

MARCIN NIEDUŻAK, MATEUSZ PIPIEŃ

EKONOMETRYCZNA ANALIZA PRAWDOPODOBIENSTWA ZAWARCIA
TRANSAKCJI WYNIKAJĄCYCH Z NAPŁYWU INFORMACJI
– WPŁYW ZAŁOŻEŃ CO DO ROZKŁADU STÓP ZWROTU NA ZMIENNOŚĆ
MIARY VPIN

1. WSTĘP

Badanie współzależności pomiędzy mechanizmem zawierania transakcji a płynnością, zmiennością cen czy wolumenem obrotów jest zagadnieniem wspólnie silnie rozwijanym. Wiedza na temat mikrostruktury rynku jest ważna z praktycznego punktu widzenia. Na proces kształtowania się cen rynkowych i obrotów mają wpływ preferencje inwestorów, ich stosunek do ryzyka, czy swobodny dostęp do informacji (por. Doman, 2011). W badaniach często przyjmuje się założenie, że inwestorzy giełdowi obserwując zachowania innych uczestników rynku pozyskują informacje na temat notowanej spółki. Liczba składanych zleceń kupna i sprzedaży może więc odzwierciedlać reakcję na pojawienie się nowych sygnałów informacyjnych. Rozważane w literaturze modele teoretyczne zachowań inwestorów najczęściej opisują zmiany poziomu cen rynkowych jako następstwa napływu nowych informacji, zmiany widełek spreadu bid-ask, zmian wolumenu, czy rodzaju i liczby składanych zleceń; (por. Easley, Kiefer, O'Hara, 1996b; Easley, Hvidkjaer, O'Hara, 2002; O'Hara, 2003; Acharya, Pedersen, 2005). Transakcje rynkowe zawierane są w kolejnych okresach pomiędzy obojętnymi na ryzyko animatorami rynku, inwestorami dokonującymi transakcji w wyniku napływu na rynek nowej informacji oraz inwestorami, których aktywność nie jest związana z napływem sygnałów informacyjnych.

Wśród polskojęzycznej literatury zawierającej szeroki przegląd modeli mikrostruktury rynku należy wymienić monografię Doman (2011). Natomiast badania nad VPIN i innymi miarami wskazującymi na możliwość zawarcia transakcji wynikających z napływu informacji dla międzybankowego kasowego rynku złotego prowadziła Bień-Barkowska (2010), Bień-Barkowska (2012) i Bień-Barkowska (2013).

Celem niniejszego opracowania jest zaprezentowanie jednej z metod określania „zawartości informacyjnej” transakcji zawieranych przez inwestorów na Giełdzie Papierów Wartościowych w Warszawie. W opracowaniu przedstawiono konstrukcję miary VPIN (ang. *volume-synchronized probability of informed trading*) – prawdopodobieństwa zawarcia transakcji wynikających z napływu nowych informacji. Miarę tę opisali Easley, Lopez de Prado i O'Hara (2011) jako nową procedurę estymacji

parametrów modelu EKOP opartą na porównaniu zmienności cen i wolumenu transakcji. Badacze w swojej metodologii przyjęli założenia o normalności rozkładu stóp zwrotu cen instrumentu finansowego w algorytmie Bulk Volume Classification (*BVC*). Zasadniczym celem artykułu jest podjęcie dyskusji nad wrażliwością miary VPIN na zmianę postaci rozkładu stóp zwrotu w algorytmie *BVC*. W niniejszym artykule rozważono następujące postacie rozkładów stóp zwrotu: rozkład normalny, rozkład *t*-Studenta z zadaną arbitralnie liczbą stopni swobody oraz rozkład *t*-Studenta z liczbą stopni swobody podlegającą estymacji.

Badanie empiryczne zostało przeprowadzone na przykładzie cen akcji spółki KGHM Polska Miedź S.A. notowanej na Giełdzie Papierów Wartościowych w Warszawie. Dobór spółki motywowany był wyłącznie faktem dużego zainteresowania inwestorów wskazanym walorem giełdowym, a także dużą płynnością tych akcji.

W kolejnej części pracy zawarto za literaturą definicję pojęcia zawierania transakcji będących skutkiem napływu na rynek nowych informacji (*ang. informed trading*). Następnie opisana została parametryzacja modelu EKOP i miary PIN. W rozdziale czwartym opisany został algorytm Bulk Volume Classification. Piąta część artykułu zawiera opis miary VPIN wykorzystanej w badaniu empirycznym oraz dyskusję nad uzyskanymi wynikami empirycznymi.

