

**TRADING STRATEGIES BASED ON PREDICTING PRICES OF FUTURES
CONTRACTS**

Jan MROZOWSKI

Promotor: Prof. zw. dr hab. **Marian NOGA**

WSTĘP

Rynki terminowe dla towarów i walut grają ważną rolę w dzisiejszym świecie finansów międzynarodowych poprzez dawanie użytkownikom możliwości minimalizowania ryzyka albo handlowania w celu generowania dochodu czy maksymalizacji zysku. Kontrakty Futures dla metali szlachetnych takich jak Złoto są używane przez organizacje korzystające ze złota w procesie produkcji, jak również przez instytucje finansowe w celu zarządzania ryzykiem dzięki udowodnionej negatywnej korelacji pomiędzy ceną metalu a zmiennością rynku.

Niedawne postępy w algorytmach uczenia maszynowego zdobyły szerokie uznanie w niektórych gałęziach przemysłu, ale ich zastosowanie do rynków terminowych nie jest obszernie omówione w literaturze naukowej. Tworzenie modeli statystycznych przewidujących ceny towarów i użycie ich w handlu jest zwykle limitowane do domów handlowych, które nie udostępniają większości swoich badań jako sposób zachowania przewagi konkurencyjnej.

Nauzer Balsara mówi w swojej książce o handlu derywatami, *Money Management Strategies for Futures Traders*¹: „Sygnał kupna jest generowany, kiedy krótsza z dwóch ruchomych średnich przekracza dłuższą; sygnał sprzedaży jest generowany, gdy krótsza ruchoma średnia spada poniżej dłuższej ruchomej średniej”. Następnie dodaje „Uzbrojony w tą informację, handlarz może oszacować wartość odcięcia, poza którą jest niskie prawdopodobieństwo, że niezrealizowana strata będzie odzyskana i handel zakończy się zyskownie.”

Niektórzy badacze² eksperymentowali z przewidywaniem cen derywatów ropy używając Convolutional Neural Network (CNN). Ich badania pokazują, że CNN produkuje lepsze rezultaty niż tradycyjne modele ekonomiczne, ale nie wystarczająco celne, żeby były przydatne. Dzieje się tak z powodu złożonych nieliniowych charakterystyk związków wychodzących od „wielu złożonych naturalnych, ekonomicznych i politycznych czynników”. Nie znaleziono podobnych badań związanych z derywatami złota.

Wszystkie nie-akademickie źródła dla handlarzy sugerują używanie maksymalnej dozwolonej straty dla handlu podczas spekulacji na nowych trendach. Nie ma jednak żadnej koncepcji „minimalnej granicy zysku”, czy minimalnej ilości czasu w jakim kontrakty są przewidywane do utrzymania się. Autor sugeruje, że stworzenie tych nowych metryk i statystyczne obliczanie prawdopodobieństwa zysku może ulepszyć strategie handlowe poprzez ignorowanie niektórych handłów na podstawie sygnałów, albo lokowanie na nie mniejszej ilości kapitału czy eksplorowanie nisko ryzykownych strategii takich jak strategia kalendarzowa.

Słowa Kluczowe:

Derywaty, Rynki Terminowe, Algorytmy, Uczenie Maszynowe, Sieci Neuronalne, Autoencodery, Bayesian Structured Time Series, Krótkoterminowe przewidywanie trendów, Przewidywanie cen terminowych.

¹ Balsara, Nauzer. (1992). *Money Management Strategies for Futures Traders*. New York: John Wiley & Sons, Inc.

² Luo, Z.; Cai, X.; Tanaka, K.; Takiguchi, T.; Kinkyo, T.; Hamori, S. Can We Forecast Daily Oil Futures Prices? Experimental Evidence from Convolutional Neural Networks. *J. Risk Financial Manag.* **2019**, *12*, 9.

CEL BADAŃ

Celne przewidywanie Derywatów Złota, jednego z najbardziej aktywnie handlowanego towaru na świecie, zawsze było ważne dla naukowców i handlarzy. Autor recenzuje rozwój rynku terminowego i strategii handlowych oraz techniki statystyczne tradycyjnie używane w prognozowaniu. Kilka dobrze znanych modeli, takich jak Black-Scholes zawiodło handlarzy i rynki podczas kryzysów finansowych w roku 2009, więc stworzenie i wykorzystanie nowych podejść jest krytyczne w celu podniesienia pewności użytkowników derywatów co do cen i procedur.

Badanie ma celu odpowiedzieć na pytanie „Czy użycie danych stworzonych z historycznych cen może zwiększyć zyskowność spekulacji przyszłych trendów wzrostowych?”. Analiza buduje modele uczenia maszynowego, które mogą być wykorzystane do przyszłych transakcji. Testowana jest moc predykcyjna modeli i stworzone zmienne. Modele mogą być testowane na innych kontraktach Futures i dostosowane odpowiednio w celu dywersyfikacji transakcji i zwiększania ich ilości. Hipotezą jest, że stworzone dane mogą pomóc zidentyfikować 90% udanych transakcji z ponad 50% celnością.

Autor ewaluuje wydajność algorytmów uczenia maszynowego podczas identyfikacji formacji nowych trendów wzrostowych w cenach Derywatów Złota na samym ich początku. Wykorzystanie metodologii uczenia maszynowego pokazuje niedobór relacji liniowych pomiędzy historycznymi cenami i ewolucją nowych trendów. Modele „Black Box”, takie jak Sieci Neuronalne i Autoencodery pozwalają handlarzom i analitykom na klasyfikację obserwacji w sposób, który może być wykorzystany do podjęcia decyzji w handlu derywatami.

OBIEKT BADAŃ

W sektorze finansowym kontrakt Futures jest typem transakcji pochodnej, która zaistniała jako prawna umowa o sprzedaż bądź kupno instrumentu bazowego za poprzednio określoną cenę (cenę wykonania) w określonym terminie (data wykonania). Kontrakty Futures wykonują transakcje używając instrumentów bazowych, którymi zwykle są dobra albo instrumenty finansowe³.

Wszystkie operacje związane z kontraktami Futures są negocjowane na specjalnych giełdach zwanych giełdami Futures które działają jako rynki dla sprzedających i kupujących. Pierwsze są nazwani dysponentami pozycji krótkiej (short position holder), drugie: dysponentami pozycji długiej (long position holder). Istnieje pewne ryzyko, że strona umowy zdecyduje się rozwiązać ją, albo anulować, jeżeli wynegocjowane ceny nie będą dla nich korzystne. Z tego powodu warunkiem realizacji transakcji jest zdeponowanie rezerwy wartości kontraktu u neutralnej trzeciej strony. Dla przykładu, w transakcjach Futures Złota, rezerwa wynosi pomiędzy dwa i dwadzieścia procent⁴.

