Testing Market Efficiency: AR(5) and Market Model Analysis of IBM and the S&P 500

Michelle Silva, Ana Dolghiu, Siddhanth Narayanan

Schulich School of Business

FINE 3110

Professor Mark Kamstra

Table of Contents

INTRODUCTION

Prior research (e.g., Fama, 1970; Smith, 2007) found little autocorrelation in daily returns for large U.S. firms, supporting weak-form market efficiency (Eugene F. Fama.) This report examines the daily return data over a 10-year period (Dec 31, 2014 – Dec 31, 2024) for IBM and the S&P 500 Index to test this null hypothesis that markets are efficient and lagged returns do not significantly predict future returns.

MATERIALS & METHODS

IBM was selected due to its size, liquidity, and relevance among institutional investors. The S&P 500 was chosen as the market benchmark as it reflects the behavior of informed market participants and comprises large-cap firms, including IBM itself. We estimated AR(5) models for both and ran an OLS regression of IBM returns on S&P 500 returns to assess market influence. Daily data came from WRDS, with regressions run in Excel and Jupyter.

RESULTS

In contrast to the hypothesis that markets are efficient, the IBM AR(5) model shows that lag 1 and lag 5 are significant in predicting today's return. Lag 1's negative coefficient suggests a mean reversion, signifying changes in return tend to reverse the next day. Lag 5's positive coefficient suggests a slight weekly pattern. Other lags are insignificant and the low R^2 in **IBM's AR(5) model** *(Appendix A*) confirms minimal return predictability, consistent with market efficiency. With only lag 1 and 5 being significant, it indicates IBM's returns are mostly driven by external factors.

The S&P 500 AR(5) model (*Appendix B*) shows slightly stronger short-term patterns: lags 1 and 4 show mean reversion, and lag 2 suggests short-term momentum.

The **OLS regression** *(Appendix C)* shows that S&P 500 returns explain 39.4% of IBM's return variation ($R^2 = 0.394$). With a beta of 0.84, IBM moves with the market but with lower volatility, reflecting a defensive stock nature.

Appendix G shows that most spikes and crashes align with earnings reports which is consistent with the idea that markets incorporate available information quickly. For example, IBM jumped 9% in 2017 after beating Q3 expectations and rose 13.6% in 2024 following strong Q4 results driven by AI growth and rising margins. In 2020, both IBM and the S&P 500 experienced sharp volatility due to the COVID-19 crash and recovery. IBM also had down days tied to missed expectations, such as in October 2021 where it crashed 9.5% (Kilgore, Tomi.) An especially noteworthy finding is that while the market model explains approximately 39.4% of IBM's return variation, a substantial 60.6% is driven by firm-specific factors.(*Appendix C*) This could be due to challenges adapting to industry shifts like cloud computing, resulting in underperformance relative to peers which isn't reflected in the regression output.

CONCLUSION

Our findings broadly support weak-form market efficiency. The S&P 500 shows no significant autocorrelation in returns to IBM, consistent with the notion that prices adjust rapidly and incorporate available information. The AR(5) model for IBM shows only weak evidence of predictability at short lags - possibly due to firm-specific effects or market microstructure noise -but not to a degree that violates market efficiency meaningfully.

Appendix

Appendix A: Regression output for a simple AR(5) model – IBM stock

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.071963404							
R Square	0.005178731							
Adjusted R Square	0.003193853							
Standard Error	0.01497729							
Observations	2512							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	5	0.002926348	0.00058527	2.60909199	0.023197598			
Residual	2506	0.562143986	0.000224319					
Total	2511	0.565070333						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.000444121	0.000299493	1.482906922	0.138224885	-0.000143159	0.001031401	-0.000143159	0.001031401
X Variable 1	-0.041894247	0.01995658	-2.09926986	0.03589291	-0.081027326	-0.002761168	-0.081027326	-0.002761168
X Variable 2	0.02218959	0.019959225	1.111746066	0.266353991	-0.016948675	0.061327855	-0.016948675	0.061327855
X Variable 3	-0.00081059	0.019961195	-0.040608302	0.967611401	-0.039952718	0.038331538	-0.039952718	0.038331538
X Variable 4	-0.028423369	0.01995475	-1.424391168	0.154457779	-0.067552859	0.01070612	-0.067552859	0.01070612
X Variable 5	0.043076067	0.019947672	2.159453362	0.030909605	0.003960456	0.082191677	0.003960456	0.082191677

