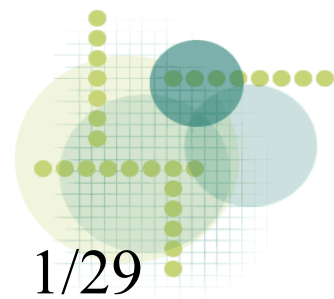


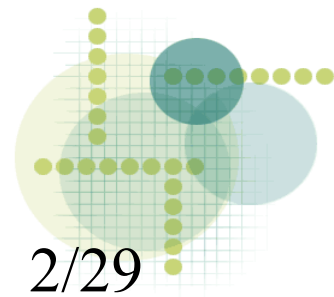
Analiza zależności między notowaniami akcji firm a artykułami prasowymi

Mateusz Kobos, 23.04.2008
Seminarium Metody Inteligencji Obliczeniowej



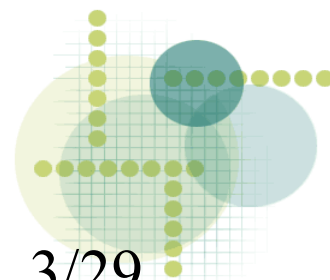
Spis treści

- Słowa najczęściej występujące wokół ekstremalnych zwrotów
- Analiza artykułów występujących wokół ekstremalnych zwrotów
- Czasowy związek między liczbą publikacji a wolumenem
- Implementacja eksperymentu opisanego w literaturze



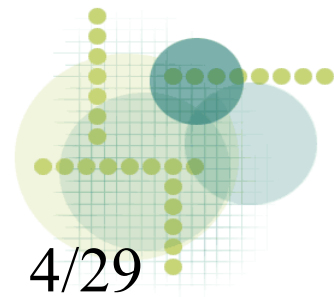
Słowa najczęściej występujące wokół ekstremalnych zwrotów

- parametry eksperymentu
 - badanie publikacji wokół 5 ekstremalnych zwrotów dziennych firmy MSFT
 - okienka czasowe: 2-15 dni
 - okienko obejmuje: a) dni przed publikacją, b) dni otaczające publikacje
 - dla porównania te same eksperymenty przeprowadzono dla 5 losowo wybranych zwrotów
- wnioski
 - brak wyróżniających się słów



Słowa występujące wokół pozytywnych i negatywnych ekstremalnych zwrotów

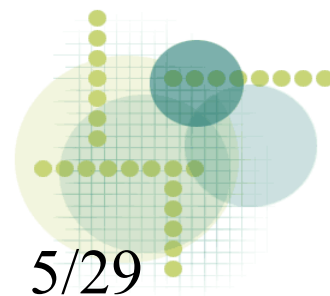
- Badano częściej występujące słowa, które występują tylko przy ekstremach:
 - tylko pozytywnych
 - tylko negatywnych
 - przy pozytywnych i negatywnych
- parametry eksperymentu:
 - notowania firmy MSFT
 - okienko czasowe: "od przeszłości do teraz"



Słowa występujące wokół pozytywnych i negatywnych ekstremalnych zwrotów



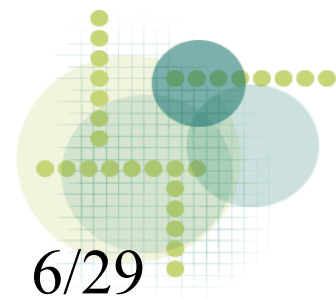
- parametry eksperymentu
 - badano:
 - 20 pierwszych ekstremów, okienko $[-1, 0]$
 - 20 pierwszych ekstremów, okienko $[0, 0]$
 - 5% pierwszych ekstremów (po 125 elementów), okienko $[-2, 0]$
 - ekstrema do „przejęcia” histogramu wielkości ekstremów (16 ekstremów), okienko $[-2, 0]$



Słowa występujące wokół pozytywnych i negatywnych ekstremalnych zwrotów



- Wnioski
 - dużo słów, które kojarzą się pozytywnie, jest przypisanych do klasy negatywnej i odwrotnie (np. słowo "increas")
 - brak spójnych zależności
 - widać, że duża część nieznaczących słów jest odsiewana
- Pomysły:
 - rozpoznawać całe wyrażenia (np. predefiniowane) z uwzględnieniem negacji



