METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XIX/3, 2018, s. 284 – 292

Received: 28.05.2018
Accepted: 05.10.2018

FILTRACJA FINANSOWYCH SZEREGÓW CZASOWYCH METODAMI NIEUJEMNEJ FAKTORYZACJI MACIERZY

Ryszard Szupiluk, Paweł Rubach https://orcid.org/0000-0001-5487-609X
Kolegium Analiz Ekonomicznych
Szkoła Główna Handlowa w Warszawie
e-mail: ryszard.szupiluk@sgh.waw.pl; pawel.rubach@sgh.waw.pl

Streszczenie: W niniejszym artykule przedstawimy metodę wielowymiarowej filtracji do eliminacji szumów oraz estymacji trendów z finansowych szeregów czasowych. Jednym z istotnych elementów procesu filtracji będzie dekompozycja szeregów czasowych przy wykorzystaniu nieujemnej faktoryzacji macierzy. Prezentowana metoda może być wykorzystana w wielu praktycznych obszarach finansów i zarządzania jak analiza techniczna rynków, systemy inwestycyjne czy modele ryzyka.

Słowa kluczowe: filtracja szeregów czasowych, estymacja trendów, nieujemna faktoryzacja macierzy

JEL classification: C32, C45, C63

WPROWADZENIE

Analiza trendów szeregów czasowych jest istotnym zagadnieniem w wielu obszarach finansów i zarządzania [Murphy 1999]. Istnieje w tym obszarze wiele rozwiązań odnoszących się do różnych metod i metodyk. W naszym ujęciu estymację trendu potraktujemy jako eliminację z podstawowych wartości instrumentów niepożądanych składników takich jak zakłócenia lub krótkookresowe fluktuacje.

Całe zagadnienie można potraktować także jako formę wielowymiarowej filtracji [Szupiluk 2013; Szupiluk, Rubach 2018]. Proces filtracji składa się z następujących etapów: dekompozycji (separacji), identyfikacji i eliminacji szumów oraz rekompozycji (mieszania). Metodą dekompozycji jaką zastosujemy w ramach niniejszego opracowania będzie nieujemna faktoryzacja macierzy (ang.

https://doi.org/10.22630/MIBE.2018.19.3.26

Non-negative Matrix Factorization – NMF) [Cichocki, Zdunek, Phan, Amari 2009].

Należy zaznaczyć jednak, iż algorytm NMF może być zastąpiony innym typem dekompozycji takim jak analiza składowych niezależnych, analiza składowych głównych lub analiza składowych rzadkich. Główną motywacją wyboru metody NMF jest jej adekwatność dla nieujemnych oraz niestacjonarnych danych oraz potwierdzona skuteczność w separacji rzeczywistych sygnałów. Z założenia przyjmiemy także, że rozważamy przypadek wielowymiarowy. Oznacza to, że interesuje nas znalezienie i wydzielenie wspólnych elementów zakłócających.

Prezentowane podejście może stanowić alternatywę dla klasycznych technik estymacji trendu takich jak dopasowanie funkcji opisującej, wygładzanie średnimi kroczącymi lub filtracja spektralna. Metoda ta, mimo eliminacji zakłóceń, pozwala na otrzymanie trendu rozumianego jako zasadnicza tendencja rozwojowa przy zachowaniu typowych dla instrumentów finansowych charakterystyk statystycznych takich jak zmienność lub rzadkość danych.

