

Regresja zmian cen akcji

Karol Oleszek

10 maja 2020

Spis treści

1	Wstęp	3
2	Cel projektu	4
2.1	Model wyboru akcji do celów inwestycyjnych	4
2.2	Zbadanie zależności pomiędzy zmianą cen, a informacjami finansowymi	4
3	Opis danych	5
3.1	Zbiór danych	5
3.2	Usuwanie braków danych	5
3.3	Transformacja zmiennej kategorycznej	5
3.4	Zmienne objaśniające	5
3.5	Rozkład zmiennej objaśnianej	5
3.6	Korelacja	6
4	Dobór zmiennych do modelu	7
5	Wybór postaci modelu	7
6	Weryfikacja statystyczna modelu	7
7	Prognoza	7
8	Interpretacja	7
9	Podsumowanie	7
10	Spis tabel	8
11	Spis rysunków	9
12	Literatura	10

1 Wstęp

Przewidywanie zmian cen akcji oraz innych instrumentów finansowych znajduje się w centrum zainteresowania inwestorów. Zmiany cen są podstawowym zjawiskiem powodującym bogacenie się lub ubożenie inwestora indywidualnego bądź instytucjonalnego, dlatego też próby zrozumienia i opisanie reguł rządzących tym zjawiskiem są kluczowe dla podejmowania skutecznych decyzji o alokacji kapitału.

Teoria rynków kapitałowych proponuje wiele różnych wyjaśnień zmienności cen: hipoteza rynku efektywnego (Bachelier [1]) zakłada, że ceny rynkowe akcji w danej chwili odzwierciedlają wszystkie dostępne informacje o spółce; autorzy i zwolennicy hipotezy krótkoterminową zmienność cen opisują jako losowy ruch wokół efektywnej wartości. Hipoteza rynku efektywnego znalazła wielu zwolenników, którzy poddawali w wątpliwość samą zasadność przewidywania cen (Cowles [4]), jak również zainspirowała powstanie indeksowych funduszy inwestycyjnych.

Hipoteza rynku efektywnego spotkała się z szeroką krytyką ze strony ekonomistów i inwestorów giełdowych, którzy wskazywali na kontrprzykłady obalające hipotezę. Współcześnie właściwie wszystkie duże organizacje finansowe używają różnego rodzaju systematycznych narzędzi do analizy i prognozy zmian cen na rynkach kapitałowych (Graham Capital Management [5]). Duże oraz wciąż rosnące znaczenie ma też algorytmiczny handel (Capgemini [2]).

Do złożonego problemu jakim jest symulacja i prognozyka zachowania rynków kapitałowych stosuje się bardzo szeroki wachlarz metod statystycznych, algorytmicznych i ekonometrycznych. Duże zastosowanie mają metody uczenia maszynowego (Shunrong Shen [6]), w tym głębokie sieci neuronowe o niekonwencjonalnych architekturach. Ponadto do prognozyki coraz częściej używa się analizy języka naturalnego (Zhaoxia Wang [7]).

Poniższa praca zawiera przekrojową regresję zmian cen akcji na rynku amerykańskim w 2019 z wykorzystaniem standardowych narzędzi ekonometrycznych. Zbiór danych służący do konstrukcji modelu zawiera dane z roku 2018, dotyczące sytuacji finansowych, kapitałowych i operacyjnych spółek, zawarte w formie wskaźników i pozycji ze sprawozdań finansowych.

2 Cel projektu

2.1 Model wyboru akcji do celów inwestycyjnych

Celem projektu jest wyznaczenie bazowego poziomu efektywności wyboru spółek, których akcje w nadchodzącym roku zyskają na wartości. Za wybór odpowiadał będzie model, który powstanie przy użyciu metody najmniejszych kwadratów i który będzie mógł służyć jako punkt odniesienia do badania efektywności innych metod predykcji.

Efektywność prognostyczna modelu zostanie zbadana przy użyciu *średniego błędu prognozy ex post*, danego wzorem:

$$ME = \frac{1}{s} \sum_{t=1}^s (y_t - y_t^P)$$

,

Gdzie:

s - ilość obserwacji w testowym zbiorze danych

y_t - prawdziwa wartość zmiennej objaśnianej

y_t^P - prognozowana wartość zmiennej objaśnianej

2.2 Zbadanie zależności pomiędzy zmianą cen, a informacjami finansowymi

Ponadto model posłuży do oceny wpływu informacji finansowych zawartych w publicznie dostępnych źródłach na przyszłą wartość spółek giełdowych. Ocena ta może być użyteczna przy podejmowaniu decyzji o tym, jakie dane zbierać na temat spółek w celu skutecznego przewidywania ich przyszłej wyceny.