Analiza artykułów występujących wokół ekstremalnych zwrotów

- Parametry: dzienne zwroty firmy MSFT
 - jedno z 6 ekstremów (2000-11-30) powstało dlatego, że cały rynek spadł (nie było żadnego specyficznego artykułu o MSFT)
 - spadek indeksu/rynku był długofalowy (parę dni), więc nie można wyeliminować tej anomalii za pomocą normalizacji dziennych zwrotów
 - pomysł: uwzględniać długofalowe zwroty
- Parametry: dzienne zwroty firmy CXTS
 - badano 4 ekstrema (2 pozytywne, 2 negatywne)
 - 2 nie dało się przewidzieć za pomocą wcześniej opublikowanych artykułów
 - 1 raczej nie dało się przewidzieć (umowa między MSFT i CXTS ale bez wyraźnego pozytywnego nastawienia autora) 7/29



Związek czasowy między liczbą publikacji a wolumenem

- Parametry eksperymentu:
 - notowania 9 firm różnych firm
 - przedział czasowy: 10 lat

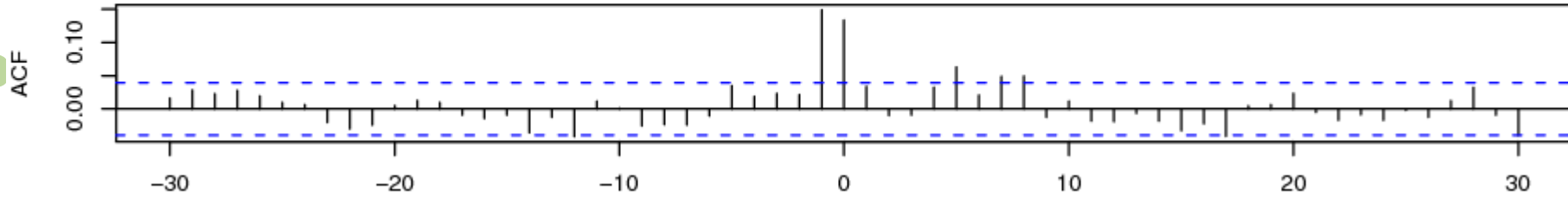
symbol	returns_min	returns_max	returns_sd	texts_no	texts_day_avg
NOVL	-0.39	0.33	0.041	92	0.04
SUNW	-0.27	0.3	0.037	384	0.15
ERTS	-0.26	0.2	0.033	100	0.04
SYMC	-0.37	0.28	0.037	77	0.03
SCUR	-0.48	0.44	0.056	6	0
MSFT	-0.16	0.2	0.022	2738	1.09
AAPL	-0.52	0.33	0.035	603	0.24
VOLV	-0.12	0.18	0.021	106	0.04
YHOO	-0.22	0.24	0.043	496	0.2

Związek czasowy między liczbą publikacji a wolumenem

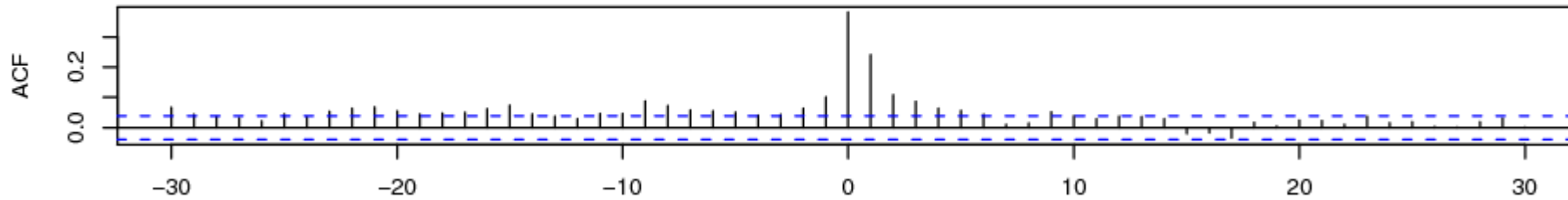
- Badano m.in. czasowe zależności:
 - liczba artykułów a wolumen
 - liczba artykułów a modułem zwrotu
 - autokorelacje liczby artykułów, wolumenu