2. INFORMED TRADING

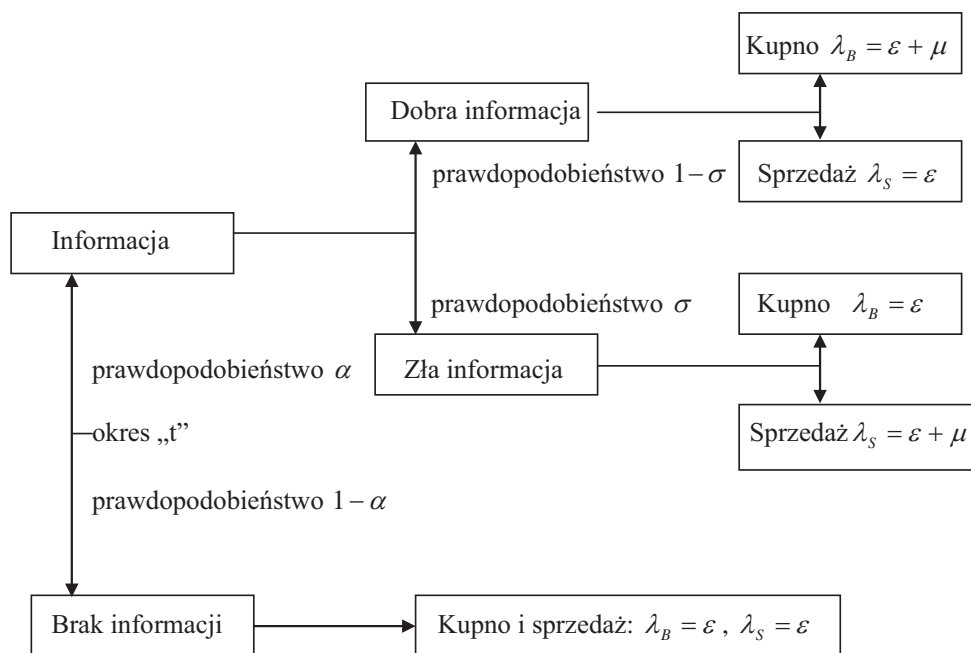
Argumenty przeczące istnieniu rynków w pełni efektywnych zwykle dotyczą anomalii obserwowanych na rynkach finansowych takich, jak efekty kalendarzowe, zmienna w czasie korelacja stóp zwrotu, czy bańki spekulacyjne. Współcześnie dominuje przekonanie, że na światowych rynkach finansowych występuje asymetria informacyjna pomiędzy inwestorami, która może być wykorzystywana do zdobywania przewagi (por. Doman, 2011). Dlatego ciągle podejmowane są badania nad zachowaniem uczestników rynku i samym procesem dyskontowania informacji w cenach rynkowych. Bień-Barkowska (2012) w swej pracy opisuje jeden z modeli informacji (*ang. information models*), który umożliwia pomiar „zawartości informacyjnej” procesu transakcyjnego. Scenariusz składanych zleceń kupna i sprzedaży stanowi odzwierciedlenie oczekiwań inwestorów co do przyszłej wartości instrumentów finansowych (por. Glosten, Milgrom, 1985; Easley, O’Hara, 1987). Ponadto autorka wskazuje, że w modelach informacji wyodrębnia się dwie grupy inwestorów:

- zawierających transakcje w oparciu o napływającą na rynek nową informację (*ang. informed traders*, w dalszej części pracy nazywanych inwestorami lepiej poinformowanymi),
- inwestorów dokonujących transakcji niemających związku z napływem sygnałów informacyjnych (*ang. uninformed traders*).

Model EKOP pochodzący od nazwisk autorów Easley, Kiefer, O’Hara oraz Paperman (1996a) jest najczęściej przywoływaną formalną parametryzacją służącą do opisu

procesu transakcyjnego. Dla ustalenia porządku zakłada się, że transakcje poszczególnych uczestników rynku zawierane są w kolejnych okresach $i = 1, \dots, I$.

Poniżej zaprezentowano diagram przedstawiający założenia modelu EKOP.



Rysunek 1. Diagram modelu EKOP

Źródło: Bień-Barkowska (2013).

Szczegółowy opis założeń modelu EKOP został zaprezentowany przez Easley i inni (2012b). Autorzy przez C_i oznaczyli cenę danego waloru giełdowego w chwili i . Na początku każdego okresu notowań na rynku kapitałowym może wystąpić przynajmniej jedno zdarzenie niosące ważną dla notowanej spółki informację. Prawdopodobieństwo wystąpienia takiego zdarzenia oznaczono przez α , a jego brak przez $1 - \alpha$. Jeśli pojawi się korzystna dla spółki informacja to inwestor posiadający wiedzę o jej znaczeniu będzie się spodziewał wzrostu ceny waloru rynkowego \bar{C}_i , natomiast przy wystąpieniu złego dla spółki zdarzenia będzie spodziewał się spadku ceny \bar{C}_i . Prawdopodobieństwo pojawienia się niekorzystnej informacji oznaczono przez σ , natomiast prawdopodobieństwo pojawienia się dobrej informacji o spółce oznaczono przez $1 - \sigma$. Zgodnie z założeniami modelu EKOP transakcje kupna i sprzedaży papierów wartościowych odbywają się zgodnie z niezależnymi procesami Poissona odpowiednio o wartościach oczekiwanych λ_B i λ_S . Inwestorzy dokonujący

transakcji w wyniku napływu nowych informacji o spółce, czy instrumencie finansowym wystawiają zlecenia kupna (dla dobrej informacji), bądź sprzedaży (dla złej) ze stałą intensywnością μ to jest średnią liczbą transakcji w ciągu rozważanego okresu. Inwestorzy zawierający transakcje nie związane z napływem sygnałów informacyjnych dokonują kupna i sprzedaży papierów wartościowych z intensywnością równą odpowiednio $\lambda_B = \varepsilon$ oraz $\lambda_S = \varepsilon$ (por. Easley i inni, 2012b). Zgodnie z Bień-Barkowską (2013) w modelu EKOP, gdy inwestorzy lepiej poinformowani posiadają nowe, niekorzystne dla spółki informacje, to transakcje kupna inicjowane są jedynie przez pozostałych uczestników rynku. W modelu fakt ten odzwierciedlony został poprzez wprowadzenie parametru $\lambda_B = \varepsilon$. Natomiast transakcje sprzedaży inicjowane są przez obie grupy inwestorów (parametr $\lambda_S = \varepsilon + \mu$). Z drugiej strony, gdy na rynku pojawia się jakaś dobra informacja to zlecenia kupna wystawiane są przez obie grupy inwestorów ($\lambda_B = \varepsilon + \mu$), a zlecenia sprzedaży (z intensywnością $\lambda_S = \varepsilon$) pochodzą wyłącznie od inwestorów dokonujących transakcji nie związanych z napływem sygnałów informacyjnych. W modelu EKOP inwestorzy lepiej poinformowani wyceniają aktualną wartość danego waloru rynkowego na podstawie napływających informacji. Jeśli wskazują one na przewartościowanie waloru spółki, to inwestorzy spodziewają się spadku ceny, a jeśli akcje są niedowartościowane to wzrostu ich kursu.