WIELOWYMIAROWA FILTRACJA TRENDU

Przyjmijmy istnienie zbioru *m* szeregów czasowych $\mathbf{x}_i$, $i=1,...,m$ reprezentujących wartości cen określonych instrumentów finansowych. Szeregi te zbierzemy w jednej wielowymiarowej zmiennej $\mathbf{X}=[\mathbf{x}_1,...,\mathbf{x}_m]^T$. Zauważmy dalej, że obserwowane wartości zawierają komponenty związane z długo-, średnio- i krótkookresowymi trendami jak również losowe fluktuacje i zaszumienia niepowiązane ze zdarzeniami ekonomicznymi. Inaczej mówiąc, możemy przyjąć, że dany rezultat jest kombinacją istotnych dla nas komponentów wspierających $\hat{\mathbf{s}}_j$, $j=1,...,p$ związanych z poszukiwanymi długookresowymi trendami oraz zakłócających $\tilde{\mathbf{s}}_l$, $l=1,...,q$ związanych z szumem informacyjnym, losowymi zachowani itp. Wszystkie te składniki związane ze specyfiką danego modelu, potraktujemy jako ukryte komponenty bazowe zawarte (zmieszane) w wielowymiarowej zmiennej $\mathbf{X}$ o wymiarach $m\times K$, gdzie *K* oznacza ilość obserwacji. W przypadku liniowego sposobu (systemu) mieszania można to zapisać jako

$$\mathbf{X}=\mathbf{AS}, \qquad (1)$$

gdzie macierz $\mathbf{A}=\lfloor a_{ij} \rfloor$ o wymiarach $m\times n$ reprezentuje system mieszający zaś macierz $\mathbf{S}=[\hat{\mathbf{s}}_1,...,\hat{\mathbf{s}}_p,\tilde{\mathbf{s}}_{p+1},...,\tilde{\mathbf{s}}_{p+q}]^T$ reprezentuje zbiór komponentów bazowych, gdzie $n=p+q$. Dla uproszczenia przyjmijmy, że $m=n$ zaś macierz $\mathbf{A}$ jest nieosobliwa. Zauważmy dalej, że identyfikując system mieszający $\mathbf{A}$ oraz komponenty bazowe **S** oraz eliminując komponenty zakłócające (stawiając odpowiednio $\tilde{\mathbf{s}}_l = \mathbf{0}$) otrzymamy

$$\widehat{\mathbf{X}} = \mathbf{A}[\hat{\mathbf{s}}_1,...,\hat{\mathbf{s}}_p,\mathbf{0}_{p+1},...,\mathbf{0}_n]^T, \tag{2}$$

gdzie $\widehat{\mathbf{X}} = [\hat{\mathbf{x}}_1,...,\hat{\mathbf{x}}_m]^T$ jest „poprawianą” wersją pierwotnych rezultatów **X**. Można to także przedstawić w analogicznej postaci wstawiając w macierzy $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2,..., \mathbf{a}_n]$ zera w kolumnie odpowiadającej eliminowanemu sygnałowi z macierzy **S**. W efekcie dla $\hat{\mathbf{A}} = [\mathbf{a}_1,..., \mathbf{a}_p,\mathbf{0}_{p+1},\mathbf{0}_{p+2},...,\mathbf{0}_n]$ otrzymamy

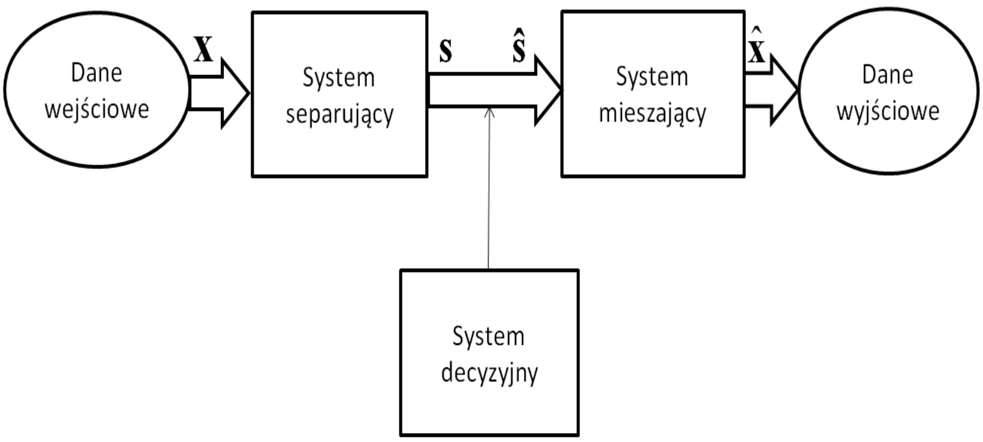
$$\widehat{\mathbf{X}} = [\mathbf{a}_1,..., \mathbf{a}_p,\mathbf{0}_{p+1},\mathbf{0}_{p+2},...,\mathbf{0}_n]\mathbf{S} = \hat{\mathbf{A}}\mathbf{S}. \tag{3}$$