Przykładowe wyniki

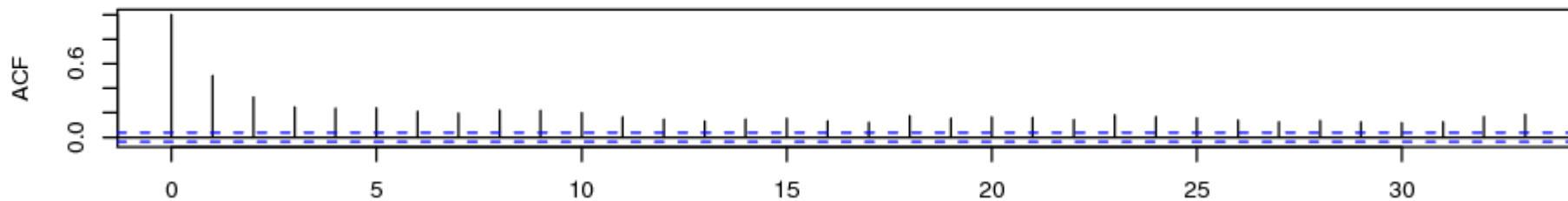
SUNW: articles_no vs volume lag



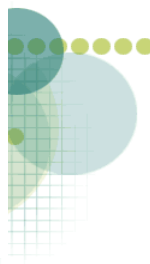
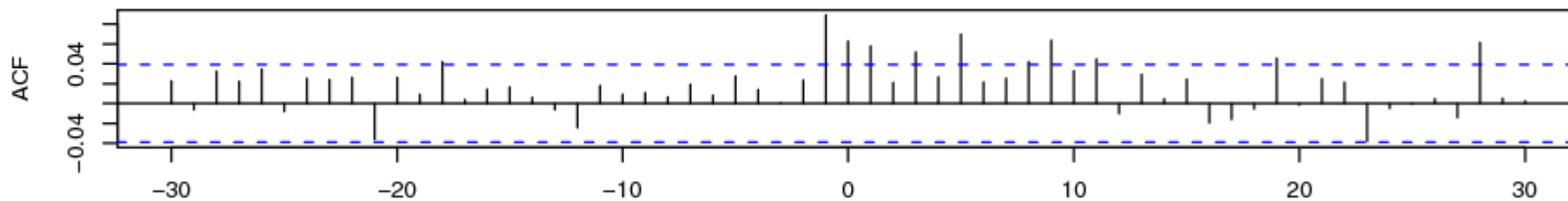
SUNW: volume vs abs(return) lag



