-.0951385	.1932158
L6.	-.1148936	.0730472	-1.57	0.117	-.2589668	.0291797
L7.	.0668329	.0727485	0.92	0.359	-.0766513	.2103171
L8.	-.0368834	.0725054	-0.51	0.612	-.1798882	.1061213
L9.	-.103722	.0718048	-1.44	0.150	-.2453448	.0379009
L10.	.0097888	.0713774	0.14	0.891	-.1309912	.1505687
L11.	-.0418288	.0714332	-0.59	0.559	-.1827187	.099061
L12.	-.0572824	.0714115	-0.80	0.423	-.1981295	.0835647
_cons	2.077519	.5036528	4.12	0.000	1.084149	3.07089

Figure 10: Regressing HIOPHIINV over 12 lags (20 years without crises)

If we compare the results we obtained in this model without the data points for the 2007-2008 Financial Crises and the Covid pandemic with the previous one where these crises weren't excluded from our analysis, we see that there are still no grounds to talk about reliable predictability. In both models LOOPLOINV is unpredictable for any lag period. The other portfolios examined before in the model with crises, namely OP2INV2, OP3INV3, OP4INV5 and HiOPHiINV, show signs of predictability for certain lag periods in the adjusted model, but the R^2 and the Prob>F returned by STATA point at the inaccuracy of the results and suggest that further information and data needs to be involved in the analysis to make sure that we can better explain the variance in the dependent variable. In conclusion, for the model without the crises, we can say that possibly it could be used to talk about long term investments (ones over a couple of years) since in this way the effects of potential shocks to the market are less severe since most of them are temporary. However, given the data we analyzed this is still just a hypothesis and without further research can't be assumed to be true. Also, certain crises are very difficult to predict like the recent, Covid, one.

12 months

Another method to check the prediction of a portfolio's return is to use the data from the past 12 months and check whether the regression of the portfolio against the lastest months data (one lag) results in anything significant. Again, we are using the data from Kenneth French (25 Portfolios Formed on Operating Profitability and Investment (5 x 5), same section) and again our confidence level is 95% and significance level is 5%. Using the Breusch-Pagan test for heteroskedasticity results that we don't reject the initial hypothesis, implying homoskedasticity for all regressions. Therefore, we run the OLS regression of a portfolio against its first lag. Having done that for all 25 portfolios the results are:
4 out of 25 (16%) portfolios show predictability.
Predictable portfolios have highest R^2 so the variance of a portfolio's return is most explained by variables in the regression model in these cases.
The ones that are said to be predictable portfolios are OP2INV4, OP2INV5, OP3INV3 and OP4INV2.

To illustrate the results of our research, we will now describe what we obtained regarding the portfolios LoOPLoINV, OP2INV2, OP3INV3, OP4INV5 and HIOPHIINV, same as in a case of regression using the data from the last 20 years.
Since the regression has only one regressor, the p value and the Prob>F value are going to be the same. Therefore, the p-value determines whether we accept the model as a whole or not.

In the LoOPLoINV portfolio's case, the model is insignificant. The lag doesn't show a significant relationship with the current value of the portfolio, with a p-value of 0.684. The R^2 is very small (0.0193)and the adjusted R^2 is negative. Therefore, we conclude that the portfolio's returns are not predictable using this model.

```
reg looploinv l1.looploinv

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.18
       Model |  8.13920292         1  8.13920292   Prob > F        =    0.6839
    Residual |  413.911551         9  45.9901724   R-squared       =    0.0193
-------------+----------------------------------   Adj R-squared   =   -0.0897
       Total |  422.050754        10  42.2050754   Root MSE        =    6.7816

------------------------------------------------------------------------------
   looploinv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   looploinv |
         L1. |   .1399178   .3325942     0.42   0.684    -.6124625    .8922982
             |
       _cons |   2.097376   2.184433     0.96   0.362    -2.844154    7.038906
------------------------------------------------------------------------------
```

Figure 11: Regressing LoOPLoINV over 1 lag (1 month)

Portfolio OP2INV2 has 'better' results that the previously described portfolio. The p-value is 0.377, R^2 is higher (0.0876), however the adjusted R^2 is still negative. Again, this model cannot be said to be predictable given our confidence level.

```
reg op2inv2 l1.op2inv2

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.86
       Model |  9.64053465         1  9.64053465   Prob > F        =    0.3770
    Residual |  100.473846         9  11.1637607   R-squared       =    0.0876
-------------+----------------------------------   Adj R-squared   =   -0.0138
       Total |  110.114381        10  11.0114381   Root MSE        =    3.3412

------------------------------------------------------------------------------
     op2inv2 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     op2inv2 |
         L1. |   -.291228   .3133919    -0.93   0.377     -1.00017    .4177138
             |
       _cons |   1.255098   1.057592     1.19   0.266    -1.137341    3.647538
------------------------------------------------------------------------------
```

Figure 12: Regressing OP2INV2 over 1 lag (1 month)