W modelu EKOP mieszanka dwuwymiarowych rozkładów Poissona transakcji kupna i sprzedaży w okresie notowań t określona jest poprzez cztery parametry α , σ , ε i μ . Easley i inni (2012b) wskazują, że parametr ε powinien być interpretowany jako normalny poziom liczby kupujących i sprzedających to jest zawierane transakcje nie są związane z napływem sygnałów informacyjnych. Ponadprzeciętna liczba transakcji kupna i sprzedaży interpretowana jest jako efekt działania inwestorów lepiej poinformowanych i w modelu EKOP odzwierciedlona została poprzez wprowadzenie μ . Parametry α i σ służą do określenia notowań cen rynkowych, dla których pojawiły się jakieś nowe informacje.

Wiedza na temat aktualnego stanu rynku oraz kondycji danej spółki wpływa na przekonania inwestorów co do przyszłej zmiany kursu, co następnie ma odzwierciedlenie w poziomie widełek bid-ask. Easley i inni (2012b) w swej pracy przez $P(i) = (P_n(i), P_b(i), P_g(i))$ oznaczają ciąg prawdopodobieństw takich, że na rynku w czasie i pojawiające się zdarzenie: nie wnosi nowych informacji (n), przynosi złą informację (b) lub dobrą informację (g). Warunkowa względem wszelkich informacji zaobserwowanych w przeszłości Ψ_{i-1} wartość oczekiwana ceny instrumentu finansowego dana jest wzorem:

$$E[C_i | \Psi_{i-1}] = P_n(i)C_i^* + P_b(i)\underline{C}_i + P_g(i)\overline{C}_i, \quad (1)$$

przy czym $C_i^* = \sigma \underline{C}_i + (1 - \sigma)\overline{C}_i$ oznacza wartość oczekiwaną waloru giełdowego względem informacji dostępnych w chwili $i-1$.

Zgodnie z Easley i inni (2012b) w modelu EKOP kurs bid stanowi warunkową wartość oczekiwaną ceny

$$Bid(i) = E[C_i | \psi_{i-1}] - \frac{\mu P_b(i)}{\varepsilon + \mu P_b(i)} (E[C_i | \psi_{i-1}] - \underline{C}_i). \quad (2)$$

Z drugiej strony kurs ask będzie równy wartości oczekiwanej ceny instrumentu pod warunkiem, że inwestor lepiej poinformowany wstrzymuje się z kupnem papierów wartościowych na rynku, to jest

$$Ask(i) = E[C_i | \psi_{i-1}] + \frac{\mu P_g(i)}{\varepsilon + \mu P_g(i)} (\overline{C}_i - E[C_i | \psi_{i-1}]). \quad (3)$$

Zgodnie z powyższymi równaniami kurs bid i ask będzie równy wartości oczekiwanej ceny instrumentu, jeśli żadna z transakcji nie została zawarta w wyniku napływu na rynek nowej informacji ($\mu = 0$). Gdy $\varepsilon = 0$ to kurs bid i ask równe są odpowiednio cenie minimalnej i maksymalnej. Jeśli natomiast w danym momencie transakcji dokonują obie grupy inwestorów to $Bid(i) < E[C_i | \psi_{i-1}] < Ask(i)$ (por. Easley i inni, 2012b).

Jeśli w dowolnym okresie i przez $\Delta(i) = Ask(i) - Bid(i)$ oznaczmy spread bid-ask to na mocy równań (2) i (3) spread możemy zapisać następująco:

$$\Delta(i) = \frac{\mu P_g(i)}{\varepsilon + \mu P_g(i)} (\overline{C}_i - E[C_i | \psi_{i-1}]) + \frac{\mu P_b(i)}{\varepsilon + \mu P_b(i)} (E[C_i | \psi_{i-1}] - \underline{C}_i). \quad (4)$$

Gdy pojawienie się dobrej lub złej informacji jest jednakowo prawdopodobne ($\sigma = 1 - \sigma$) wartość oczekiwana spreadu będzie wyrażać się wzorem:

$$\Delta = \frac{\alpha \mu}{\alpha \mu + 2\varepsilon} (\overline{C}_i - \underline{C}_i). \quad (5)$$

Pierwszy czynnik równania (5) w modelu EKOP nazwano miarą prawdopodobieństwa zawarcia transakcji wynikających z napływu nowych informacji (*ang. probability of informed trading – PIN*), to jest

$$PIN = \frac{\alpha \mu}{\alpha \mu + 2\varepsilon}, \quad (6)$$

przy czym $a\mu + 2\varepsilon$ określa wszystkie dostępne w danej chwili zlecenia, a $a\mu$ opisuje rynkowe zlecenia inwestorów mających dostęp do sygnałów informacyjnych (por. Easley i inni, 2012b).

3. ALGORYTM BULK VOLUME CLASSIFICATION

Do estymacji parametrów modelu EKOP niezbędny jest podział ogółu transakcji na te, które zostały zainicjowane przez stronę kupującą i te inicjowane przez sprzedających. Taki podział można uzyskać poprzez zastosowanie jednej z technik klasyfikacji transakcji opartych na zmienności cen i wolumenu. W literaturze światowej szeroko stosowana jest technika zwana „Algorytmem Lee-Ready’ego” od nazwisk jej twórców (Lee, Ready, 1991) oraz inne algorytmy takie, jak „test tickowy” (*ang. tick test*) oraz „test notowania” (*ang. quotes test*), czy algorytm Bulk Volume Classification (por. Easley i inni, 2013).