Swoje początki Futures ma w 1972 roku, gdzie były głównie używane do negocjacji cen produktów rolnych. W późniejszych latach były głównie używane w transakcjach surowców naturalnych, takich jak ropa.

³ Chew, D. H. *Corporate Risk Management*. Columbia University Press, 2008. p. 23

⁴ Valdez, S., *An Introduction To Global Financial Markets* (3rd ed.). Basingstoke: Macmillan Press, 2000. p. 34-36

Z czasem, ten typ kontraktów rozwinął się i dzisiaj operują na instrumentach bazowych takich jak waluty, stopy procentowe i akcje, które odgrywają dużą rolę w ogólnym rynku Futures⁵.

Oryginalnie, głównym celem kontraktów Futures było złagodzenie ryzyka związanego z cenami i ruchem kursu wymiany poprzez pozwalanie stronom na ustalenie cen lub oszacowanie transakcji, które miały być sfinalizowane w późniejszym czasie, z wyprzedzeniem. Były pomocne, kiedy strony oczekiwały płatności z góry, w obcych walutach.

Strategie Handlu Futures

Z szerokiego wachlarza strategii omawianych w literaturze i opisywanych w badaniu zostały wyłonione dwie w celu dalszej eksploracji i użyciu w połączeniu z zaproponowanymi modelami:

- Swing Trading – z tym podejściem, Inwestor celowo zostawia transakcje otwarte na koncie przez okres więcej niż jednego dnia, czasami znacznie dłużej (czasem nawet kilka tygodni). Swing Trading zakłada wykorzystanie “wachadeł”, wyraźnie zaznaczonych części wykresu. W zasadzie jest używana z większymi przedziałami, dziennymi czy tygodniowymi. Inne zasady to zasady zarządzania kapitałem, optymalizacja strat i zysków.
- Strategia kalendarzowa (Calendar Spread) – równie ważne jest zdefiniowanie strategii kalendarzowej w terminach często używanych na rynku Futures. Jest również znana jako intracontract, intracommodity, intermonth albo time spread (rozkład czasowy) i zawiera kupowanie tej samej ilości przeciwstawnych kontraktów wygasających w różnych miesiącach.

PROPONOWANE ROZWIĄZANIA

Proponowana Strategia

Powszechna strategia na otwarcie pozycji długiej w złocie została stworzona poprzez studiowanie związków między ruchomymi średnimi różnych długości przed wytworzeniem się nowych trendów. Kiedy ruchoma średnia z 4 dni przecina średnią z 9 dni podczas wzrostu, może wytworzyć się nowa pozytywna anomalia w następujących dniach. Dzień, kiedy to się zdarza jest dalej zwany dniem sygnału.

Transakcje otwiera się w cenie rynkowej w dzień sygnału z celem utrzymania kontraktu przez sześć lub więcej dni, jeżeli wystąpi trend wzrostowy. Minimalny cel zysku jest mierzony poprzez uśrednienie różnicy między wysoką i niską ceną dzienną przez ostatnie 9 dni. Ta sama miara jest używana do określenia maksymalnej straty z transakcji i jest używana do zamknięcia pozycji w ciągu pierwszych sześciu dni, jeżeli jest taka konieczność. To podejście jest średnio zyskowe, jako że w ~30% przypadków materializuje się nowy trend wzrostowy ze średnim zyskiem będącym znacznie wyższym niż średnia strata.

⁵ Chew, Donald H. *op. cit.*, 2008, p 26

Cel Badania

Autor chciałby udoskonalić to podejście i stworzyć probabilistyczny model, który może być używany jako wskazówka ryzyka proponowanych handlów w momencie wejścia na pozycję. Model powinien zawierać dostępne historyczne ceny, ilość i tendencje rynkowe, które mogą być zaczerpnięte z dostępnych zmiennych. Są to impet zmiennych dla różnych długości czasu, proporcje aktualnej ceny do maximum i minimum w okresach czasu i inne wskaźniki techniczne używane przez handlarzy. Autor chce zbadać moc predykcyjną tych wtórnych zmiennych i stworzyć metody wcielające dane w proces zarządzania decyzyjnego[decyzjami].

Celem jest zdeterminowanie czy wtórne historyczne zmienne mogą lepiej przewidywać nowe trendy wzrostowe w Futures Złota i zwiększyć zyskowność tych inwestycji. Wyniki tego badania pomogą handlarzom lepiej zrozumieć moc predykcyjną wtórnych zmiennych i pomogą z decyzjami o wchodzeniu na pozycje i lokowaniu kapitału. Funkcje modelu zostaną zaprojektowane na podstawie dostępnych danych Futures Złota. To badanie będzie odkrywać związek między historycznymi cenami i cenami Futures dla dat z ustalonymi kryteriami.

Zmienne i zakres

Dla tego badania, autor brał pod uwagę tylko daty, w których zdarzyło się krzyżowanie 4 i 9-cio dniowych ruchomych średnich. Ceny z 6 dni po przecięciu używane są w celu klasyfikacji, które transakcje osiągnęły minimalny próg zysku. Jeżeli cena jest ponad progiem zmienna zależna wynosi 1, inaczej wynosi 0.

Dostępne dzienne ceny od 1975 roku zostały wydobyte z Barchart.com używając API:

- Cena otwarcia
- Cena wysoka
- Cena niska
- Cena zamknięcia
- Liczba transakcji (Volume)
- Liczba otwartych kontraktów (Open Interest)

Funkcje danych zostały stworzone dla zmiennych niezależnych poprzez kalkulację różnych typów impetów dostępnych zmiennych i proporcji ostatnich znanych cen do minimum i maximum różnych długości czasu. Poniżej są grupy zaprojektowanych w oparciu o funkcje, niezależnych zmiennych:

- Tempo zmiany (momentum, impet) dla wszystkich dostępnych zmiennych używając 1, 4, 9, 15, 30 i 60 dni i jego średniej ruchomej używając 4, 9, 15, 30 i 60 dni
- Tempo zmiany (momentum, impet) dla ceny zamknięcia ważonej przez Open liczbę transakcji, otwartych kontraktów i oba. Jego ruchoma średnia używając 4, 9, 15, 30 i 60 dni
- Stosunek ceny zamknięcia do minimum i maximum ceny zamknięcia w ostatnich 4, 9, 15, 30, 60, 90, 180 i 360 dniach
- Stosunek ceny zamknięcia do Wstęgi Bollingera i jej ruchomej średniej dla 4, 9, 15, 30, 60 i 90 dni
- Dzienna różnica między ceną kontraktu Futures i fizycznym złotem i jej ruchoma średnia dla 4, 9, 15, 30, 60 i 90 dni