Cały proces filtracji można przedstawić jako

$$\widehat{\mathbf{X}} = \hat{\mathbf{A}}\mathbf{S} = \hat{\mathbf{A}}\mathbf{A}^{-1}\mathbf{X}. \tag{4}$$

Kluczowym zagadnieniem w rozważanej procedurze jest znalezienie takiej transformacji, która rzeczywiście prowadzi do dekompozycji danych **X** na interesujące nas komponenty wspierające i zakłócające. W prowadzonych rozważaniach identyfikację komponentów bazowych dokonamy za pomocą algorytmu nieujemnej faktoryzacji. Cały proces przedstawiono na rysunku 1.

Rysunek 1. Proces filtracji finansowych szeregów czasowych



Źródło: opracowanie własne

NIEUJEMNA FAKTORYZACJA MACIERZY

Podstawową ideę nieujemnej faktoryzacji macierzy (NMF) można dla modelu (1) sprowadzić do faktoryzacji danej macierzy **X** jako iloczynu dwóch macierzy nieujemnych **A** i **S** [Cichocki, Zdunek, Phan, Amari 2009]. Przez nieujemność macierzy **A** rozumie się nieujemność jej elementów, co będzie zapisywane jako $\mathbf{A} > 0$. W praktyce **X** oznacza nieujemną macierz danych obserwowanych, **A** oraz **S** zaś są nieujemnymi macierzami faktoryzującymi reprezentującymi odpowiednio system mieszający oraz sygnały źródłowe. Poszukiwanie macierzy **A** oraz **S** wiąże się z minimalizacją określonej funkcji celu. Jedną z takich funkcji może stanowić funkcja dywergencji alpha Amariego postaci [Amari, Nagaoka 2000; Cichocki 2010]:

$$D_A^{(\alpha)}(\mathbf{X} \| \mathbf{AS}) = \frac{1}{\alpha(\alpha-1)} \sum_{it} \left(x_{it}^{\alpha} [\mathbf{AS}]_{it}^{1-\alpha} - \alpha x_{it} + (\alpha-1)[\mathbf{AS}]_{it} \right). \tag{5}$$

Minimalizacja tej funkcji prowadzi do następującego macierzowego algorytmu NMF

$$\mathbf{S} \leftarrow \mathbf{S}.\times \left(\mathbf{A}^{\mathrm{T}} \left((\mathbf{X})./(\mathbf{AS}) \right)^{.\alpha} \right)^{.(1/\alpha)}, \tag{6}$$

$$\mathbf{A} \leftarrow \mathbf{A}.\times \left(\left((\mathbf{X})./(\mathbf{AS}) \right)^{.\alpha} \mathbf{S}^{\mathrm{T}} \right)^{.(1/\alpha)}, \tag{7}$$

gdzie $.\times$ oznacza mnożenie tablicowe element na element, podobna notacja (matlabowska) odnosi się do dzielenia i potęgowania.