Appendix B: Regression output for simple AR(5) model – S&P 500 index

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.167337244							
R Square	0.028001753							
Adjusted R Square	0.026062411							
Standard Error	0.011079674							
Observations	2512							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	5	0.00886247	0.001772494	14.43879013	5.72098E-14			
Residual	2506	0.307634504	0.000122759					
Total	2511	0.316496974						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.000548938	0.000222342	2.468889746	0.013619544	0.000112945	0.000984931	0.000112945	0.000984931
X Variable 1	-0.127226889	0.019965493	-6.372338791	2.20818E-10	-0.166377446	-0.088076331	-0.166377446	-0.088076331
X Variable 2	0.060700957	0.02008027	3.022915396	0.002528883	0.021325333	0.10007658	0.021325333	0.10007658
X Variable 3	-0.015803096	0.0201072	-0.785942153	0.431975683	-0.055231527	0.023625335	-0.055231527	0.023625335
X Variable 4	-0.069657403	0.020072338	-3.470318373	0.000528628	-0.109017472	-0.030297333	-0.109017472	-0.030297333
X Variable 5	0.028085804	0.019959869	1.407113661	0.159517697	-0.011053723	0.067225331	-0.011053723	0.067225331

Appendix C: Regression output for a simple market data model for IBM and S&P 500

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.627929755							
R Square	0.394295777							
Adjusted R Square	0.394054941							
Standard Error	0.011674945							
Observations	2517							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	0.223156339	0.223156339	1637.191623	3.946E-276			
Residual	2515	0.342805438	0.000136304					
Total	2516	0.565961776						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.84444E-05	0.000232918	0.1221218	0.902812314	-0.000428287	0.000485176	-0.000428287	0.000485176
X Variable 1	0.838789372	0.020730184	40.46222464	3.946E-276	0.798139394	0.879439349	0.798139394	0.879439349

Appendix D: Python output for a simple AR(5) model – IBM stock

```
[3]: import pandas as pd
     import numpy as np
     import statsmodels.api as sm
     from statsmodels. tsa.ar_model import AutoReg, ar_select_order
     from statsmodels.tsa.api import acf, pacf, graphics
     data_path = r'C:\Users\michs\Downloads\qzw1gr7kaes9sc27 (1).csv'
     t_data = pd.read_csv(data_path)
     t_data.describe()
     y = t_data.RET
     mod = AutoReg(y, 5, old_names=False)
     res = mod.fit()
     print(res.summary())
```

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:                    RET   No. Observations:                 2517
Model:                     AutoReg(5)   Log Likelihood                6992.095
Method:               Conditional MLE   S.D. of innovations              0.015
Date:                Tue, 30 Sep 2025   AIC                         -13970.190
Time:                        18:11:07   BIC                         -13929.388
Sample:                             5   HQIC                        -13955.381
                                 2517
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0004      0.000      1.485      0.138      -0.000       0.001
RET.L1        -0.0419      0.020     -2.102      0.036      -0.081      -0.003
RET.L2         0.0222      0.020      1.113      0.266      -0.017       0.061
RET.L3        -0.0008      0.020     -0.041      0.968      -0.040       0.038
RET.L4        -0.0284      0.020     -1.426      0.154      -0.067       0.011
RET.L5         0.0431      0.020      2.162      0.031       0.004       0.082
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.4040           -1.0414j            1.7481           -0.3984
AR.2           -1.4040           +1.0414j            1.7481            0.3984
AR.3            0.7192           -1.7962j            1.9349           -0.1894
AR.4            0.7192           +1.7962j            1.9349            0.1894
AR.5            2.0293           -0.0000j            2.0293           -0.0000
-----------------------------------------------------------------------------
```

Appendix E: Python output for a simple AR(5) model – S&P 500 index

```python
import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels. tsa.ar_model import AutoReg, ar_select_order
from statsmodels.tsa.api import acf, pacf, graphics
data_path = r'C:\Users\michs\Downloads\qzw1gr7kaes9sc27 (1).csv'
t_data = pd.read_csv(data_path)
t_data.describe()
y = t_data.sprtrn
mod = AutoReg(y, 5, old_names=False)
res = mod.fit()
print(res.summary())
```

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:                 sprtrn   No. Observations:                 2517
Model:                     AutoReg(5)   Log Likelihood                7749.269
Method:               Conditional MLE   S.D. of innovations              0.011
Date:                Tue, 30 Sep 2025   AIC                         -15484.538
Time:                        18:17:57   BIC                         -15443.737
Sample:                             5   HQIC                        -15469.729
                                 2517
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0005      0.000      2.472      0.013       0.000       0.001
sprtrn.L1     -0.1272      0.020     -6.380      0.000      -0.166      -0.088
sprtrn.L2      0.0607      0.020      3.027      0.002       0.021       0.100
sprtrn.L3     -0.0158      0.020     -0.787      0.431      -0.055       0.024
sprtrn.L4     -0.0697      0.020     -3.474      0.001      -0.109      -0.030
sprtrn.L5      0.0281      0.020      1.409      0.159      -0.011       0.067
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.3614           -0.9689j            1.6710           -0.4016
AR.2           -1.3614           +0.9689j            1.6710            0.4016
AR.3            1.0909           -1.7408j            2.0544           -0.1609
AR.4            1.0909           +1.7408j            2.0544            0.1609
AR.5            3.0212           -0.0000j            3.0212           -0.0000
```