Metody Modelowania

Po pierwsze, oceniana jest binarna dystrybucja w celu sprawdzenia czy jakiegokolwiek z algorytmów uczenia maszynowego może dać wynik odpowiedni dla zaproponowanej strategii. Metody użyte:

- Regresja Logistyczna (Logistic Regression)
- Drzewa Decyzyjne (Decision Trees)
- Las Losowy (Random Forest)
- Wzmocnienie Gradientowe (Gradient Booster Classifier)
- Sieć Neuronowa (Neural Network)

Po drugie, podejście szeregu czasowego, używając strukturalnych modeli szeregów czasowych jest zdefiniowany przez dwa równania. Równanie obserwacyjne [indeksujące?] wiąże zaobserwowaną daną y_t do wektora zmiennych ukrytych α_t znanych jako stan

$$y_t = Z^T \alpha_t + E_t.$$

Równanie obserwacyjne opisuje jak stan ukryty ewoluje w czasie.

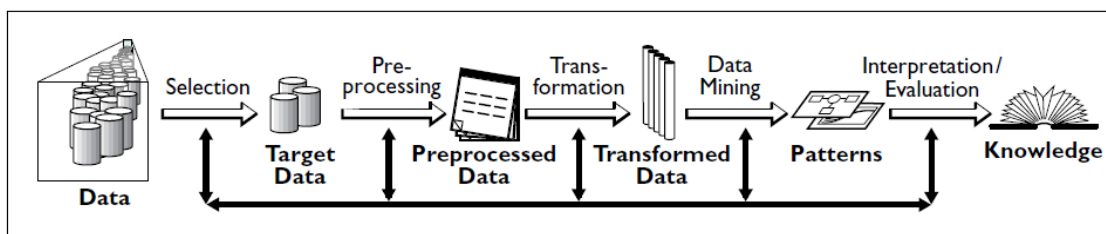
$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t.$$

Terminy błędów E_t i η_t są funkcjami Gaussa i niezależne od wszystkiego innego. Macierze Z_t , T_t i R_t są parametrami strukturalnymi. To podejście jest rozważane z powodu dostępności dobrze rozwiniętych algorytmów MCMC dla inferencji Bayesowskiej z parametrami zależnymi od czasu (paczka BSTS R).

BADANIA EKSPERYMETNALNE

Model bazowy

Badanie wykorzystuje strukturę Knowledge Discovery and Data Mining (KDD), zilustrowaną poniżej.



Rysunek 1. KDD Framework

Struktura zakłada znalezienie źródła danych, przetwarzanie wstępne i transformację danych, następnie analiza opisowa i modelowanie, na końcu interpretacja wyników. Najlepsze wyniki prowadzą do wiedzy, na podstawie której można działać. Wszystkie kroki opisane są w poniższych sekcjach.

Opis Danych

Barchart Inc. został zidentyfikowany jako potencjalne źródło. Ich darmowe oprogramowanie API zostało użyte do pobrania wstępnych danych. Zawierały one ponad 11,000 dziennych obserwacji cen Futures złota jak również ilość transakcji i otwartych pozycji z podziałem na poszczególne dni. Każdy dzień ma 4 punkty cenowe: otwarcia, niska, wysoka i zamknięcia. Dane zostały pobrane w ustrukturyzowanych arkuszach używając interfejsu R. Ceny punktowe dla złota zostały pobrane oddzielnie z tego samego źródła i włączone do danych.

Tworzenie zmiennej zależnej

Użycie monitorowanych algorytmów uczenia maszynowego wymaga posiadanie dla każdej obserwacji zmiennej zależnej jako miary wyniku i zmiennych niezależnych jako predyktorów. Dane ceny dziennej zostały użyte do stworzenia zmiennych zależnych i niezależnych.

Średnie ruchome ceny zamknięcia zostały obliczone używając okresów 9 i 4 dni, następnie dokonano kalkulacji różnicy pomiędzy nimi używając krótszej średniej jako bazy. Wszystkie obserwacje, w których różnica zmieniła się na pozytywną z negatywnej poprzedniego dnia zostały zidentyfikowane jako Data Sygnału. Dzienna miara zmienności została wybrana jako średnią różnicą pomiędzy wysoką i niską ceną dla 9 dni. Ta miara została użyta do określenia minimalnego zysku i maksymalnej dopuszczalnej straty z transakcji. Ona została dodana do ceny zamknięcia w poszczególnej dacie sygnału. Cena Celu została obliczona poprzez dodanie celu minimalnego zysku do ceny zamknięcia w dacie sygnału. Ta wyliczona cena Celu została następnie porównana z faktyczną ceną zamknięcia w 6 dniu od daty sygnału. Jeżeli cena rzeczywista była ponad ceną Celu to transakcja była traktowana jako zyskowa i zmienna zależna była kodowana jako 1. Jeżeli warunek nie był spełniony, zmienna zależna kodowana była jako 0.

Tworzenie zmiennych niezależnych

Wśród handlarzy i analityków popularne jest użycie Rate of Change (ROC), znane również jako Momentum, do zdefiniowania trendów w cenach i decyzjach rynkowych. Mając tego świadomość, ROC zostało obliczone dla wszystkich zmiennych używając popularnie wykorzystywanych okresów 1, 4, 9, 15, 30, 90, 180 i 360 dni. Te nowe zmienne zostały użyte do stworzenia średnich ruchomych dla nich. Dla zmiennych cenowych zostały również obliczone ważone średnie ruchome używając liczby otwartych kontraktów oraz transakcji jako wag.

Dodatkowe funkcje związane z potencjalnymi poziomami oporu cenowego i nastroju rynkowego zostały stworzone:

- Stosunek ceny zamknięcia do minimalnej i maksymalnej ceny okresów różnej długości: 9, 30, 60, 90 i 180 dni.
- Dzielne różnice między cenami Future zamknięcia i punktowe ceny złota i jego ruchome średnie dla 4, 9, 15, 30, 60 i 90 dni.

Łącznie, 163 zmienne niezależne zostały stworzone i dołączone do zmiennych zależnych. Z powodu braku pewnych danych i fakt, że niektóre ze zmiennych niezależnych nie dały się obliczyć dla pierwszego roku

dostępnych danych, 634 z dostępnych 650 obserwacji zostało wybranych do analizy. Zostały rozdzielone do zestawów Training i Testing używając stosunku 70/30.

