

41st Conference
Multivariate Statistical Analysis (MSA'2023)



Organizing Committee / Komitet organizacyjny

Organizing Committee Chair / Przewodnicząca Komitetu Organizacyjnego

Dr hab. Alina Jędrzejczak, prof. UŁ

Scientific Secretaries / Sekretarze naukowci

Dr Kamila Trzcńska

Dr Elżbieta Zalewska

University of Łódź

Uniwersytet Łódzki

Faculty of Economics and Sociology Wydział Ekonomiczno-Socjologiczny

Department of Statistical Methods Katedra Metod Statystycznych

Scientific Committee / Komitet naukowy

Scientific Committee Chair / Przewodniczący Komitetu Naukowego

Prof. dr hab. Czesław Domański, Uniwersytet Łódzki

Scientific Committee Members / Członkowie Komitetu Naukowego

Prof. Francesca Greselin,
University of Milano Bicocca

Prof. Marie Huskova,
Charles University in Prague

Prof. dr hab. Krzysztof Jajuga,
Uniwersytet Ekonomiczny we Wrocławiu

Prof. dr hab. Mirosław Krzyśko,
Uniwersytet im. Adama Mickiewicza w Poznaniu

Prof. dr hab. Włodzimierz Okrasa,
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

Prof. dr hab. Józef Pociecha,
Uniwersytet Ekonomiczny w Krakowie

Prof. dr hab. Mirosław Szreder,
Uniwersytet Gdański

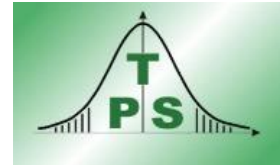
Prof. dr hab. Janusz Wywiół,
Uniwersytet Ekonomiczny w Katowicach

Co-Organizers / Współorganizatorzy

**Institute of Statistics and Demography
of the University of Łódź /
Instytut Statystyki i Demografii
Uniwersytetu Łódzkiego**



**Polish Statistical Association, Branch in Łódź /
Polskie Towarzystwo Statystyczne, Oddział w Łodzi**



Honorary Patronage / Patronat honorowy

Rector of University of Łódź / Rektor Uniwersytetu Łódzkiego



Patronat Rektora
Uniwersytetu Łódzkiego

**The President of Statistics Poland /
Prezes Głównego Urzędu Statystycznego w Polsce**



**Polish Academy of Sciences /
Polska Akademia Nauk**



Partner / Partner

StatSoft Polska Sp. z o. o.

StatSoft Polska

LabMasters



Bank Pekao



Invited Lectures / Wykłady zaproszone

Przemysław Biecek, Warsaw University of Technology & University of Warsaw

Title: Explanatory Models Analysis

Abstract:

Anscombe's quartet is a great illustration of the need to enrich models with data visualisation. Through a simple example, it shows that data visualisation is very useful, or even essential, in the statistical analysis of data.

Today, after 50 years, we are witnessing a real revolution in the number and variety of predictive models that are being developed for the most diverse applications. These models are increasingly complex with a number of coefficients that effectively make them impossible to understand. Problems with understanding a model occur with as few as 10 coefficients, but today's DNN models often have thousands, millions or billions of coefficients.

In this talk, I will present the reasons why it is useful to visualise predictive models not only to understand them, but also to improve them. Then I will discuss situations when this is needed, most often enforced by emerging regulations such as the AI-act being discussed in the European Union. Then I will present selected tools that can be used to visualise predictive models, which are increasingly being developed under the umbrella of eXplainable Artificial Intelligence (XAI). These techniques can be used to visualise a single model, a pair of models (e.g. during champion-challenger analysis) or a larger set of models (e.g. during Rashomon set analysis).

Francesca Greselin, University of Milano-Bicocca

Title: Measuring income inequality via percentile relativities

Abstract:

"The rich are getting richer" implies that the population income distributions are getting more right skewed and heavily tailed. For such distributions, the mean is not the best measure of the center, but the classical indices of income inequality, including the celebrated Gini index, are all mean-based. In view of this, Professor Gastwirth sounded an alarm back in 2014 by suggesting to incorporate the median into the definition of the Gini index, although noted a few shortcomings of his proposed index. In the present paper we make a further step in the modification of classical indices and, to acknowledge the possibility of differing viewpoints, arrive at three median-based indices of inequality. They avoid the shortcomings of the previous indices and can be used even when populations are ultra-heavily tailed, that is, when their first moments are infinite. The new indices are illustrated both analytically and numerically using parametric families of income distributions, and further illustrated using capital incomes coming from 2001 and 2018 surveys of fifteen European countries. We also discuss the performance of the indices from the perspective of income transfers.