Jest kwestią otwartą dobór parametru α. Dla określonych wartości tego parametru dywergencja (5) jest tożsama z innymi popularnymi dywergencjami [Cichocki 2010]. W szczególności interesujące są tu przypadki graniczne $\alpha \to 1$ oraz $\alpha \to 0$, dla których otrzymuje się odpowiednio uogólnioną dywergencję Kullbacka–Leiblera oraz dualną uogólnioną dywergencję Kullbacka–Leiblera. W naszym przypadku zaproponujemy powiązanie tego parametru z dodatkową charakterystyką jaką jest wariancja stóp zwrotu

$$\alpha = \min \sum_{i \in P} \mathrm{var}(z_i), \tag{8}$$

gdzie $z_i(k) = \ln\left(s_i(k+1) / s_i(k) \right)$, zaś P oznacza zbiór komponentów kontrolnych wziętych pod uwagę przy wyliczaniu wariancji. Wybór komponentów kontrolnych jest arbitralny i zależy od celu inwestycyjnego. Wielowymiarowa natura metody NMF wymaga analizy wielu instrumentów, jednak część z nich może pełnić rolę zewnętrznego środowiska informacyjnego dla zestawu nas interesującego. Zasadne wydaje się ograniczenie do charakterystyk kontrolnych instrumentów dla nas istotnych. W skrajnym przypadku oznacza to, że parametr alfa wyznaczamy w oparciu o wariancję tylko jednego instrumentu (istotnego dla nas).

ALGORYTM FILTRACJI I UWAGI METODOLOGICZNE

Pełna postać algorytmu filtracji jest następująca:

1. Zbierz wybrane instrumenty finansowe w jednej macierzy **X**
2. Za pomocą algorytmu NMF dokonaj separacji (dekompozycji) macierzy **X** na macierz komponentów bazowych **S** oraz macierz mieszającą **A**
3. Zidentyfikuj komponenty wpierające oraz zakłócające oraz wyeliminuj te ostatnie.
4. Dokonaj transformacji mieszającej odwrotnej do separującej na zbiorze komponentów wspierających (i wyzerowanych zakłócających).

W ramach prezentowanej metody zaznaczyć należy kilka kwestii metodologicznych. Przede wszystkim prezentowana koncepcja w swojej podstawowej postaci bazuje zasadniczo na zależnościach algebraicznych i od strony formalnej nie wymaga żadnych założeń probabilistycznych w szczególności takich jak stacjonarność czy ergodyczność. Dodatkowo eksplorowana nieujemność jest niejako naturalnie adresowana do wartości cen instrumentów (a nie stóp zwrotu). Jej skuteczne zastosowanie wiąże się wyłącznie z pewnym domyślnie przyjętym założeniem o stabilności zjawiska, dla którego ma miejsce zastosowanie. Jeżeli jednak dysponujemy wiedzą o charakterystyce probabilistycznej przetwarzanych danych może być ona włączona do rozwiązania przez dodanie odpowiednich warunków ograniczających do funkcji celu (5). Możemy także powiązać bezpośrednie dzielenie algorytmu NMF z charakterystykami statystycznymi np. wariancją logarytmicznych stóp zwrotu jak to podaliśmy wyżej. W efekcie mamy eksplorowane jednocześnie wartości cen oraz stóp zwrotu.

Kolejną kwestią jest postać przyjętego modelu mieszania (transformacji) komponentów bazowych w wartości cen instrumentu finansowego. Przyjęty liniowy model generujący, jest tzw. modelem roboczym, pozwalającym na wyprowadzenie algorytmu, nie przesądza to jednak o skuteczności jego działania dla wyłącznie tego przypadku.

Ostatnim zagadnieniem jest klasyfikacja komponentów bazowych jako wspierających lub zakłócających. W przypadku niewielu instrumentów może być to dokonane poprzez pełny (mechaniczny) test eliminacji wszystkich możliwych kombinacji komponentów bazowych i wybór najbardziej efektywnej. Przyjmiemy, że do takiego przypadku ograniczamy się w niniejszym opracowaniu. Oczywiście w przypadku większej liczby instrumentów takie pełne przeszukanie nie jest możliwe ze względów obliczeniowych i konieczne są pewne aprioryczne kryteria oceny tychże komponentów.