Modelowanie

Analiza Opisowa

Struktury Weight of Evidence (WOE) i Information Value (IV) zostały użyte do analizy eksploracyjnej i odsiewu zmiennych dla klasyfikatora binarnego, będącego zmienną zależną. Według Kim Larsen ⁶ WOE i IV pozwalają na:

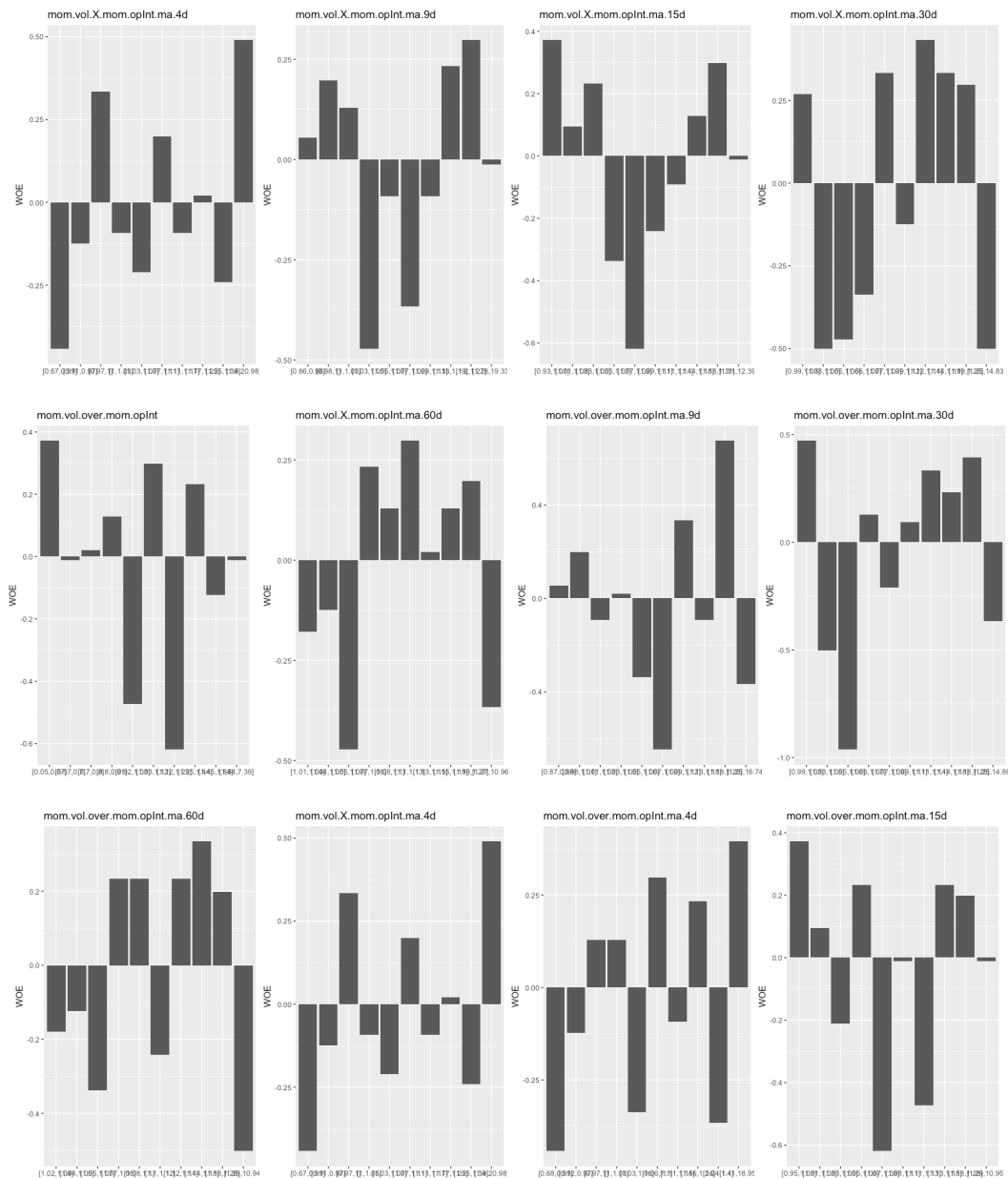
- Wzięcie pod uwagę niezależnej kontrybucji do wyniku każdej zmiennej.
- Wykrycie liniowych i nie-liniowych zależności.
- Rankowanie zmiennych w kategoriach “jednowymiarowych” sił predykcyjnych.
- Wizualizację korelacji między zmiennymi prognozującymi i wynikiem binarnym.

WOE i IV grają dwie różne role podczas analizy danych:

- WOE opisuje związek między predyktorem i zmienną celu binarnego.
- IV mierzy siłę tego związku.

Używając zbioru danych Testing, WOE i IV zostały obliczone. Poniższa tabela przedstawia 12 najwyższych zmiennych na podstawie IV. Ich dystrybucja sugeruje nie-liniowy związek między zmiennymi zależnymi i niezależnymi.

⁶ Larsen, Kim. (2016, April 07). Uplift Models: Optimizing the Impact of Your Marketing. Retrieved July 15, 2019, from: https://www.predictiveanalyticsworld.com/sanfrancisco/2016/uplift_modeling.php



Rys. 2. Information Value and Weight of Evidence

W celu zrozumienia czy ten sam wzór pojawia się w zbiorze danych Testing, Net Weight of Evidence (NWOE) i Net Information Value (NIV) zostały obliczone poprzez deaferencjację WOE i IV zbiorów danych Testing i Training. Poniższa tabela przedstawia 10 najwyższych zmiennych na podstawie NIV. Z grafów wynika że predykcyjna wartość zmiennych w Testing jest o wiele mniejsza od oryginalnych wartości obliczonych ze zbioru danych Training. Końcowy NIV jest mniejszy niż połowa początkowo oczekiwanej wartości. To sugeruje, że wnioski ze zbioru Training mogą nie zostać zwalidowane oraz, że lepsze wyniki mogą zostać osiągnięte z modeli nienadzorowanych takich jak Sieci Neuronalne.