Test tickowy klasyfikuje daną transakcję na zainicjowaną przez kupującego (sprzedającego), jeśli cena bezpośrednio poprzedzającej transakcji była niższa (wyższa) od ceny w obecnej transakcji. Jeśli cena dwóch następujących po sobie notowań była taka sama, to kierunek transakcji wyznaczała ostatnio odnotowana zmiana ceny. Zaprezentowana metoda jest bardzo prosta do wdrożenia, ponieważ jedyną niezbędną informacją jest sekwencja cen.

Test notowania określa transakcję jako zainicjowaną przez kupującego (sprzedającego), jeśli aktualnie zawarta transakcja jest bliższa do ostatnio odnotowanej ceny ask (bid). Jeśli aktualna cena jest równa średniej mid (*ang. midquote*), wówczas do ustalenia kierunku transakcji wykorzystywana jest wcześniej odnotowana zmiana cen.

Algorytm Lee-Ready’ego uważany jest za bardziej trafny od jego poprzedników. Kierunek transakcji ustala się przy wykorzystaniu łącznie testu tickowego i testu notowań. Główną wadą tej metody jest problem z klasyfikacją transakcji zawieranych na początku każdego okresu notowań.

Easley i inni (2012a) przedstawili probabilistyczną metodę klasyfikacji transakcji zwaną algorytmem Bulk Volume Classification (*BVC*). W zadanym okresie zmienność cen i wolumenu zawiera informację na temat oczekiwań inwestorów co do aktualnej wartości notowanych akcji. Badacze poprzez V określili sumę transakcji (*ang. volume bars* nazywane w dalszej części paczkami wolumenu transakcji) utworzoną w ten sposób, że V zawiera dokładną liczbę akcji będących przedmiotem obrotu kolejno następujących po sobie transakcji. W przypadku, gdy wolumen ostatniej transakcji wchodzącej w skład danej paczki był większy niż wymagany to „nadmiar” liczby akcji stanowił początek kolejnej paczki V . W ten sposób autorzy podzielili całą próbę na równe co do liczby akcji paczki wolumenu V . Jeśli ostatnia paczka nie była „pełna” to zostawała ona wyłączona z dalszych analiz. Paczki V oznaczmy kolejno przez $\tau = 1, \dots, n$. W algorytmie *BVC* każda paczka wolumenu (por. Easley i inni, 2012b) określona została jako paczka kupna (*ang. buy volume bucket*) albo paczka sprze-

daży (*ang. sell volume bucket*) wykorzystując zmianę ceny papieru wartościowego w obrębie z góry przyjętego interwału czasowego. Paczki kupna V_τ^B i sprzedaży V_τ^S wyznacza się następująco:

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P_i}}\right), \quad (7)$$

$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot [1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P_i}}\right)] = V - V_\tau^B, \quad (8)$$

gdzie: $t(\tau)$ oznacza indeks ostatniego elementu w paczce wolumenu τ , Z określa dystrybuantę przyjętego rozkładu prawdopodobieństwa, $\sigma_{\Delta P_i}$ wskazuje odchylenie standardowe zmian cen w obrębie paczki V_τ .

Jeśli ceny następujących po sobie elementów paczki są równe to opisana powyżej procedura dzieli wolumen V na połowę po równo na V_τ^B i V_τ^S . Jeśli natomiast cena wzrasta, to większa część wolumenu z paczki V uzupełnia paczkę kupna niż sprzedaży, a różnica zależy od wielkości wzrostu tej ceny. Zauważmy, że wskazany powyżej podział wolumenu transakcji ściśle zależy od przyjętego rozkładu prawdopodobieństwa stóp zwrotu oraz od wielkości wolumenu. Opisany w niniejszej pracy podział transakcji został następnie wykorzystywany do określenia nierównowagi informacyjnej występującej na rynku. Dla paczki wolumenu τ niech nierównowaga będzie równa $OI_\tau = |V_\tau^B - V_\tau^S|$. Wartość oczekiwana $E[OI_\tau]$ nie zależy od wielkości paczki V i służy do szacowania aktualnej nierównowagi informacyjnej występującej wśród inwestorów (por. Easley i inni, 2012b).

4. VPIN

Parametry strukturalne modelu EKOP α , σ , ε i μ mogą być estymowane w oparciu o liczbę zleceń kupna i sprzedaży. Brak dostępu do tego typu historycznych danych przysparza jednak wiele problemów zarówno badaczom jak i praktykom rynkowym. Wynika on przede wszystkim z wielkości kosztu dostępu do danych oraz ich kompletności. Z uwagi na ten fakt w niniejszej pracy zaprezentowano podejście, które bazuje na cenach i wolumenie transakcji to jest szeroko dostępnych danych rynkowych. Zakłada się, że każda transakcja zawarta na kilku papierach wartościowych traktowana jest jako kilka transakcji na pojedynczym papierze wartościowym po tej samej cenie równej cenie wyjściowej transakcji. Takie podejście pozwala na włączenie do analizy także wielkości giełdowych transakcji.

Easley i inni (2008) pokazali, że wartość oczekiwana nierównowagi informacyjnej występującej na rynku wyraża się wzorem $\frac{1}{n} \sum_{\tau=1}^n |V_{\tau}^B - V_{\tau}^S| \approx \alpha\mu$, natomiast oczekiwana liczba ogółu transakcji jest równa:

$$\frac{1}{n} \sum_{\tau=1}^n (V_{\tau}^B + V_{\tau}^S) = V = \alpha(1 - \delta)(\varepsilon + \mu + \varepsilon) + \alpha\delta(\mu + \varepsilon + \varepsilon) + (1 - \alpha)(\varepsilon + \varepsilon) = \alpha\mu + 2\varepsilon.$$

Czynnik $\alpha(1 - \delta)(\varepsilon + \mu + \varepsilon)$ odzwierciedla wolumen paczki, w której pojawiła się jakaś korzystna dla notowanej spółki informacja. Czynnik $\alpha\delta(\mu + \varepsilon + \varepsilon)$ określa wolumen, gdy pojawiła się zła informacja, a czynnik $(1 - \alpha)(\varepsilon + \varepsilon)$ wolumen transakcji nie związanych z napływem żadnych sygnałów informacyjnych (por. Easley i inni, 2012b).