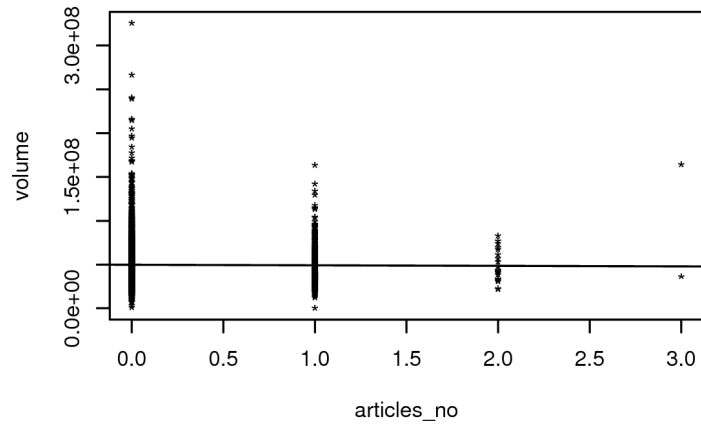
SUNW: volume ACF lags



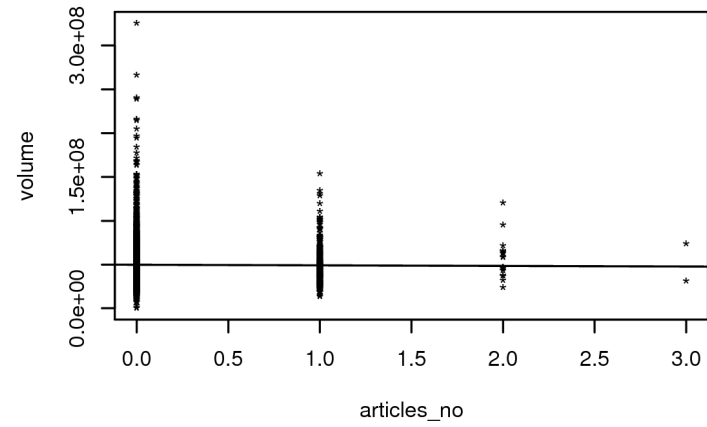
SUNW: articles_no vs abs(return) lag



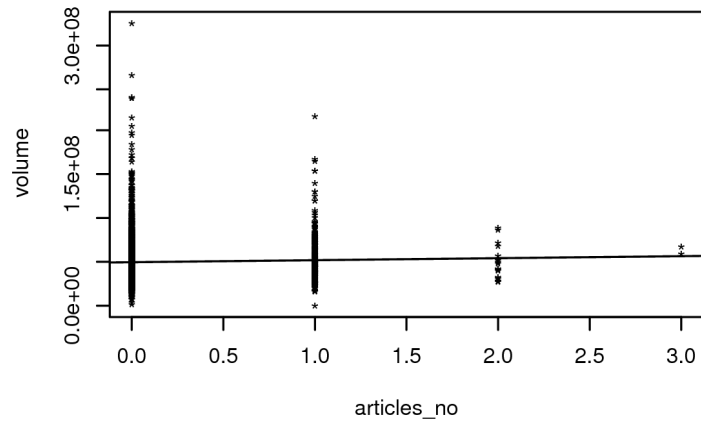
SUNW: articles_no vs. volume lag=-3



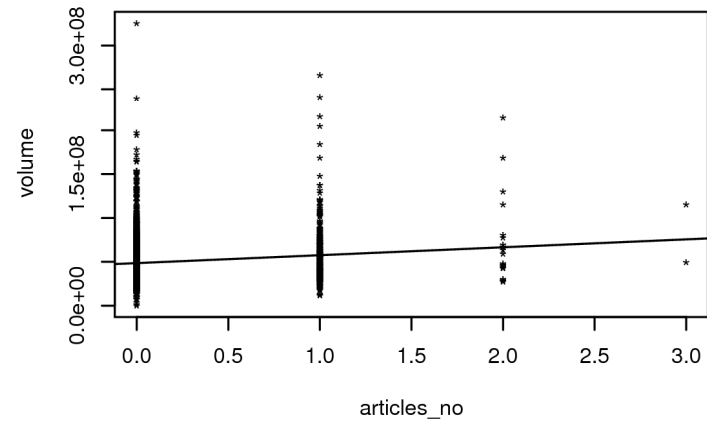
SUNW: articles_no vs. volume lag=-2



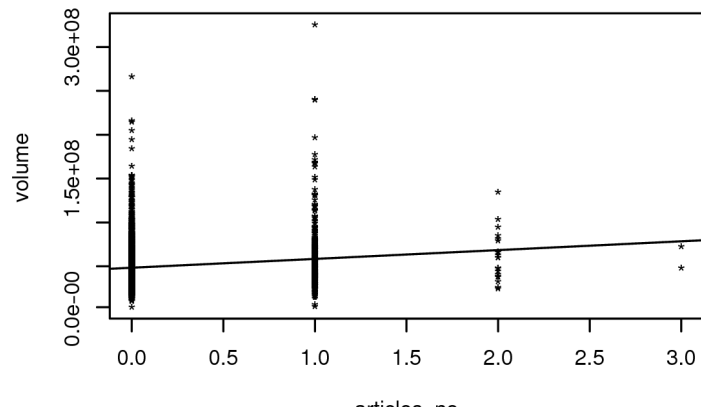
SUNW: articles_no vs. volume lag=-1



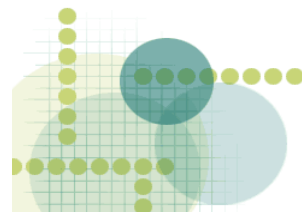
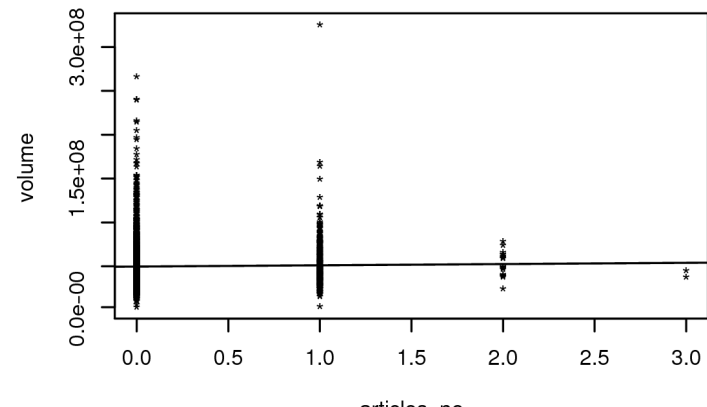
SUNW: articles_no vs. volume lag=0