	Cluster	Variable	IV	PENALTY	AdjIV
62	45	mom.49d.ma30cl	0.21430317	0.12541399	0.0888891803
52	35	mom.90.180.ma30cl	0.24042815	0.15998139	0.0804467672
76	59	cl.max15	0.27075677	0.22341167	0.0473450940
72	55	cl.max60	0.09868773	0.05796630	0.0407214349
29	19	mom15c9ma	0.17362132	0.13898644	0.0346348816
74	57	cl.max30	0.13911815	0.10484595	0.0342722027
69	52	cl.min180	0.12282260	0.08897672	0.0338458741
131	91	ma.fut.spot9d	0.17035145	0.13852668	0.0318247693
130	90	ma.fut.spot15d	0.24107789	0.21012616	0.0309517249
139	94	mom1c	0.06906315	0.03937413	0.0296890180

Tabela 3. 10 zmiennych z największą AdjIV

Wybór zmiennych

Zmienne niezależne zostały stworzone z wysokimi poziomami korelacji, jako że te same oryginalne zmienne zostały użyte w różny sposób, z oczekiwaniem, że niektóre z nich mogą mieć marginalnie lepsze moce predyktywne niż inne. Wysoki poziom korelacji powoduje problemy z większością nadzorowanych MAL-ów, więc musiała zostać wykonana redukcja wymiarowa.

W celu osiągnięcia tego i wybrania najbardziej użytecznych predyktorów, dokonano klastrowania zmiennych, używając poprzednio wyliczonego IV. Podzieliło to zmienne we wzajemnie wykluczające się zbiory takie jak:

- Korelacje między zmiennymi przypisanymi do tego samego zbioru są zmaksymalizowane.
- Korelacje między zmiennymi w różnych zbiorach są zminimalizowane.

Używając różnych poziomów odcięcia dla korelacji zmiennych zostały stworzone 2 zbiory danych:

- 24 zmienne z korelacją poniżej 0.8, najbardziej dopasowane do algorytmów czułych na współliniowość takich jak regresja i drzewa
- 123 zmienne z korelacją poniżej 0.92, najbardziej dopasowane do algorytmów nie czułych na współliniowość takich jak Sieci Neuronalne (NN)

Regresja Logistyczna

Binarna Regresja Logistyczna jest typem regresji, w której binarna zmienna odpowiedzi jest związana z zestawem dyskretnych albo ciągłych zmiennych wyjaśniających. Została przetestowana z dwudziestoma czterema zmiennymi wyjaśniającymi, pomimo analizy eksploracyjnej niedającej pobudek do istnienia związku liniowego. Najlepszy model miał trzy znaczące zmienne i nie udało mu się przewidzieć więcej niż 2% jako "1" w Training i Testing.

Użycie Regresji Logistycznej zostało rozszerzone poprzez użycie Principal Components (PC) od stu dwudziestu trzech zmiennych i Regresji Regularyzowanej, która penalizuje współczynniki niedodające żadnej wartości. Tylko jeden z osiemnastoma PC, które wyjaśniały 80% wariacji zostało wybrane jako znaczące. Wszystkie współczynniki zostały stłumione do "0" w regresji Regularyzowanej. Żaden nie przewidywał ponad 3% jako "1" w Testing i Training.

Drzewko Decyzyjne

Drzewko Decyzyjne CART jest metodologią klasyfikacji typu, w której drzewo jest tworzone przez rekurencyjne partycjonowanie danych i dopasowanie modelu predykcyjnego wewnątrz każdej partycji. W celu uniknięcia ponad-dopasowania, minimalny podział został ustawiony na 20 w tym badaniu; drzewo rozdzieli się tylko kiedy jest co najmniej 20 obserwacji w każdej gałęzi. Rezultat był obiecujący w etapie Training z 29% obserwacji przewidywanych jako "1" z celnością 74%. Nie powtórzyło się to w zbiorze Ing, z celnością predykcji "1" spadającą do 30% zgodnie z randomowym wyborem. Ten rezultat był oczekiwany, na podstawie różnicy między WOE i IV pomiędzy zbiorami danych Training i Testing.

Las Losowy

Las Losowy jest metodą zbioru, która konstruuje wiele drzewek decyzyjnych i wytwarza jedno drzewo, które jest modelem indywidualnych drzewek. Ilość skonstruowanych drzew została ustawiona na 200, z powodu relatywnie małego zbioru danych. Różne wartości maksymalnej głębokości od 2 do 15 zostały przetestowane i, bazując na najlepszej 5-cio krotnej walidacji krzyżowej, drzewo z głębokością ustawienia wynoszącą 5 poskutkowało w 29% predykcji jako "1" z 87% celnością. Podobnie do CART ten wynik przywrócił 28% celność z 30% obserwacjami przewidzianymi jako 1.

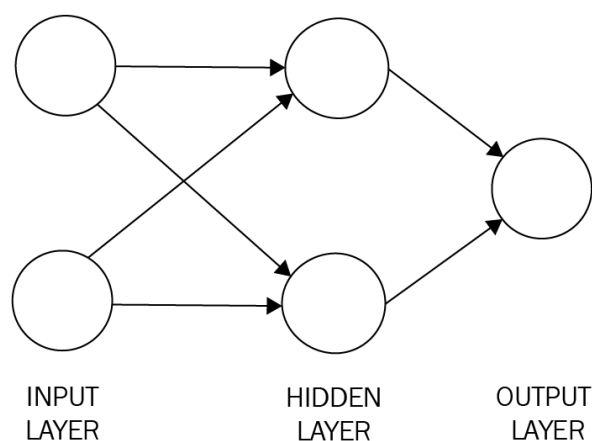
Wzmocnienie Gradientowe

Wzmocnienie Gradientowe (GB) jest metodą zbioru, która produkuje model predykcyjny w formie słabych modeli predykcyjnych, typowo drzew decyzyjnych. Model schodzi po kolejnych stadiach, dopóki nie znajdzie optymalnej wartości dla współczynników. Podobnie jak Las Losowy, ilość drzewek została ustawiona na 200 z różnymi ustawieniami głębi. W przeciwieństwie do Lasu Losowego, tempo uczenia było wymagane dla procesu wzmocnienia. Różne ustawienia głębi i tempa uczenia zostały wypróbowane, jednakże żadne z nich nie zdołały przewidzieć jakiegokolwiek "1" w Training. Wzmocnienie Gradientowe zostało więc odrzucone jako opcja.

Sieci Neuronalne

Sieci Neuronalne (NN) są systemem komputerowym bazującym na kolekcji jednostek albo węzłów, zwanymi sztucznymi neuronami. Rysunek 4. Perceptron wielowarstwowy (MLP) jest typem NN składającym się z co najmniej trzech warstw: wejścia, wyjścia i przynajmniej jednej warstwy ukrytej. MLP używa nadzorowanej

techniki uczenia zwanej „backpropagation for training”⁷. Jego różne warstwy i nie-liniowe funkcje aktywacyjne pozwalają mu na rozróżnienie, danych które nie są linearnie rozdzielne⁸.