Podejście mechaniczne i aprioryczne jest jednak zasadne głównie w przypadkach kiedy mamy jasno zdefiniowane kryterium jakości filtracji.

Z punktu widzenia praktyki inwestycyjnej ze względu na jednoczesne występowanie trendów o różnych okresach (długo-, średnio- i krótkookresowych) konieczny jest zwykle bezpośredni nadzór badawczy (lub rozbudowany system rozpoznawania trendów i formacji) pozwalający określić jaki zestaw komponentów daje interesujący wynik.

EKSPERYMENT PRAKTYCZNY

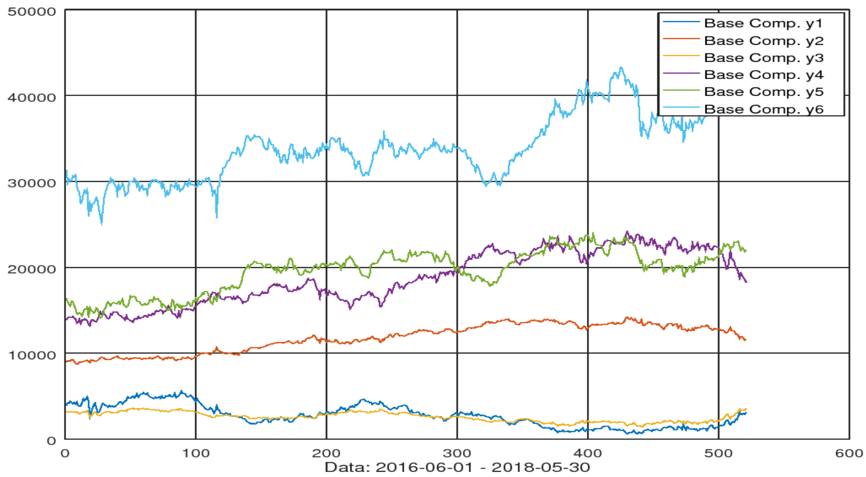
W niniejszej części zaprezentujemy działanie przedstawionej koncepcji na rzeczywistych danych. Estymacja trendu zwykle wiąże się z pewnymi dodatkowymi warunkami lub arbitralnymi decyzjami takimi jak wybór zakresu danych, liczba wykorzystanych instrumentów oraz biznesowy lub ekonomiczny cel prowadzonej analizy. W naszym przykładzie tym dodatkowym warunkiem będzie taka eliminacja fluktuacji i zakłóceń aby uzyskany przebieg trendu stanowił sygnał wyprzedzający przebieg podstawowy. Inaczej mówiąc oczekujemy, iż będzie osiągał on wcześniej punkty zwrotne (minima lub maksima) dając użyteczną inwestycyjnie informację.

Przyjmiemy, że interesującym nas instrumentem będzie indeks SP 500, jednak ze względu na wielowymiarową naturę proponowanej metody zostaną także wykorzystane inne indeksy. Te dodatkowe indeksy można interpretować jako dodatkową (zewnętrzną) informację wprowadzaną do analizy podstawowego instrumentu.

Eksperyment przeprowadzono na wartościach sześciu wybranych indeksów giełdowych: WIG 20, SP 500, FTSE 250, NIKKEI 225, DAX (Deutsche Boerse), BUX (Budapest Stock Exchange). Obserwowano dzienne wartości zamknięcia, a w przypadku wystąpienia w tygodniu święta państwowego, przyjmowano wartość indeksu z dnia poprzedzającego. Obserwowane wartości pochodzą z okresu 1 czerwca 2016 – 30 maja 2018. Badania przeprowadzono dla parametru $\alpha = 0.95$.