Easley i inni (2011) wskazali, że podział transakcji w obrębie okresu notowań na wolumen kupna i sprzedaży, to jest $V_{\tau}^B + V_{\tau}^S = V$ dla każdego τ służy do wyliczenia miary *Volume-synchronized probability of informed trading* – *VPIN* następująco:

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = \frac{\alpha\mu}{V} \approx \frac{\sum_{\tau=1}^n IO_{\tau}}{nV} = \frac{\sum_{\tau=1}^n |V_{\tau}^B - V_{\tau}^S|}{nV}. \quad (9)$$

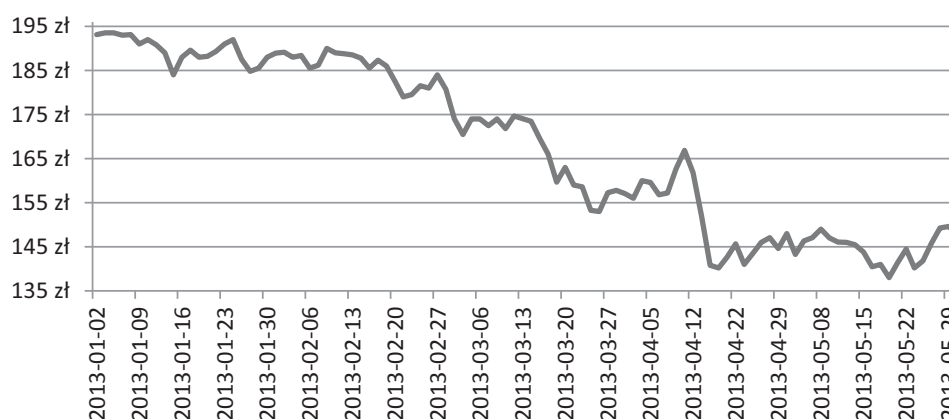
Miara *VPIN* zależy od wielkości przyjętego wolumenu V , nierównowagi informacyjnej występującej w każdej paczce V oraz liczby paczek wolumenu – n .

5. WYNIKI EMPIRYCZNE

W niniejszej pracy użyto notowań ze 101 dni transakcyjnych akcji spółki KGHM Polska Miedź S.A. w okresie od 2 stycznia 2013 roku do 31 maja 2013 roku obejmujących łącznie 171 187 obserwacji. Na rysunku 2 przedstawiono przebieg zmienności ceny zamknięcia akcji KGHM Polska Miedź S.A.

Wykorzystana w badaniu baza danych zawierała notowania transakcyjne występujące w nierównomiernych odstępach czasowych. W celu zredukowania liczby obserwacji i ułatwienia określenia „zawartości informacyjnej” transakcji zawieranych na rynku surowe dane poddano wstępnemu przekształceniu. Próba zawierała wyłącznie notowania zaobserwowane pomiędzy godziną 9:00 a 17:35. Następnie całą próbę podzielono na pięciominutowe odstępy czasu (czyli 103 pięciominutowe okresy w ciągu jednej sesji). Za cenę w każdym pięciominutowym podokresie przyjęto odpowiednio średnią arytmetyczną cen transakcji zaobserwowanych w ciągu wskazanych pięciu minut, a za wolumen – sumę wolumenu tych transakcji. W tabeli 1 zaprezentowano

wano podstawowe statystyki opisowe zmiennych uzyskanych po wykonaniu opisanej powyżej agregacji surowych danych.



Rysunek 2. Kurs zamknięcia akcji KGHM Polska Miedź S.A. w okresie od 02.01.2013 do 31.05.2013

Źródło: opracowanie własne na podstawie danych z <http://www.gpwinfostrefa.pl>.

Tabela 1.

Statystyki opisowe przyrostu cen akcji KGHM Polska Miedź S.A., odchylenia standardowego oraz wolumen transakcji w obrębie pięciominutowych okresów czasu

Zmienna	Średnia	Mediana	Wartość minimalna	Wartość maksymalna	Odchylenie standardowe
Przyrost średniej ceny	-0,0044	0	-4,3464	4,5335	0,2719
Odchylenie	0,09	0,07	0,00	2,59	0,08
Przyrost wolumenu	27	-45	-367810	401860	17904

Źródło: opracowanie własne.

Kolejnym krokiem był podział zagregowanej próby na pakiety wolumenu V. Za V przyjęto dokładnie 10 000 transakcji na pojedynczej akcji. Następnie stosując algorytm BVC każdą taką paczkę podzielono na paczki wolumenu kupna i sprzedaży. Ta metoda oczywiście pomija część notowań, natomiast podtrzymuje wartość informacyjną nierównowagi informacyjnej występującej w danym momencie na rynku. W niniejszej pracy zastosowano algorytm BVC oddzielnie dla różnych postaci rozkładu stóp zwrotu cen instrumentu, a mianowicie dla rozkładu normalnego oraz dla rozkładu *t*-Studenta. Dla drugiego rozkładu przyjęto liczbę stopni swobody na poziomie 0,25 zgodnie z propozycją Easley, Lopez de Prado i O'Hara (2013) oraz

jako alternatywę liczbę stopni swobody podlegającą estymacji. Dla każdego pięciominutowego okresu liczbę stopni swobody oszacowano wykorzystując formułę kurtozy rozkładu t -Studenta $K = 3 \frac{\nu - 2}{\nu - 4}$, gdzie ν oznacza liczbę stopni swobody rozkładu t -Studenta, a K kurtozę. Estymację liczby stopni swobody przeprowadzono na podstawie obliczeń kurtozy próbkowej. Jeżeli kurtoza z próby była równa 3 to arbitralnie przyjęto 100 stopni swobody. Natomiast za każdą oszacowaną liczbę stopni swobody mniejszą niż 4 przyjęto 4 stopnie swobody. Zasadność założeń co do liczby stopni swobody nie będzie stanowić przedmiotu dyskusji w prezentowanych artykule. Zagadnienia te będą rozwijane w ramach kontynuacji badań.