Rysunek 4. Diagram typowej sieci neuronalnej

MLP z ilością warstw pomiędzy jeden i sześć zostały przetestowane z różnymi funkcjami aktywacyjnymi dla warstw. Najlepsze rezultaty dla zbioru danych z stu dwudziestoma trzema zmiennymi zostały wyprodukowane przez model z trzema ukrytymi warstwami oraz funkcją aktywacyjną Rectifier Linear Unit (ReLU). Tabela 4.3.7.2. Rezultatem była celność ponad 50% w identyfikacji “1” w bazach Training i Testing. Ogólny procent zidentyfikowanych “1” był niski, w obu przypadkach pomiędzy 11% i 14%. Obniżanie odcięcia prawdopodobieństwa nie pomogło, jako że zmiękczenie przewidzianego “1” było towarzyszące spadkowi celności do poziomowych Wyboru Losowego.

Zdarzenie (Wybór Losowy)	Testing		Training	
	Przewidywane “0” (Dokładność %)	Przewidywane “1” (Dokładność %)	Przewidywane “0” (Dokładność %)	Przewidywane “1” (Dokładność %)
0 (70%)	132 (76%)	11	280 (73%)	24
1 (30%)	41	17 (61%)	104	25 (51%)
Przewidywany % od całości	84%	14%	89%	11%

Tabela 5. Macierz błędu sieci neuronalnej z 123 zmiennymi

⁷ Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016) *Deep Learning*. MIT Press. p. 196.

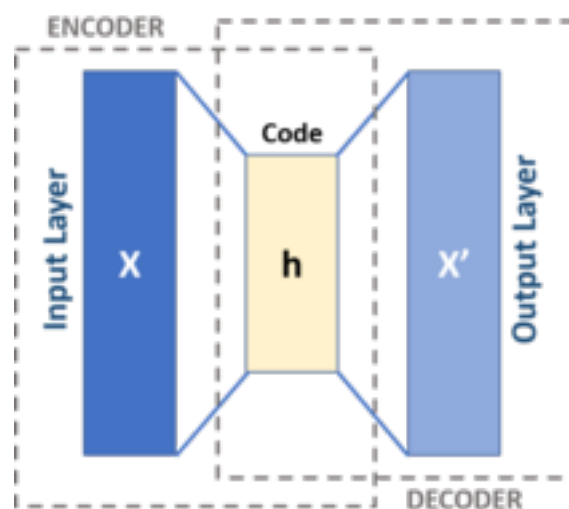
⁸ Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems, 2(4), 303–314.

Model z tymi samymi parametrami został wytrenowany używając zbiorów danych z 24 zmiennymi. Zyskał rezultat nieco gorszy w kwestii celności predykcji 1; jednakże przewidział znacznie więcej ich w zestawach Training i Testing. Celność wynosiła 55% dla Training i 43% dla Testing z proporcją przewidzianego “1” wynoszącą 19% i 21%.

Zdarzenie (Wybór Losowy)	Testing		Training	
	Przewidywane “0” (Dokładność %)	Przewidywane “1” (Dokładność %)	Przewidywane “0” (Dokładność %)	Przewidywane “1” (Dokładność %)
0 (70%)	115 (74%)	24	267 (76%)	37
1 (30%)	41	18 (43%)	83	45 (55%)
Przewidywany % od całości	79%	21%	81%	19%

Tabela 6. Macierz błędów sieci neuronalnej z 24 zmiennymi

Dążąc do dalszych ulepszeń, inne typy NN zostały przetestowane z modelem Deep Learning (DL), dając interesujące rezultaty. Proces, zwany autokodowaniem, przedstawionym na rysunku 6, często używany dla detekcji anomalii. Jest to symetryczna feedforward Sieć Neuronowa z celem rekonstrukcji swoich wejść zamiast predykcji wartości celu. Dla tego badania, użyto modelu z bazą danych Training składającego się tylko z obserwacji, w których zmienna zależna wynosiła 0. Został potem wytrenowany wejściami równymi wynikom w celu nauczenia się jak wygląda „normalny” proces. Potem, “1” z setów Training i Testing zostały wprowadzone do modelu i zostały zmierzone ich granice błędów.



Rysunek 7. Diagram Typowego Autoencodera

Poniżej znajduje się rekonstrukcja błędów z podziałem na klasę dla Autoencodera z czterema gęstymi warstwami i dwudziestoma zmiennymi.



Rysunek 8. Błąd Rekonstrukcji Autoencodera z 24 zmiennymi

Najlepsze wyniki z tego modelu osiągnięte zostały używając odcięcia Błędu Rekonstrukcji o wartości 0.02. Prawie 20% obserwacji leżało poniżej linii odcięcia z proporcją "1" wynoszącą tylko 8%. Wciąż miał 35% celności w rozpoznawaniu "1" w pozostałej populacji, ale celność rozpoznawania "0" poniżej linii jest najlepszą znaną w tej analizie.

Ocena wydajności modeli:

Pośród wszystkich metod klasyfikacji, tylko Sieci Neuronowe wyprodukowały rezultaty z celnością indentyfikacji "1" większą niż 40%, podczas predykcji 20% obserwacji jako "1" i celności w indentyfikacji "0" ponad 90% dla 20% obserwacji. Model z 123 zmiennymi osiągnął 50%+ celności w zestawach Training i Testing, ale proporcja przewidzianych "1" wynosiła 11% i 14%. Z powodu Sieci Neuronalnych będących metodą „Black Box”, zrozumienie ważności zmiennych nie może zostać osiągnięte.

Walidacja założeń

Nie zrobiono żadnych założeń o formie prawdziwej funkcjonalnej zależności oraz oszacowania formy funkcji. Te modele nie potrzebowały liniowej zależności pomiędzy zależną i niezależną zmienną. W rzeczywistości, te modele wspierały wiele typów zależności. Również, zmienne nie musiały być heteroscedastyczne dla każdego poziomu zmiennych niezależnych. Zamiast tego, analiza skupiła się tylko na możliwościach predykcyjnych modeli.

Walidacja Modeli

Tradycyjne metody Weryfikacji i Walidacji (V&V), takie jak cross-walidacja nie mogą być zastosowane do Sieci Neuronowych z powodu ich struktury „Black Box”⁹. Jediną walidacją modelu stworzonego z użycia zbioru danych Testing było zastosowanie wstrzymania zbioru danych Testing.