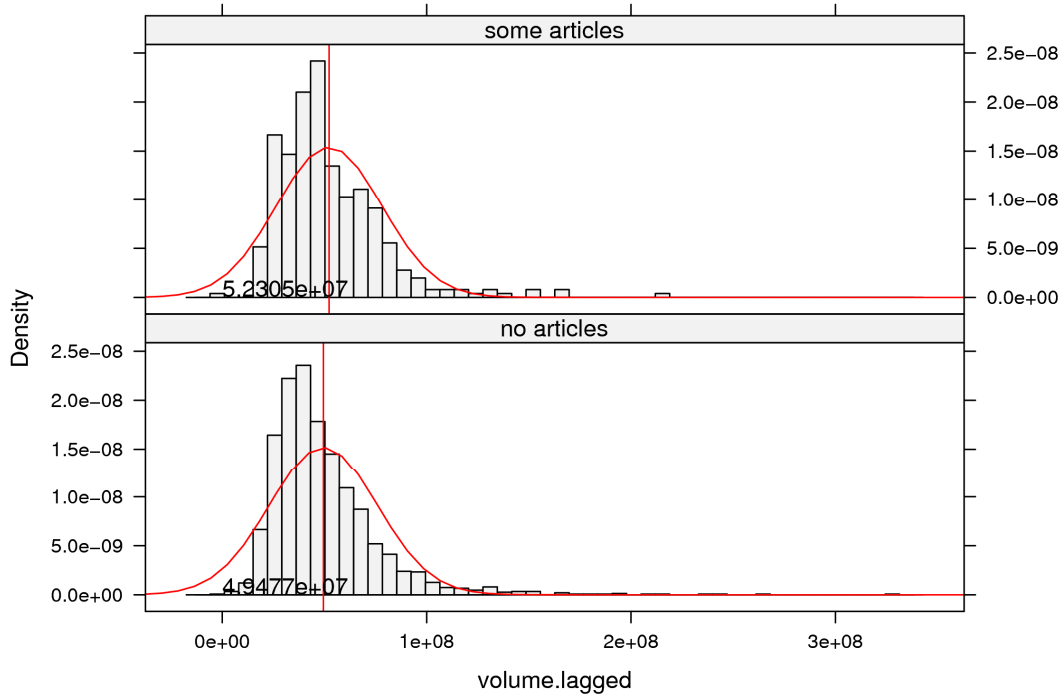
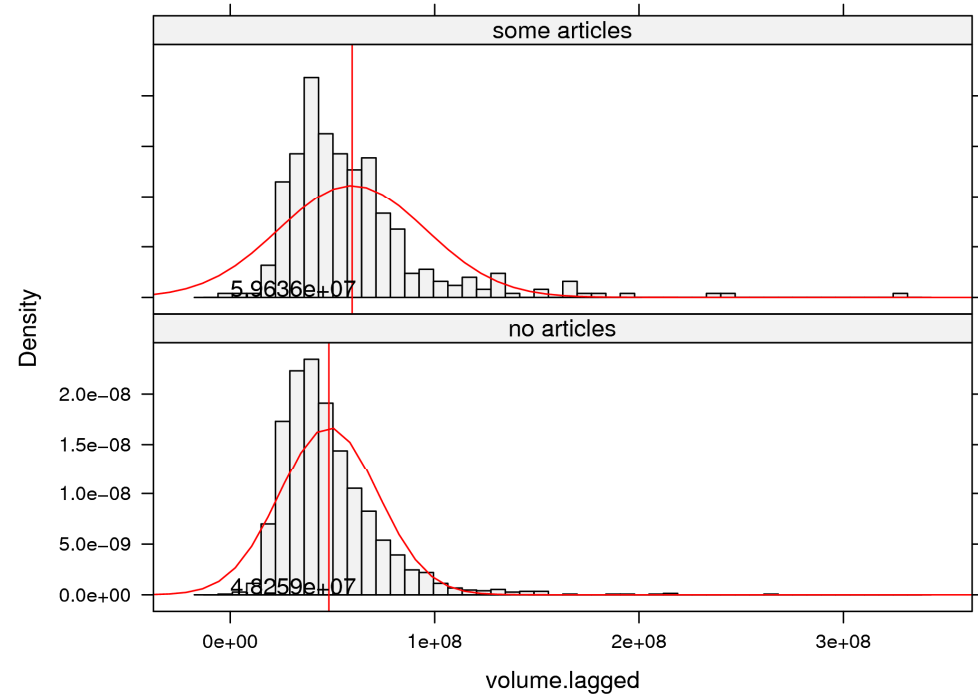