W pierwszym kroku dane indeksów zebrano w jednej zmiennej wielowymiarowej, a następnie dokonano dekompozycji na komponenty bazowe. Przebieg wartości komponentów bazowych uzyskanych w wyniku transformacji NMF przedstawia rysunek 2.

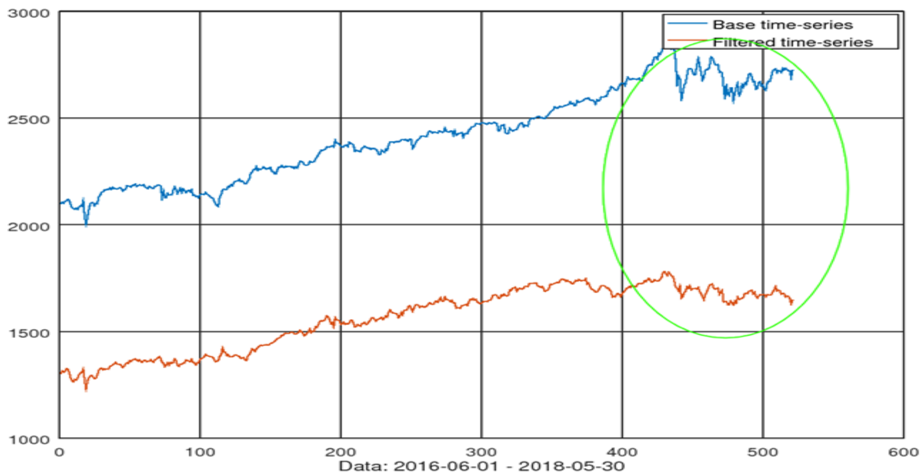
Rysunek 2. Wartości komponentów bazowych dla obserwowanych danych indeksów



Źródło: opracowanie własne

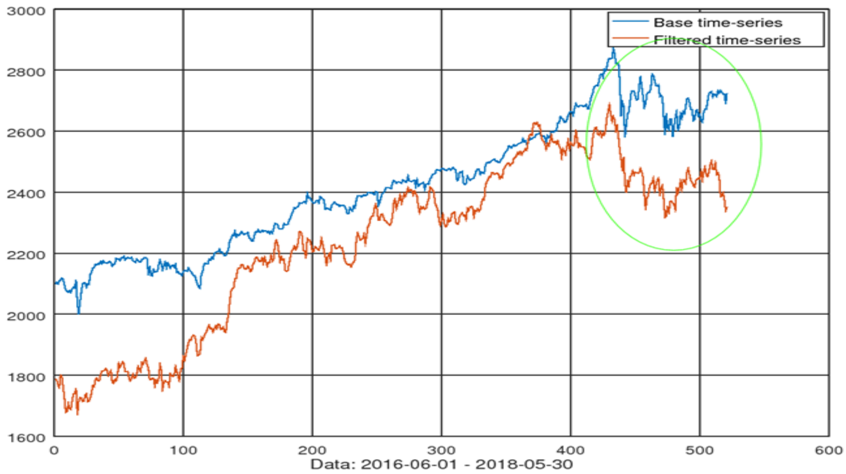
W kolejnych krokach usuwano wybrane komponenty bazowe z szeregów czasowych poszczególnych indeksów oraz dokonywano rekompozycji prowadzącej od komponentów bazowych do przefiltrowanych wartości indeksów.