Podejście Szeregów Czasowych

Rozumowanie dla zdefiniowania predykcji w macierzy pomyłek jest następująca:

- Jeżeli predykcja jest ponad celem cenowym a faktyczna wchodzi ponad cel cenowy, jest prawdziwie pozytywna.
- Jeżeli predykcja jest ponad celem cenowym a faktyczna wchodzi poniżej cel cenowy, jest fałszywie pozytywna.
- Jeżeli predykcja jest poniżej celu cenowego a faktyczna wchodzi poniżej celu cenowego, jest prawdziwie negatywna.
- Jeżeli predykcja jest poniżej celu cenowego a faktyczna wchodzi powyżej celu cenowego, jest fałszywie negatywna.

Model został po raz pierwszy wytrenowany ze wszystkimi danymi poczynając w 1975 do pierwszego sygnału w 2005. Potem, został odbudowany dla każdego sygnału do końca 2014 i została przeprowadzona Walidacja Postępowa używając danych z 2015-2019. Każdy komponent Trendu i Regresji został użyty osobno, potem zostały rankingowane na podstawie mocy predykcyjnej. Dla ostatecznego modelu najlepszy komponent Trend został najpierw wybrany, następnie komponenty Trend zostawały dodane do momentu, kiedy model przestał się ulepszać. Następnie komponenty Regresja zostały dodane w ten sam sposób.

Ostateczny model składa się z dwóch komponentów trendowych (Trend Lokalny Ceny Zamknięcia i Trend Lokalny Ceny Wysokiej) i dwóch komponentów regresyjnych (opóźnienie 3-dniowe Ceny Zamknięcia i 3-dniowe opóźnienie Ceny Wysokiej). Celność predykcji dla okresów Training i Walidacja jest znacznie wyższa niż dla wszystkich innych modeli wcześniej branych pod uwagę. Tabela 9. Bezpośrednie porównanie jest jednak niemożliwe, z powodu różnic w podejściach i tego, jak zmienna zależna jest użyta dla modelowania.

Zdarzenie (Wybór Losowy)	Testing		Training	
	Przewidywane "0" (Dokładność %)	Przewidywane "1" (Dokładność %)	Przewidywane "0" (Dokładność %)	Przewidywane "1" (Dokładność %)
0 (70%)	33 (86%)	4	99 (92%)	10
1 (30%)	5	16 (75%)	9	30 (75%)
Przewidywany % od całości	66%	34%	73%	27%

Tabela 9. Macierz błędów modelu BSTS z 4 składnikami

⁹ Tim Menzies, Charles Pecheur. Verification and Validation and Artificial Intelligence. *Advances in Computers*, 2005,65.

Opisowa Analiza Wyników

Jak wyjaśnialiśmy w naszej sekcji analizy opisowej, związek między zależną i niezależnymi zmiennymi jest słaby i nieliniowy. To przeszkodziło nam w zrozumieniu ważności pewnych zmiennych predykcyjnych.

Wyniki Modelowania

Moc predykcyjna danych historycznych:

Przez konstrukcję wielu modeli predykcyjnych używających różnych technik uczenia maszynowego, dane cen historycznych pokazały, że mają średnią moc predykcyjną dla tworzenia się nowych trendów wzrostowych.

Najlepszy model Sieci Neuronowych z powodzeniem przewidział 35% wszystkich przypadków "1" w zestawie Training i 31% w zestawie Testing. W tym przypadku, informacja o historycznej cenie nie może być użyta jako jedyny predyktor w tworzeniu nowego trendu wzrostowego. Jednakże, może być użyty jako poradnik do potencjalnej zyskowności i pomóc z decyzjami, jak wiele kapitału ulokować na każdą transakcję.

Podejście serii czasowych przewidziało ponad 70% udanych transakcji i może być użyte jako podstawa dla strategii handlowej, ale jej wydajność musi być monitorowana w celu zapewnienia ciągłej celności.

Znaczenie zmiennych

Dwadzieścia cztery zmienne wybrane używając metodologii IV i WOE wydawały się mieć podobne, jeżeli nie lepsze, siły predykcyjne jak sto dwadzieścia trzy zmienne. Metodologie IV i WOE mogą więc być użyte niezależnie od natury związku między przewidywanymi zmiennymi i predyktorami. Indywidualne rezultaty wybranych 24 zmiennych nie są możliwe do uzyskania z powodu natury "Black Box" Sieci Neuronowych.

PODSUMOWANIE

Ta analiza spotkała się z wieloma wyzwaniami. Nie było istniejących badań, a użyte metody były eksperymentalne. Choć celność ponad 40% jest na pewno lepsza niż Wybór Losowy z 30%, ogólna ilość przypadków nie jest wystarczająco wysoka, żeby model mógł zostać użyty jako jedyne narzędzie decyzyjne.

To badanie pomaga zrozumieć relację między danymi historycznymi i tworzeniem się nowych trendów. Pomimo małego zbioru danych, moc predykcyjna jaką historyczne ceny mają na nowe trendy została udowodniona. Model może być używany jako poradnik do lokowania kapitału przez handlarzy.

Najlepszy model binarny z tego badania osiągnął celne predykcje 33% zdarzeń z celnością 43% a najlepszy model Szeregów Czasowych osiągnął celne predykcje 75% zdarzeń z celnością 75%. W oparciu o te

miary, możemy odrzucić hipotezę, że historyczne ceny mogą zostać użyte do przewidzenia 90% zdarzeń z celnością ponad 50%.

Praktyczne wykorzystanie

Algorytmy uczenia maszynowego w tym badaniu mogą zostać użyte jako rozumowanie do alokacji kapitału. Modele Autoencoder mogą zostać użyte do identyfikowania okoliczności, w których możliwość sukcesu jest ekstremalnie niska (8%) i albo wykluczyć je z rozważań, albo użyć nisko-ryzykownych strategii takich jak strategia kalendarzowa. Jeżeli obserwacja nie jest wykluczona przez model Autoencodera, NN może zostać użyte do alokacji kapitału na bazie prawdopodobieństwa sukcesu. W obecnej formie, obserwacje z obliczonym prawdopodobieństwem poniżej 0.5 mają 26% szansy na sukces, gdzie te z prawdopodobieństwem powyżej 0.5 mają 42% szansy. Znaczy to, że szansa na obserwacje z grupy prawdopodobieństwa 0.5+ do bycia zyskownymi jest 1.6 razy wyższa. Do osiągnięcia zbalansowanego podejścia do ryzyka, handlarz może kupić dwa kontrakty za każdym razem prawdopodobieństwo jest poniżej 0.5 i trzy kontrakty, kiedy jest ponad 0.5.

Model szeregów czasowych BSTS może zostać użyty do podejmowania decyzji inwestycyjnych bieżąco, poprzez wykonywanie transakcji za każdym razem, gdy sugeruje je model.