Title: Reference points in multiobjective, multistage decision making

Abstract:

Multiple-criteria decision-making (MCDM) is a sub-discipline of operations research that explicitly evaluates multiple conflicting criteria in decision making. One branch of MCDM is multiobjective dynamic programming. In the presentation multiobjective, multistage decision processes are taken into consideration. Vector multistage criterion function consists of multistage components. Multistage components consist of stage criteria. Process realization a^* is Pareto optimal, if it does not exist any other process realization which is better than a^* . Such a definition is widely applied in MCDM.

There are some shortcomings when applying such an approach to control multistage multiobjective decision processes. Decision makers are often interested not only in multistage values of criteria, but stage values of criteria are also equally or even more important for them. This is the motivation for my up-to-date research related to application of reference points to the analysis of these processes.

Reference points are not new in MCDM. The first step was looking for solutions which are the closest to the ideal point or another point in the decision space pointed by a decision maker. Such an approach can be found in the compromise programming as well as in the goal programming. Applying ideal point method we often find solutions which are not the same as solutions which are the most distant from the anti-ideal (nadir) point. This is the reason why ideal and anti-ideal solutions were considered simultaneously. The next step was to use two sets of reference points – the first of them consists of “good” objects and the second one consists of “bad” objects. Such an approach can be named Bipolar approach.

Multistage Bipolar Method, which is original MCDM method derived for control multiobjective, multistage decision processes. Stage realizations of the process are taken into considerations as well as the multistage realizations. The decision maker is asked to give stage sets of good and bad objects. The method consists of the following steps:

Step 1. Stage process realizations are compared with reference objects. Well known procedures, as Electre, Promethee and APH can be employed. Then the position of each process realization with respect to the bipolar stage reference system is established.

Step 2. Relationships in the set of multistage process realizations are established.

Step 3. Multistage process realizations are assigned to predefined classes.

Step 4. The final solution is pointed from the multistage process realization belonging to the class with the smallest number.

Step 5. The procedure of testing Pareto-optimality for the obtained realization is performed. If this solution is not Pareto-optimal, the set of better Pareto optimal solutions is generated. It is up to the decision maker to make a final choice.

Papers / Referaty

Sesja IIIa

Elżbieta Gołata, Uniwersytet Ekonomiczny w Poznaniu

Tytuł: Konsekwencje pomiaru ludności w ocenie płodności kobiet w Polsce

Streszczenie: Celem referatu jest przedstawienie konsekwencji pomiaru ludności dla oceny płodności kobiet w Polsce. Dodatkowym zamierzeniem jest ukazanie wpływu przyjętych definicji ludności dla oceny zróżnicowania płodności według wieku oraz w ujęciu terytorialnym. W badaniu wykorzystano dane GUS, Eurostatu oraz Human Fertility Database. Kontynuacja innych badań i analiz demograficznych posłużyły do wskazania zmiany wzorców płodności kolejnych generacji.

Badanie przeprowadzono w kontekście oceny jakości wyników powszechnego spisu ludności i mieszkań 2021 oraz Rozporządzenia Komisji Europejskiej ws. europejskich statystyk ludności i mieszkań. W badaniu wykorzystano dane GUS, Eurostatu oraz Human Fertility Database.

W badaniu wykorzystano metody analizy demograficznej, przekrojowej i kohortowej. Odwołano się do podstawowych zasad konstrukcji współczynników demograficznych w ujęciu poprzecznym oraz wzdłużnym. Przeliczono wszystkie cząstkowe współczynniki płodności wykorzystując szacunek liczby ludności rezydującej na podstawie wyników NSP 2021.

Rezultaty szacunków wskazują że współczynnik dzietności kobiet w Polsce, przy przyjęciu krajowej definicji ludności, w 2021 roku był zaniżony o 4,3%. Największe zniżenie w skali kraju dotyczyło kobiet w wieku 35-39 lat (6,3%), a w przekroju województw - zamieszkujących w Podkarpackim (11%).

Grażyna Trzpiot, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Wybrane środowiskowe implikacje na jakość życia mieszkańców miast

Streszczenie: Artykuł ma na celu wskazanie wybranych środowiskowych implikacji dla jakości życia mieszkańców miast. Zwrócono uwagę na zależność między zanieczyszczeniem powietrza a przedwczesnymi zgonami w Unii Europejskiej i w Polsce. Następnie na zależność pomiędzy hałasem a zdrowiem psychicznym i fizycznym mieszkańców miast oraz na zagrożenia związane z klimatem a jakością zdrowia mieszkańców miast. Zmiany klimatyczne a wrażliwość populacji w wybranych polskich miast zbadano wyznaczając miary wrażliwości demograficznej dużych miast w Polsce.