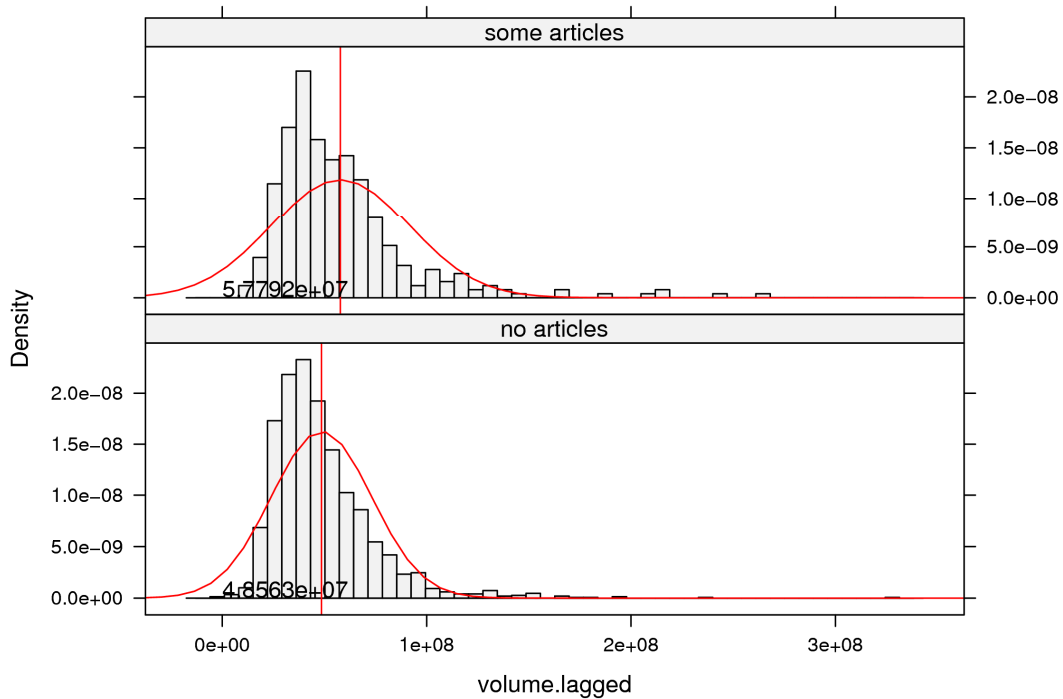
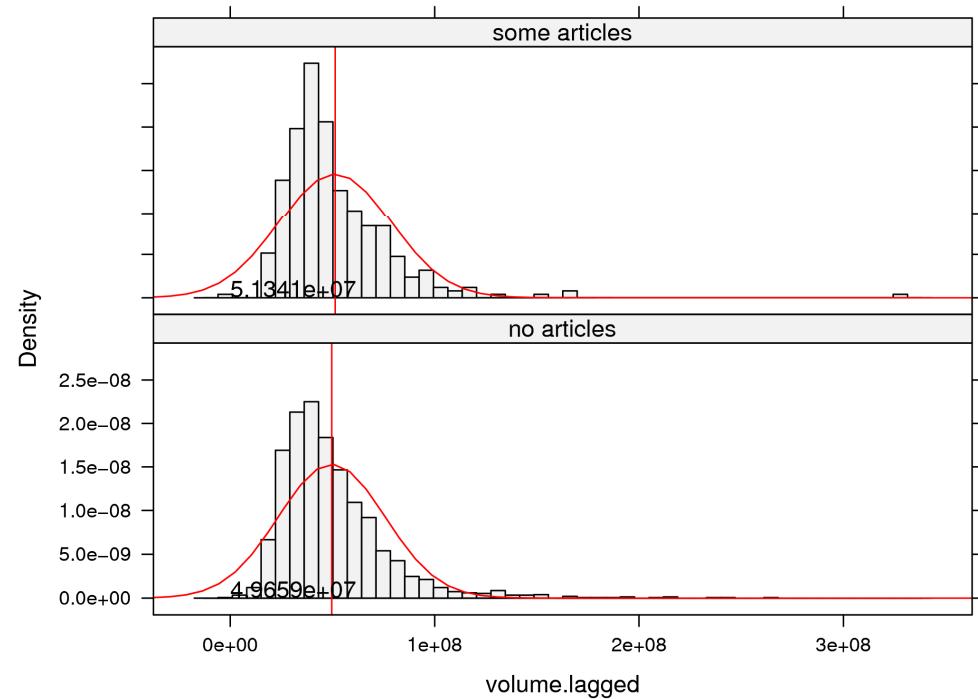