Dalszy Rozwój

Badanie może być rozszerzone to zbadania tego samego podejścia dla innych kontraktów Futures. Dalsze badania mogą również dalej eksplorować strukturę warstw ukrytych Sieci Neuronowych i dostosowywać ich hiperparametry. Z powodu natury zbioru danych będącej seriami czasowymi, podejścia serii czasowych z użyciem Sieci Neuronowych z Pamięcią Długotrwałą i dalsze rozszerzenia prawdopodobieństwa subiektywnego powinny być zbadane.

BIBLIOGRAFIA

- A Review of Warren Buffett's Investing Strategy*. (2019). Retrieved from <https://www.investopedia.com/investing/warren-buffetts-investing-style-reviewed/>
- Balsara, Nauzer. (1992). *Money Management Strategies for Futures Traders*. New York: John Wiley & Sons, Inc.
- Bernstein, W. (2010). *The four pillars of investing*. New York: McGraw Hill.
- Best Forex Scalping Strategies*. (2019) Retrieved from <https://www.theforexarmy.com/forex-scalping-strategy/>
- Bollerslev T. (1986), Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa.
- Charles Duchock, (1990), "Evidence of Efficiency in United States Futures Oil Prices," dissertation, United States International University.
- Chew, D. H. *Corporate Risk Management*. Columbia University Press, 2008.
- Competitivefutures.com. (2019). [online] Available at: <https://www.competitivefutures.com/wp-content/uploads/2015/07/Case-Shiller-Q1-2013.jpg> [Accessed 5 Jun. 2019].
- Cox, D. R. (1955). "Some Statistical Methods Connected with Series of Events". *Journal of the Royal Statistical Society*. 17 (2): 129–164. doi:10.2307/2983950
- Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
- Dezfouli, A. and Bonilla, E. V. (2015). Scalable inference for Gaussian process models with black-box likelihoods. In *Advances in Neural Information Processing Systems*.
- Di Lorenzo, R. (2012). *How to Make Money by Fast Trading*. New York: Springer.
- Doerr C., Blenn N., and Van Mieghem P. (2013). Lognormal infection times of online information spread. *PLOS ONE*, 8.
- Engle R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1006.

- Futuresforecasts.com. (2019). *eMini S&P500 Futures Prices & Chart Forecasts*. [online] Available at: <http://www.futuresforecasts.com/emini-sp-500-futures.html> [Accessed 5 Jun. 2019].
- Gannon, N. (2010). *Investing strategies for the high net-worth investor*. New York, NY: McGraw-Hill.
- Gelman A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016) *Deep Learning*. MIT Press. p. 196.
- Granger C.W.J., Swanson N.R. (1997), An Introduction to stochastic unit-root process, *Journal of Econometrics*, 80, 35-62.
- Jones C.R., Marriott J.M. (1999), A Bayesian analysis of stochastic unit root models, *Bayesian Statistics*, 6, s. 785-794.
- Jorgenson, Dale. *Capital Theory and Investment Behavior*. *American Economic Review* 53, no. 2 (1963): 247-259.
- Larsen, Kim. (2016, April 07). Uplift Models: Optimizing the Impact of Your Marketing. Retrieved July 15, 2019, from: https://www.predictiveanalyticsworld.com/sanfrancisco/2016/uplift_modeling.php
- LeBaron, D., & Vaitilingam, R. (2002). *Dean LeBaron's treasury of investment wisdom*. New York: Wiley.
- Lee, S. H., Crawford, M. M., and Wilson, J. R. (1991). Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. *Communications in Statistics Simulation*, 20(2):777–809.
- Leybourne S.J., McCabe B.P.M., Mills T.C. (1996), Randomized unit root processes for modelling and forecasting financial time series: theory and applications, *Journal of Forecasting*, 15, 253-270.
- Leybourne S.J., McCabe B.P.M., Tremayne A.R. (1996), Can economic time series be differenced to stationarity? *Journal of Business and Economic Statistics*, 14, 435-446.
- Lloyd, C., Gunter, T., Osborne, M. A., and Roberts, S. J. (2015). Variational inference for Gaussian process modulated Poisson processes. *International Conference on Machine Learning*.
- Luo, Z.; Cai, X.; Tanaka, K.; Takiguchi, T.; Kinkyō, T.; Hamori, S. Can We Forecast Daily Oil Futures Prices? Experimental Evidence from Convolutional Neural Networks. *J. Risk Financial Manag.* **2019**, *12*, 9.
- Lynn, G. (2006). *A dictionary of derivatives and financial engineering*. Oakland, CA: Mason Pub.
- Markham, Jerry W., (1987), *The History of Commodity Futures Trading and Its Regulation*, Praeger

Redhead, Keith (1997). *Financial Derivatives: An Introduction to Futures, Forwards, Options and Swaps*. London: Prentice-Hall

Research.cs.aalto.fi. (2019). *GPstuff - log Gaussian Cox process demo*. [online] Available at: https://research.cs.aalto.fi/pml/software/gpstuff/demo_lgcp.shtml [Accessed 5 Jun. 2019].

Ross, J. (2004). *Day trading*. Cedar Park, TX: Ross Trading, Inc.

Scalping: Small Quick Profits Can Add Up. (2019). Retrieved from <https://www.investopedia.com/articles/trading/05/scalping.asp>

Schap, K. (2005). *The complete guide to spread trading*. New York: McGraw-Hill.

Simma, A. and Jordan, M. I. (2010). Modeling events with cascades of poisson processes. In UAI, pages 546–555.

Sollis R., Leybourne S.J., Newbold P. (2000), Stochastic unit roots modelling of stock price indices, *Applied Financial Economics*, 10, 311-315.

Tim Menzies, Charles Pecheur. Verification and Validation and Artificial Intelligence. *Advances in Computers* , **2005**,65,153-201.

Titsias, M. K. (2009). Variational learning of inducing variables in sparse Gaussian processes,. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics.

Tong H. (1990), *Non-linear time series: a dynamical system approach*, Oxford University Press, Oxford.

Unofficialgoogledatascience.com. (2019). *Fitting Bayesian structural time series with the bsts R package*. [online] Available at: <http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html> [Accessed 5 Jun. 2019].

Van Blokland, P. (1995). *Using the futures market to lower the farm management risks of producing for unknown market prices in the USA*. Gainesville, FL: University of Florida, Institute of Food and Agricultural Sciences, Food and Resource Economics Dept.

Vanhatalo J., Riihimaki J., Hartikainen J., Jylanki P., Tolvanen V., Vehtari A. (2013). Gpstuff: Bayesian modeling with gaussian processes. *J. Mach. Learn. Res.*, 14(1):1175–1179.