Appendix F: Output for a simple market data model for IBM

[6]:

OLS Regression Results

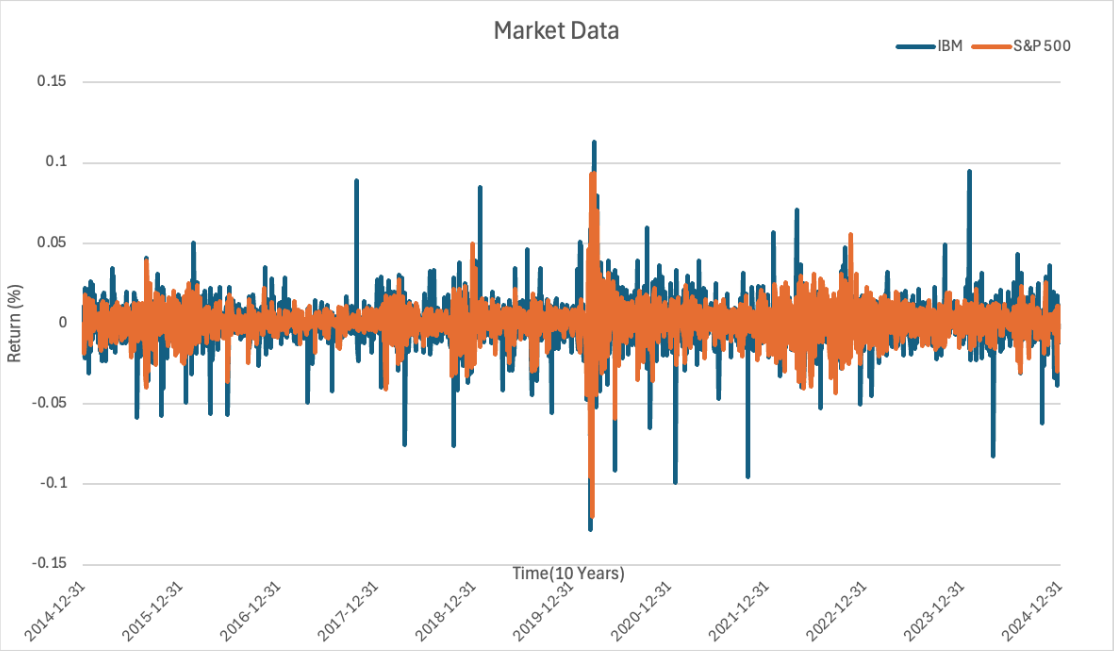
Dep. Variable:	RET	**R-squared:**	0.394
Model:	OLS	**Adj. R-squared:**	0.394
Method:	Least Squares	**F-statistic:**	1637.
Date:	Tue, 30 Sep 2025	**Prob (F-statistic):**	3.95e-276
Time:	18:32:25	**Log-Likelihood:**	7631.0
No. Observations:	2517	**AIC:**	-1.526e+04
Df Residuals:	2515	**BIC:**	-1.525e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.844e-05	0.000	0.122	0.903	-0.000	0.000
sprtrn	0.8388	0.021	40.462	0.000	0.798	0.879

Omnibus:	622.593	**Durbin-Watson:**	1.979
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	22014.670
Skew:	-0.450	**Prob(JB):**	0.00
Kurtosis:	17.460	**Cond. No.**	89.1

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Appendix G : Market Data line plot



Sources

1) Eugene F. Fama. “Efficent Capital Markets: II” The *Journal of Finance,* December 1991, http://www-2.rotman.utoronto.ca/~kan/3032/pdf/TestsOfMarketEfficiency/Fama_JF_1991.pdf
2) Wharton Research Data Services. “Wharton Research Data Services.” *WRDS*, 2024, wrds-www.wharton.upenn.edu
3) Kilgore, Tomi. “IBM’s Stock Surge Is One for the History Books.” *MarketWatch*, 19 Oct. 2017. www.marketwatch.com/story/ibms-stock-surge-is-one-for-the-history-books-2017-10-18. Accessed 7 Oct. 2025.
4) Python : anaconda, jupyter lab : python (conda.env.base)