SUNW: articles_no vs. volume lag=1



SUNW: articles_no vs. volume lag=2



SUNW: volume lag=-1**SUNW: volume lag=1****SUNW: volume lag=0****SUNW: volume lag=2**

Związek czasowy między liczbą publikacji a wolumenem

- Obserwacje
 - występuje zależność między l. artykułów a dzisiejszym i jutrzejszym wolumenem
 - dla 5/9 spółek jest wyraźnie widoczna zależność
 - dla 2/9 spółek jest słabo widoczna zależność
 - AAPL
 - MSFT - przyczyną może być b. duża liczba artykułów.
 - dla 2/9 spółek brak zależności
 - VOLV
 - SCUR - przyczyną może być b. mała liczba artykułów
 - dość silna autokorelacja wolumenu

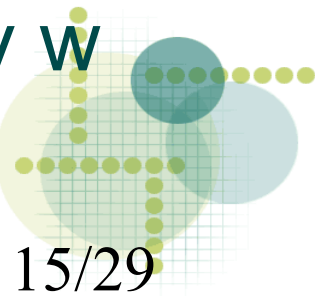


Związek czasowy między liczbą publikacji a wolumenem

- Wnioski
 - nie wiadomo jak bezpośrednio zastosować wiedzę i jutrzejszym wolumenie do przewidywania jutrzejszego zwrotu

Implementacja eksperymentu opisanego w literaturze

- artykuł: Mahesh Pritamani, Vijay Singal, „**Return predictability following large price changes and information releases**”, Journal of Banking and Finance, p. 631-656, 2001
 - dokładny opis eksperymentu
 - dobre wyniki
- idea: dla danej firmy określenie pewnych dat jako ważnych, zwiastujących duże zwroty w przyszłości (horyzont 20 dni)



Warunki eksperymentu

	artykuł	mój eksperyment
przedział czasowy	1990-1992	2004-05-03:2007-04-30
dane tekstowe	Wall Street Journal, Dow Jones News Wire	Wall Street Journal
notowania	dokładne logi (?) transakcji z Institute for the Study of Security Markets (ale nie jest wykorzystywany cały potencjał tych danych)	dzienne ceny pochodzące z finance.yahoo.com
skąd pochodzą spółki?	giełdy: NYSE, AMEX	indeks: NYSE Composite
normowanie zwrotu i wolumenu firmy	value-weighted CRSP NYSE/AMEX index	indeks: NYSE Composite

Interpretacja kluczowy określeń

- „public announcement”, “release of news”, “news report”
 - zakładam: publikacja artykułu na temat firmy
- „adjusted return” (zwrot unormowany względem indeksu)
 - *zwrot z akcji – zwrot z indeksu*
- „daily return”
 - zakładam: dla dnia nr n :
 - *cena zamknięcia(n) – cena zamknięcia($n-1$)*
 - w artykule:
 - „returns are calculated at mid-point of closing bid-ask prices”



Podstawowy wybór danych

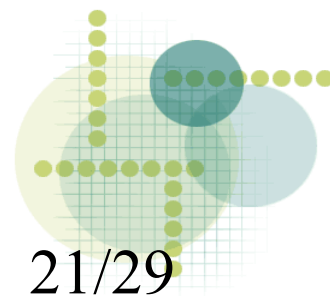
- wybór spółek z określonego przedziału czasowego
 - artykuł: giełdy NYSE/AMEX
 - eksperyment: indeks NYSE Composite (1709 spółek)
- wybór wydarzeń
 - wydarzenie („event”) - dzień, w którym unormowany dzienny zwrot znajduje się więcej niż 3 odchylenia standardowe od średniej unormowanych zwrotów z 250 ostatnich dni dla danej spółki
 - artykuł: korzystają z „abnormal return” unormowanego względem indeksu (23459 wydarzeń)
 - eksperyment: zwrot unormowany względem indeksu (10720 wydarzeń)

Wydarzenia - opis

- Na wydarzenia są nakładane kolejne filtry („screens”) eliminujące część z nich
- w opisie znajduje się liczba wydarzeń pozostałych po zastosowaniu kolejnego filtra podana w artykule i uzyskana w eksperymencie
- Wydarzenie jest pozytywne, jeśli *unormowany zwrot* > 0
 - w horyzoncie czasowym 20 dni należy spodziewać się dodatniego zwrotu
 - Analogicznie dla wydarzenia negatywnego