Opisany powyżej podział wolumenu transakcji został wykorzystany do oszacowania miary VPIN. Ponieważ wybór sposobu klasyfikacji wolumenu ma istotny wpływ na estymację miary VPIN w zaprezentowanym badaniu podjęto próbę określenia wrażliwości zmian wartości miary VPIN uzyskanych przy założeniu różnych rozkładów stóp zwrotu cen w algorytmie BVC. Ostatecznie każda miara VPIN uzyskana została na mocy wzoru (9) dla piętnastu następujących po sobie wolumenów V_{τ}^B i V_{τ}^S określających stopień nierównowagi transakcyjnej występującej na rynku.

Opisany powyżej proces przekształcania surowych danych został powtórzony oddzielnie dla rozkładu normalnego, rozkładu t -Studenta z 0,25 stopni swobody oraz rozkładu t -Studenta dla zmiennej liczby stopni swobody obliczonej z kurtozy próbkowej. Zasadniczym celem takiego podejścia była próba odpowiedzi na pytanie jak miara VPIN zależy od rodzaju arbitralnie przyjętego rozkładu prawdopodobieństwa w algorytmie BVC. W tabeli 2 zaprezentowano podstawowe statystyki opisowe miary VPIN dla trzech rozważanych rozkładów.

Tabela 2.

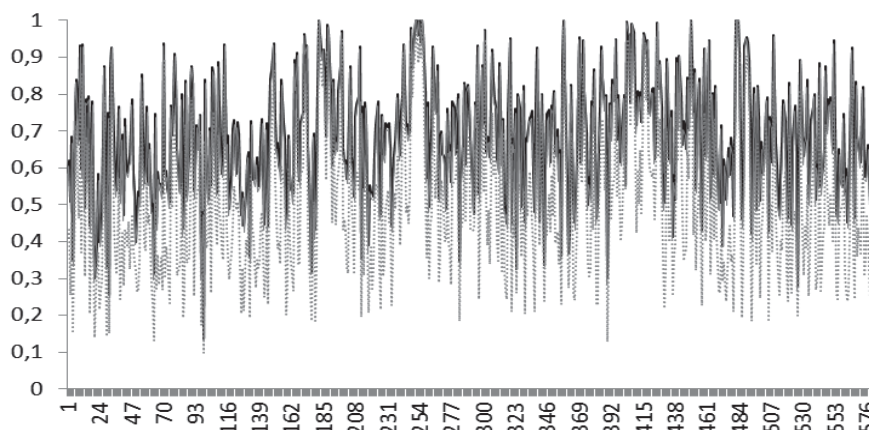
Wartości statystyk opisowych miary VPIN w zależności od przyjętego rozkładu stóp zwrotu cen akcji KGHM Polska Miedź S.A.

Rozkład	Średnia	Mediana	Wartość minimalna	Wartość maksymalna	Odchylenie standardowe	Współczynnik zmienności	Współczynnik ekscesu
Normalny	0,682691	0,680310	0,137730	1	0,159832	23,41213	-0,363133
t -Student z estymowaną liczbą stopni swobody	0,671002	0,666233	0,135287	1	0,162676	24,24372	-0,417305
t -Student z 0,25 stopni swobody	0,496400	0,464566	0,097450	0,999733	0,196307	39,54610	-0,354539

Źródło: opracowanie własne.

Nietrudno zauważyć, że przyjęcie warunkowego rozkładu t -Studenta z 0,25 stopniami swobody daje zdecydowanie odmienne wartości: średniej, mediany oraz odchylenia standardowego VPIN niż dwa pozostałe rozkłady. Średnie prawdopodobieństwo zawarcia transakcji przez inwestora lepiej poinformowanego w przypadku rozkładu t -Studenta z 0,25 stopniami swobody było równie 49,64% i około o jedną piątą niższe niż średnie prawdopodobieństwo przy pozostałych rozkładach. Analogicznie było w przypadku mediany. Wartości uzyskane dla rozkładu normalnego i rozkładu t -Studenta ze zmienną liczbą stopni swobody nieznacznie się różniły, przy czym w przypadku drugiego rozkładu zarówno średnia, mediana jak i wartość minimalna przyjmowała nieco niższe wartości. Odchylenie standardowe miary VPIN było najniższe przy przyjęciu rozkładu normalnego, co oznacza większą stabilność w zadanym okresie. Fakt ten potwierdza również najniższy współczynnik zmienności dla tego rozkładu. Zwróćmy uwagę, że w rozważanym okresie dla każdego przyjętego rozkładu wystąpiła co najmniej jedna obserwacja dla której prawdopodobieństwo zawarcia transakcji wynikającej z napływu nowej informacji było praktycznie pewne. Nie wystąpiła natomiast sytuacja w której żadna transakcja zawarta na rynku nie byłaby związana z napływem sygnałów informacyjnych.

Na rysunku 3 przedstawiono przebieg zmienności miary VPIN w okresie od 2 stycznia do 31 maja 2013 roku oszacowanej przy wykorzystaniu algorytmu BVC z rozkładem normalnym oraz wariantami rozkładu t -Studenta.