Słowa kluczowe: miary wrażliwości demograficznej, demografia miast, środowisko

Title: Selected environmental implications on the quality of life of urban inhabitants

Abstract: The article has the aim to point out selected environmental implications on the quality of life of urban residents. Attention is given to the relationship between air pollution and premature deaths in the European Union and Poland. Then on the relationship between noise and the mental and physical health of urban residents, and on the risks of climate change and the quality of health of urban residents. Climate change and population sensitivity in selected Polish cities was examined by determining measures of demographic sensitivity of large cities in Poland.

Keywords: measures of demographic sensitivity, urban demographics, environment

Tytuł: Wielowymiarowa ocena warunków życiowych i dobrobytu ludzi starszych w Unii Europejskiej

Streszczenie: Jednym z głównych problemów demograficznych jest starzenie się społeczeństw. Dotyczy to w szczególności rozwiniętych państw europejskich. Proces ten jednak nie przebiega we wszystkich krajach jednakowo i ma różny wpływ na ich rozwój społeczno-gospodarczy. Celem badania jest ocena i porównanie warunków życiowych i dobrobytu ludzi starszych w krajach europejskich. W badaniu wykorzystano dane z Eurostatu. Analizę przeprowadzono w latach 2015 oraz 2020 na podstawie dziewięciu wskaźników. Warunki życiowe i dobrobyt ludzi starszych oceniono za pomocą zmiennej syntetycznej, utworzonej z wykorzystaniem metod skalowania wielowymiarowego. Na jej podstawie zbudowano ranking państw UE w obu badanych latach. Dodatkowo, za pomocą metody k-średnich dokonano grupowania państw według wartości analizowanych wskaźników. W obu badanych latach najwyższe pozycje w rankingach zajmowały Irlandia i Luksemburg. Wysoką pozycję w rankingach zajmowały też Austria, Holandia, Finlandia i Szwecja. Z drugiej strony, warunki życia i dobrobyt ludzi starszych były najniższe w Bułgarii, Grecji, Rumunii i Chorwacji. Sytuacja ludzi starszych poprawiła się w największym stopniu w Niemczech, a najbardziej pogorszyła się w Danii i Hiszpanii. Wyniki analizy skupień pokazują, że dysproporcje pomiędzy państwami o najlepszej i najgorszej sytuacji ludzi starszych nie ulegają zmniejszeniu. Sześć z dziewięciu analizowanych wskaźników uległo w badanych pięciu latach poprawie, natomiast trzy (odsetek osób starszych zagrożonych ubóstwem, dysproporcja pomiędzy dochodami najbogatszych i najbiedniejszych grup osób starszych oraz odsetek osób starszych w ludności ogółem) uległy pogorszeniu. Dlatego polityka społeczno-ekonomiczna powinna dotyczyć w największym stopniu tych właśnie wskaźników.

Słowa kluczowe: starzenie się społeczeństw, skalowanie wielowymiarowe, porządkowanie liniowe, analiza skupień

Title: Multivariate assessment of the living conditions and well-being of elderly people in the European Union

Abstract: One of the main demographic problems is the ageing of populations. This applies in particular to highly developed European countries. However, this process is not taking place equally in all countries and has a different impact on their socio-economic development. The aim of the study is to assess and compare the living conditions and well-being of elderly people in the European Union countries. The study used data from Eurostat. We carried out the analysis for years 2015 and 2020 on the basis of nine indicators. We assessed the living conditions and well-being of older people by using a synthetic variable created by applying the multidimensional scaling technique. On its basis, we constructed rankings of the EU countries in the two analysed years. In addition, by using the k-means method, we clustered the countries with respect to the values of the analysed indicators. Ireland and Luxembourg held the highest rankings in both analysed years. Austria, the Netherlands, Finland and Sweden were also high in the rankings. On the other hand, the living conditions and well-being of older people were the lowest in Bulgaria, Greece, Romania and Croatia. The situation of elderly people improved the most in Germany, and deteriorated the most in Denmark and Spain. The results of the cluster analysis show that the disparity between the countries with the best and worst situations of elderly people has not declined. Six of the nine indicators analysed improved over the analysed five years, while three (the proportion of elderly people at risk of poverty, the disparity between the incomes of the richest and poorest groups of elderly people and the proportion of elderly people in the total population) worsened. Socio-economic policies should therefore address these indicators to the greatest extent possible.