Rysunek 3. Rzeczywiste wartości indeksu SP 500 oraz szereg otrzymany w wyniku usunięcia z indeksu piątego elementu bazowego



Źródło: opracowanie własne

Rysunek 4. Rzeczywiste wartości indeksu SP 500 oraz szereg otrzymany w wyniku usunięcia z indeksu trzeciego elementu bazowego



Źródło: opracowanie własne

Rysunki 3 oraz 4 przedstawiają oryginalne wartości indeksu SP500 oraz jego wartości po filtracji NMF. W obydwu przypadkach widzimy wcześniejszą zmianę trendu na wartościach filtrowanych. Przebieg z rysunku 3 otrzymany został przez eliminację komponentu piątego zaś dla szeregu z rysunku 4 przy eliminacji komponentu trzeciego. Z faktu zgodności trendów szeregów oryginalnego i filtrowanych a następnie wcześniejszej zmiany trendów filtrowanych możemy wnioskować, że komponenty uznane za wspierające są faktycznie powiązane z długookresowym przebiegiem indeksu. Zwróćmy także uwagę, że komponenty eliminowane zawarte są w całym szeregu oryginalnym jednak ich usunięcie nie zmienia oraz nie odkształca istotnie trendu w fazie zgodności (około 400 pierwszych obserwacji). Może to odróżniać tą formę filtracji od filtracji dokonywanych na bazie analizy spektralnej.

Jednocześnie, mimo procesu filtracji uzyskany przebieg zachowuje typowy dla finansowych szeregów giełdowych fraktalny „wygląd". Może więc stanowić podstawę do dalszych analiz np. z wykorzystaniem technik symulacyjnych.

PODSUMOWANIE

Rozwiązania prezentowane w tym artykule łączą dwa aspekty badawcze: wielowymiarową filtrację szeregów czasowych oraz nieujemną faktoryzację macierzy. Ich wykorzystanie w kontekście instrumentów finansowych pozwala na wsparcie decyzji inwestycyjnych podejmowanych m.in. w strategiach opartych na analizie technicznej. Cała koncepcja stanowi jednak otwarty schemat badawczy, w ramach którego wykorzystywane mogą być inne dekompozycje wielowymiarowe.

Z punktu widzenia ekonomicznego/finansowego przedstawione opracowanie ogranicza się do aspektu metodycznego i metodologicznego. W takim rozumieniu można przyjąć, że mamy do czynienia przede wszystkim z pewnym retrospektywnym przetworzeniem podstawowych instrumentów co może być użyteczne choćby w praktycznych systemach inwestycyjnych.

Należy jednak zaznaczyć, że dalszą intencją autorów jest usytuowanie tej metody w szerszej optyce ekonomicznej w szczególności w nurtach kwestionujących hipotezę racjonalnych oczekiwań oraz efektywnego rynku. W takim rozumieniu otrzymane wyniki mogą nabrać głębszej interpretacji ekonomicznej oraz uzyskać istotny walor instrumentalny. Jest to jeden z obszarów dalszych prac w ramach mniejszej problematyki.

BIBLIOGRAFIA

Amari S., Nagaoka H. (2000) Methods of Information Geometry. Oxford University Press, New York.

Cichocki A., Amari S. (2010) Families of Alpha-Beta-and Gamma-Divergences: Flexible and Robust Measures of Similarities, Entropy, 12, 1532-1568.

Cichocki A., Zdunek R., Phan A.-H., Amari S. (2009) Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis. John Wiley.

Murphy J. J. (1999) Analiza techniczna rynków finansowych. WIG-Press.

Szupiluk R. (2013) Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining. Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.

Szupiluk R., Rubach P. (2018) Extreme Value Model for Volatility Measure in Machine Learning Ensemble. [w:] Rutkowski L., Scherer R., Korytkowski M., Pedrycz W., Tadeusiewicz R., Zurada J. (red.) Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science, 10841, Springer, Cham.

FILTRATION OF FINANCIAL TIME SERIES USING NON-NEGATIVE MATRIX FACTORIZATION METHODS

Abstract: In this paper, we will present a method of multivariate filtration that may be used to eliminate noise and estimate trends in financial time-series. A significant element of the filtration process is the decomposition of time-series using nonnegative matrix factorization. The presented method may be applied in many practical aspects of finance and management, in particular for use in technical analysis of financial markets, trading systems or risk models.

Keywords: time-series filtration, trend estimation, non-negative matrix factorization