Rysunek 3. Przebieg zmienności miary VPIN cen akcji KGHM Polska Miedź S.A. dla dwóch wariantów rozkładu t -Studenta oraz rozkładu normalnego.

Czarna linia na wykresie oznacza miarę VPIN uzyskaną przy założeniu normalnego rozkładu stóp zwrotu, szara ciągła linia oznacza VPIN indukowany na podstawie rozkładu t -Studenta z estymowaną liczbą stopni swobody, zaś szara przerywana linia prezentuje przebieg zmienności VPIN dla rozkładu t -Studenta z liczbą stopni swobody $v=0,25$ zgodnie z jedną z propozycji Easley, Lopez de Prado, O'Hara (2013).

Źródło: opracowanie własne.

Wartości miary VPIN przy założeniu rozkładu normalnego jako rozkładu stóp zwrotu cen akcji cechują się dużą zmiennością w całym analizowanym okresie. Prawdopodobieństwo pojawienia się transakcji dokonywanych przez inwestorów lepiej poinformowanych najczęściej oscyluje pomiędzy 0,5 a 0,9. Oznacza to, że większość transakcji rynkowych zawierana była w wyniku napływu nowych informacji. Podając analizie krótsze okresy czasu można wskazać kilka wyraźnych trendów malejących (na przykład pomiędzy 180 i 220 miarą VPIN to jest mniej więcej od 1 do 8 marca 2013 roku, czyli okresu krótkiej stabilizacji kursu akcji po uprzednim ich silnym spadku), czy znacznie rosnących (od 230 do 260 wartości VPIN, czyli od 11 do 18 marca 2013 roku w którym to akcje KGHM Polska Miedź S.A. silnie traciły na wartości). Tendencje rozwojowe miary VPIN w wyznaczonym okresie odzwierciedlają odpowiednio wzrost bądź spadek liczby transakcji zawieranych przez inwestorów w wyniku napływu sygnałów informacyjnych. Zdecydowane odbicie się od trendu miary VPIN (około 35% wartości bezpośrednio sąsiadujących) miało miejsce dla 389 wartości VPIN (dokładnie 12 kwietnia 2013 roku) co najprawdopodobniej było odzwierciedleniem bardzo silnego w tym dniu spadku cen akcji oraz rozpoczynającej się stabilizacji kursu.

Wartości miary VPIN przy założeniu rozkładu t -Studenta z szacowaną liczbą stopni swobody cechuje się również dużą zmiennością w analizowanym okresie podobnie, jak przy rozkładzie normalnym z tą różnicą że wartości miary oscylują jednak mniej więcej w przedziale od 0,5 do 0,8. Oznacza to, że w badanym okresie na rynku zdecydowanie przeważały transakcje wynikające z napływu nowych informacji. Podobnie jak uprzednio opierając się o wartości VPIN można wskazać kilka podokresów zdecydowanego wzrostu (spadku) liczby transakcji zawieranych przez inwestorów lepiej poinformowanych. Analogicznie jak dla rozkładu normalnego od 180 do 220 oszacowanej wartości VPIN miał miejsce wyraźny spadek wartości VPIN, czyli stopniowego wycofywania się z rynku przez inwestorów lepiej poinformowanych. Można również wskazać okresy, w których liczba tych inwestorów w porównaniu do pozostałych uczestników rynku zmieniała się gwałtownie z okresu na okres (około 100 wartości VPIN raz miara była bliska 0,1, aby następnie przyjąć wartości około 0,8).

Wartości miary VPIN dla rozkładu t -Studenta z liczbą stopni swobody równą 0,25 cechują się większą stabilnością. Jednak w tym przypadku mamy do czynienia z częstymi silnymi skokami wartości VPIN zarówno dodatnimi jak i ujemnymi. W kilku przypadkach prawdopodobieństwo pojawienia się transakcji wynikających z napływu nowych informacji było praktycznie pewne. Miara VPIN przyjmowała wartości z przedziału od 0,1 do 1,0 co oznacza duże zróżnicowanie liczby transakcji zawieranych przez inwestorów lepiej poinformowanych w poszczególnych okresach czasu. Przyjęcie tak niskiej liczby stopni swobody w rozkładzie t -Studenta spowodowało wygaszenie poszczególnych skoków miary VPIN w kolejnych okresach oraz pojawienie się bardzo wyraźnych trendów. Kosztem jest jednak nieregularny przebieg zmienności miary VPIN i zdecydowany wzrost rozpiętości przyjmowanego prawdo-

podobieństwa zawarcia transakcji przez inwestorów lepiej poinformowanych. Zgodnie z powyższym przyjęcie rozkładu t -Studenta z 0,25 stopniami swobody w algorytmie BVC powoduje uzyskanie zdecydowanie odmiennych wartości miary VPIN (niższe i bardziej zróżnicowane w całym badanym okresie), niż w dwóch pozostałych rozkładach stóp zwrotu cen akcji.

6. PODSUMOWANIE

W niniejszej pracy omówiono procedurę estymacji parametrów modelu EKOP zaproponowaną przez Easley i inni (2011). Badanie aktywności inwestorów zawierających transakcje wynikające z napływu nowych informacji na Giełdzie Papierów Wartościowych w Warszawie oparto o analizę zmian wartości miary VPIN. W artykule zastosowano metodę klasyfikacji wolumenu zaproponowaną przez Easley i inni (2012a) zwaną Bulk Volume Classification oddzielnie dla trzech postaci rozkładu stóp zwrotu: rozkładu normalnego, rozkładu t -Studenta z liczbą stopni swobody $v=0,25$ oraz rozkładu t -Studenta z liczbą stopni swobody uzyskaną z kurtozy próbkowej. Wartości miary VPIN przy przyjęciu rozkładu normalnego i rozkładu t -Studenta z szacowaną liczbą stopni swobody były do siebie zbliżone. Arbitralne przyjęcie niskiej liczby stopni swobody dla rozkładu t -Studenta nieco obniża wartości VPIN w porównaniu do pozostałych typów rozkładu. Ponadto przebieg zmienności VPIN w przypadku wszystkich zaprezentowanych rozkładów cechuje duże podobieństwo w każdym kolejnym podokresie.