The regression of the portfolio OP3INV3's returns looks much more promising. The coefficient of the first lag has a significant effect, given our assumptions. R^2 is much higher with a value of 0.3756 and the adjusted R^2 is positive. The p-value of the lag has a value of 0.045 and the interpretation of the coefficient is that given an increase in the lag by one unit, the current portfolio's return would decrease by -0.6435, ceteris paribus. The interesting thing to note is that the portfolio didn't show the predictability using the model with 12 lags and the data from the last 20 years.

```
reg op3inv3 l1.op3inv3

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      5.41
       Model |  55.5310948         1  55.5310948   Prob > F        =    0.0450
    Residual |  92.2971061         9   10.255234   R-squared       =    0.3756
-------------+----------------------------------   Adj R-squared   =    0.3063
       Total |  147.828201        10  14.7828201   Root MSE        =    3.2024

------------------------------------------------------------------------------
     op3inv3 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     op3inv3 |
         L1. |  -.6434944   .2765346    -2.33   0.045    -1.269059   -.0179297
             |
       _cons |   2.577195   1.046473     2.46   0.036     .2099086    4.944481
------------------------------------------------------------------------------
```

Figure 13: Regressing OP3INV3 over 1 lag (1 month)

The result of the regression of OP4INV5 doesn't give us reliable results when it comes to the predictability of returns of the portfolio. The p-value is almost 1 and the R^2 is 0 (adjusted R^2 negative). We simply reject this model for the given portfolio.

```
reg op4inv5 l1.op4inv5

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.00
       Model |  .006075949         1  .006075949   Prob > F        =    0.9890
    Residual |  273.589609         9  30.3988454   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =   -0.1111
       Total |  273.595685        10  27.3595685   Root MSE        =    5.5135

------------------------------------------------------------------------------
     op4inv5 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     op4inv5 |
         L1. |   .0045339   .3206933     0.01   0.989    -.7209247    .7299925
             |
       _cons |   .1019837   1.667197     0.06   0.953    -3.669478    3.873445
------------------------------------------------------------------------------
```

Figure 14: Regressing OP4INV5 over 1 lag (1 month)

Recalling that the portfolio's return was deemed to be predictable using the model with 12 lags and 20 years, we state that it isn't not predictable using this model. The p-value is 0.437, way higher than our significance level and the adjusted R^2 is negative.

```
reg hiophiinv l1.hiophiinv

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.66
       Model |  18.0131781         1  18.0131781   Prob > F        =    0.4366
    Residual |  244.648007         9  27.1831118   R-squared       =    0.0686
-------------+----------------------------------   Adj R-squared   =   -0.0349
       Total |  262.661185        10  26.2661185   Root MSE        =    5.2137

------------------------------------------------------------------------------
   hiophiinv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   hiophiinv |
         L1. |  -.2374718   .2917201    -0.81   0.437    -.8973885     .422445
             |
       _cons |   .2444155   1.579868     0.15   0.880    -3.329494    3.818325
------------------------------------------------------------------------------
```

Figure 15: Regressing HiOPHiINV over 1 lag (1 month)

The '12 months' model shows that more portfolio's return can be predictable, compared to other models. However, the model gives inconsistent results to the other models. Therefore, it is difficult to say whether it is actually predictable or not.

Conclusion

In conclusion, we can say that based on our research, it is inefficient to try to predict the returns of a portfolio at a given time period, based on data gathered over preceding time periods. The linear regression models we used returned low R^2 and mostly high Prob>F values, which are both indicative of the imprecision of the results. Whenever some predictability was established, it was either unreliable because of the overall inaccuracy of the regression (again, R^2 and Prob>F), or it was inconsistent and subjective with respect to the given portfolio; for different portfolios different lag periods turned out to be significant, so we couldn't establish a pattern to be applied in practice. Overall, our conclusion is that using the simple OLS regression model, we were unable to find any significant evidence/model, which would support our initial presumption that operating profitability (OP) and Investment (Inv) are determinants for the future of the firm's stock.

Sources

Leitner, Yaron. "Stock Prices and Business Investment." Federal Reserve Bank Philadelphia, Federal Reserve Bank Philadelphia, 18 Apr. 2007.

Sholichah, Fatmawati. "The Effects of Profitability and Solvability on Stock Prices: Empirical Evidence from Indonesia." Korea Science, Korea Distribution Science Association.

Chen, Qi, Itay Goldstein, and Wei Jiang. "Price Informativeness and Investment Sensitivity to Stock Price," Review of Financial Studies, 20:3 (2007), pp. 619-50.

Patton, M. (2022, April 14). How Stocks Performed During The Past 6 Recessions. Forbes. Retrieved May 28, 2022, from: https://www.forbes.com/sites/mikepatton/2020/03/30/how-stocks-performed-during-the-past-6-recessions/?sh=7743abb3339d

Cox, J. (2021, July 19). It's official: The Covid recession lasted just two months, the shortest in U.S. history. CNBC. Retrieved May 28, 2022, from: https://www.cnbc.com/2021/07/19/its-official-the-covid-recession-lasted-just-two-months-the-shortest-in-us-history.html

Kenneth French website:
`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`