Miarą tej oceny będzie współczynnik determinacji R^2 , dany wzorem:

$$R^2 = \frac{\sum_{t=1}^n (y_t^P - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

,

Gdzie:

n - ilość obserwacji w uczącym zbiorze danych

y_t - prawdziwa wartość zmiennej objaśnianej

y_t^P - prognozowana wartość zmiennej objaśnianej

\bar{y} - średnia arytmetyczna zmiennej objaśnianej

3 Opis danych

3.1 Zbiór danych

Zbiór danych użyty w projekcie pochodzi z internetowej platformy Kaggle (Carbone [3]). Zawiera on zmienną objaśnianą Y - procentową zmianę ceny akcji danej spółki w 2019 roku, oraz zmienne objaśniające $X_i, i = 1 \dots k$ - $k-1$ wskaźników finansowych i pozycji z formularza $10-K$ ¹, a także zmiennej kategorycznej oznaczającej sektor gospodarki rozważanej spółki.

3.2 Usuwanie braków danych

Dane zostały zebrane przy użyciu interfejsu programistycznego *Financial Modeling Prep API* i zawierały pewne braki wynikające z różnic w dokumentach źródłowych. Dla celów analizy usunięte zostały wszystkie obserwacje, w których brakowało więcej niż 50 wartości oraz wszystkie zmienne, w których co najmniej 10% obserwacji nie miało przypisanej wartości. Po tej transformacji, w zbiorze danych pozostało 4122 obserwacji oraz 179 zmiennych (4392 x 222, 9,98% braków przed transformacją). Wciąż brakujące 0,82% wartości zostało zastąpionych średnimi arytmetycznymi odpowiednich zmiennych.

3.3 Transformacja zmiennej kategorycznej

Kategoryczna zmienna objaśniająca *Sector*, która przyjmowała 11 różnych wartości (Consumer Cyclical, Energy, Technology, Industrials, Financial Services, Basic Materials, Communication Services, Consumer Defensive, Healthcare, Real Estate, Utilities) została przekształcona na 10 zmiennych zero-jedynkowych. Po tej operacji zbiór danych składał się ze 188 zmiennych.

3.4 Zmienne objaśniające

fdasfa

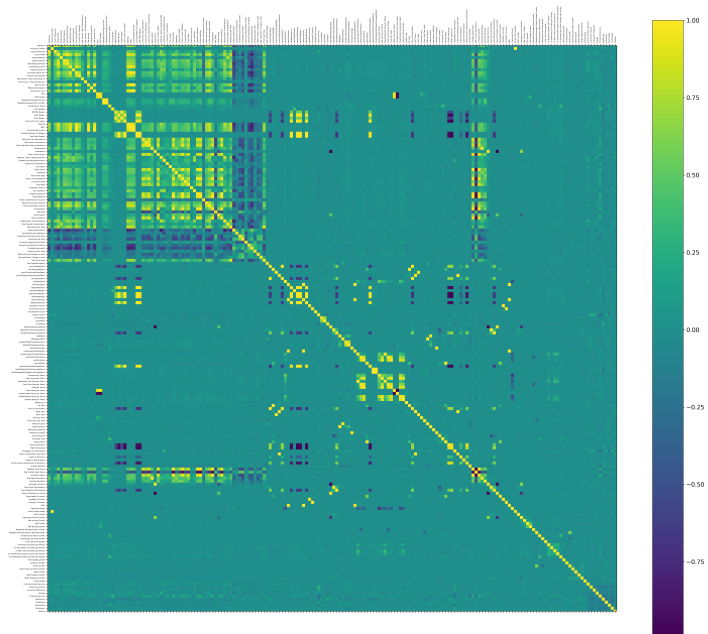
3.5 Rozkład zmiennej objaśnianej

fdasfads

¹Form $10-K$ jest to coroczne podsumowanie finansowe składane przez amerykańskie spółki giełdowe do *U.S. Securities and Exchange Commission*, federalnej agencji nadzoru finansowego.

3.6 Korelacja

To jest tekst o korelacji. fda



Rysunek 1: Macierz korelacji

4 Dobór zmiennych do modelu

fadsfda

5 Wybór postaci modelu

6 Weryfikacja statystyczna modelu

7 Prognoza

8 Interpretacja

9 Podsumowanie

10 Spis tabel

Spis tabel

11 Spis rysunków

Spis rysunków

1	Macierz korelacji	6
---	-----------------------------	---

12 Literatura

References

- [1] Louis Bachelier. *Théorie de la Spéculation*. 1900.
- [2] Caggemini. *High Frequency Trading: Evolution and the Future*. 2012.
- [3] Nicolas Carbone. *200+ Financial Indicators of US stocks (2014-2018)*. 2019.
- [4] Alfred Cowles. *Can stock market forecasters forecast?* 1932.
- [5] L.P. Graham Capital Management. *Systematic Global Macro: Performance, Risk and Correlation Characteristics*. 2013.
- [6] Tongda Zhang Shunrong Shen Haomiao Jiang. *Stock Market Forecasting Using Machine Learning Algorithms*. 2012.
- [7] Zhiping Lin Zhaoxia Wang Seng-Beng Ho. *Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment*. 2018.