Etapy filtrowania („screening”)

- brak wydarzeń w ciągu ostatnich 20 dni
 - artykuł: 13000 wydarzeń
 - eksperyment: 7934 wydarzeń
- brak split-ów, dywidend lub „equity issues” (?) w ciągu ostatnich 60 dni i dzisiaj
 - artykuł: 10689 wydarzeń
 - eksperyment: 3520 wydarzeń
- *cena zamknięcia dzisiaj > 10\$*
 - artykuł: 6127 wydarzeń
 - eksperyment: 2945 wydarzeń



Etapy filtrowania („screening”)

- firma „handlowana” w ciągu ostatnich 60 dni
 - artykuł: 4886 wydarzeń
 - eksperyment: 2850 wydarzeń (*wolumen* > 0 w ciągu ostatnich 60 dni)
- Wyrzucenie spółek z dolnego kwintyla pod względem wielkości spółki
 - artykuł: ubyło tylko 13 wydarzeń
 - eksperyment: nie zaimplementowane

Etapy filtrowania („screening”)

- unormowany zwrot wolumenu znajduje się w 90. percentylu dla rozkładu za ostatnie 60 dni
 - eksperyment: 2057 wydarzeń
- publikacja artykułu na temat firmy dzisiaj
 - artykuł: 1/3 wydarzeń spełnia ten warunek
 - eksperyment: 8 wydarzeń, ale brano pod uwagę tylko 100 spółek

Etapy filtrowania („screening”)

- przypisanie artykułu do jednej z 7 klas tematycznych. Nie odrzucamy wydarzenia, jeśli artykuł należy do jednej z klas: 1. „earnings announcement by management”, 2. „forecast of earnings by management”, 3. „analyst recommendations”
 - eksperyment: nie zaimplementowane

Wyniki eksperymentu

- zwroty 20-dniowe (wyniki z artykułu są statystycznie istotne)
 - bez filtra związanego z artykułami (eksperyment: 1978 wydarzeń)

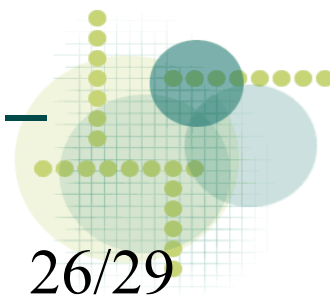
	adjusted_sell	adjusted_buy	raw_sell	raw_buy
sum	-0.6400	4.7172	9.7369	16.2638
avg	-0.0008	0.0041	0.0117	0.0142

- z filtrem liczby artykułów (eksperyment: 8 wydarzeń (ale brano pod uwagę tylko 100 spółek); artykuł: 1256 wydarzeń)

	adjusted_sell	adjusted_buy	raw_sell	raw_buy
sum	0.0092	-0.0400	-0.1042	0.0287
avg	0.0046	-0.0067	-0.0521	0.0048
artykuł: avg	-0.0168	0.0198		

Wyniki eksperymentu - komentarz

- eksperyment „bez filtra związanego z artykułami”
 - małe średnie znormalizowane zwroty (ale czy statystycznie istotne?)
 - średni zwykły zwrot dla wydarzeń negatywnych (sell) jest dodatni i zbliżony dla wartości dla wydarzeń pozytywnych (buy)!
- eksperyment „z filtrem liczby artykułów”
 - średni zwykły zwrot dla wydarzeń negatywnych (sell) jest dodatni i a dla pozytywnych (buy) ujemny!
 - liczba wydarzeń jest b. mała liczba przykładów – trudno wyciągać wnioski z takiego zb. danych



Wyniki z artykułu

- Zwroty 20-dniowe po zastosowaniu wszystkich filtrów (882 wydarzeń)

klasa artykułu	avg(abnormal_sell)	avg(abnormal_buy)
1	-0.0260	0.0344
2	-0.0196	0.0416
3	-0.0239	0.0036

- Zwroty 20-dniowe na innym przedziale czasowym (1993-1994) (762 wydarzeń)

koszty transakcji?	avg(abnormal_sell)	avg(abnormal_buy)
Nie	-0.0240	0.0218
Tak	-0.0151	0.0104

Pomysły

- Sprawić, by po zastosowaniu filtrów zostawało więcej wydarzeń (słabsze filtry, lub więcej spółek (najlepiej wszystkie pochodzące z indeksu))
- Normować zwroty zgodnie z opisem w artykule



Dziękuję za uwagę!