Przedmiotem kolejnych badań autorów będzie analiza wrażliwość miary VPIN na zmianę podstawowych założeń modelu EKOP i algorytmu BVC innych niż opisane w niniejszym opracowaniu. Ponadto autorzy będą kontynuować rozważania na temat wpływu informacji na obserwowane na rynku stopy zwrotu notowanych papierów wartościowych oraz wolumen transakcji.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

- Acharya V. V., Pedersen L. H., (2005), Asset Pricing with Liquidity Risk, *Journal of Financial Economics*, 77 (2), 375–410.
- Bień-Barkowska K., (2010), Przepływ zleceń a kurs walutowy – badanie mikrostruktury międzybankowego kasowego rynku złotego, *Bank i Kredyt*, 41 (5), 6–39.
- Bień-Barkowska K., (2013), Informed and Uninformed Trading in the EUR/PLN Spot Market, *Applied Financial Economics*, 23 (7), 619–628.
- Bień-Barkowska K., (2012), Model sekwencyjnego zawierania transakcji – zastosowanie do analizy procesu transakcyjnego na kasowym rynku złotego, *Metody Ilościowe w Badaniach Ekonomicznych*, Tom XIII/3, 42–51.

- Doman M., (2011), *Mikrostruktura giełd papierów wartościowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, 108.
- Easley D., Hvidkjaer S., O'Hara M., (2002), Is Information Risk a Determinant of Asset Returns?, *Journal of Finance*, 5, 2185–2221.
- Easley D., Lopez de Prado M. M., O'Hara M., (2012a), Bulk Classification of Trading Activity, *Johnson School Research Paper Series*, March, 2012, 1–40.
- Easley D., Lopez de Prado M. M., O'Hara M., (2012b), Flow Toxicity and Liquidity in a High Frequency World, *Review of Financial Studies*, February, 8–11.
- Easley D., Lopez de Prado M. M., O'Hara M., (2011), Flow Toxicity and Volatility in a High Frequency World, *Working paper*, SSRN, February, 1–39.
- Easley D., Engle R., O'Hara M., Wu L., (2008), Time-Varying Arrival Rates of Informed and Uninformed Trades, *Journal of Financial Econometrics*, 6, 171–207.
- Easley D., Kiefer N., O'Hara M., Paperman J., (1996a), Liquidity, Information and Infrequently Traded Stocks, *Journal of Finance*, 51, 1405–1436.
- Easley D., Kiefer N. M., O'Hara M., (1996b), Cream-skimming or Profit-sharing? The Curious Role of Purchased Order Flow, *Journal of Finance*, 51 (3), 811–833.
- Easley D., Lopez de Prado M. M., O'Hara M., (2013), Bulk Classification of Trading Activity, *Working paper*, SSRN-id1989555, March.
- Easley D., O'Hara M., (1987), Price, Trade Size and Information in Securities Markets, *Journal of Financial Economics*, 119, 69–90.
- Glosten L. R., Milgrom P. R., (1985), Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics*, 14, 71–100.
- Lee C., Ready M., (1991), Inferring Trade Direction from Intraday Data, *Journal of Finance*, 46 (2), 733–746.
- O'Hara M., (2003), Presidential Address: Liquidity and Price Discovery, *Journal of Finance*, 58 (4), 1335–1354.

EKONOMETRYCZNA ANALIZA PRAWDOPODOBIENSTWA ZAWARCIA TRANSAKCJI
WYNIKAJĄCYCH Z NAPŁYWU INFORMACJI – WPŁYW ZAŁOŻEŃ CO DO ROZKŁADU
STÓP ZWROTU NA ZMIENNOŚĆ MIARY VPIN

Streszczenie

Niniejsze opracowanie stanowi próbę syntetycznego przybliżenia fluktuacji intencji i oczekiwań inwestorów dokonujących transakcji wynikających z napływu nowej informacji na rynek (*ang. informed trading*). Na podstawie zmienności wolumenu i cen akcji KGHM Polska Miedź Spółki Akcyjnej notowanej na Giełdzie Papierów Wartościowych w Warszawie oszacowano miarę prawdopodobieństwa zawarcia transakcji wynikających z napływu nowych informacji – VPIN (*volume-synchronized probability of informed trading*). Otrzymane wyniki badania mogą stanowić podstawę do analityczno-empirycznej weryfikacji, jak pojawiająca się na rynku informacja wpływa na zachowania inwestorów.

Słowa kluczowe: nowa informacja, mikrostruktura rynku, model EKOP, prawdopodobieństwo zawarcia transakcji wynikających z napływu nowych informacji

ECONOMETRIC ANALYSIS OF PROBABILITY OF INFORMED TRADING -THE IMPACT
OF DISTRIBUTION OF MARKET PRICE RETURN ON VPIN CHANGE

A b s t r a c t

This article is as an attempt to introduce fluctuation of expectations of investors who make transactions resulting from the influx of new information on the market(*informed trading*). The Volume-synchronized probability of informed trading was estimated on the basis of price variation of KGHM Polska Miedź S.A. shares listed on the Warsaw Stock Market. The results of this analysis might serve as an analytical and empirical basis for further verification on how new information affects investors' behavior.

Keywords: new information, market microstructure, EKOP model, volume synchronized probability of informed trading