Keywords: population ageing, multidimensional scaling, linear ordering, cluster analysis

Tytuł: Estymacja nieznanymi miar obiektów w chemicznym układzie wagowym ze skorelowanymi błędami pomiarów

Streszczenie: W pracy przedstawiony został problem wyznaczania nieznanymi miar obiektów w modelu chemicznego układu wagowego. Układy te są rozpatrywane przy założeniu, że błędy eksperymentalne są jednakowo skorelowane i mają takie same wariancje. Zależności między parametrami układów wagowych są rozpatrywane z punktu widzenia kryteriów optymalności. Badane są układy, w których iloczyn wariancji estymatorów jest możliwie najmniejszy, tj. układy D-optymalne. Określone zostaną warunki, przy spełnieniu których rozważany układ jest D-optymalny lub wysoce D-efektywny w danej klasie. Podane zostaną warunki konieczne i dostateczne zapewniające istnienie tych układów oraz przykłady metod ich konstrukcji.

Słowa kluczowe: chemiczny układ wagowy, kryterium D-optymalności

Sesja IIIb

Jacek Wesołowski, GUS i Politechnika Warszawska

Title: Discrete parametric statistical graphical models

Abstract: Typically statistical graphical models are (1) either continuous and then Gaussian, that is purely parametric, or (2) discrete and then purely non-parametric. In this talk we introduce discrete graphical models of negative multinomial or multinomial type which are purely parameteric, which allows to reduce considerably size of unknown variables. We also discuss the Bayesian counterpart by introducing a conjugate prior which is a graphical version of the popular Dirichlet distribution. The talk is based on a joint work with B. Kolodziejek (Politechnika Warszawska) and X. Zeng (Universite de Strasbourg).

Marcin Szymkowiak , Maciej Beręsewicz, Uniwersytet Ekonomiczny w Poznaniu, Urząd Statystyczny w Poznaniu

Tytuł: Wnioskowanie dla prób nielosowych z wykorzystaniem estymatorów kalibracyjnych wartości globalnych i kwantyli

Streszczenie: Głównym celem prezentacji jest próba łącznego wykorzystania kalibracji dla wartości globalnej zaproponowanej przez Deville'a i Särndala (1992) i kalibracji dla kwantyli wprowadzonej przez Harmsa i Duchesne'a (2006). Podejście to uwzględnia nie tylko równania kalibracyjne dla wartości globalnych, ale bierze także pod uwagę odpowiednie ograniczenia na znane kwantyle ciągłych zmiennych pomocniczych obserwowanych w zbiorze danych. Jest to podejście, które pozwala w jednym kroku poprawić oszacowania średnich, wartości globalnych i odpowiednich kwantyli dla interesującej nas zmiennej. W badaniu symulacyjnym z wykorzystaniem prób nielosowych wykazano, że proponowane podejście charakteryzuje się lepszymi własnościami od innych metod istniejących w literaturze (na przykład pseudo-empirical likelihood method, inverse probability weighting, doubly robust estimators).

Title: Inference for non-probability samples using calibration estimators for totals and quantiles

Abstract: In this presentation, we combine calibration for population totals proposed by Deville and Särndal (1992) with calibration for population quantiles introduced by Harms and Duchesne (2006). This approach extends the calibration equations for totals by adding relevant constraints on quantiles of continuous auxiliary variables observed in the data. It is a multipurpose solution, i.e. it makes it possible to improve

estimates of means, totals and quantiles for a-variable of interest in one step. A simulation study using nonprobability samples showed that the proposed approach has better properties than other methods existing in the literature (for instance pseudo-empirical likelihood method, inverse probability weighting, double robust estimators).

Andrzej Dudek, Wrocław University of Economics and Business

Tytuł: Przenoszenie stylu obrazu i dźwięku z wykorzystaniem sieci neuronowych – podstawy statystyczne

Streszczenie: Prezentacja przedstawia algorytmy przenoszenia stylu przy użyciu sieci neuronowych (NST), techniki łączącej algorytmy uczenia maszynowego, widzenie komputerowe i przetwarzanie dźwięku, która pozwala na połączenie artystycznego stylu jednego obrazu z treścią innego lub sklonowanie timbru głosu. Przenoszenie stylu przy użyciu sieci neuronowych zdobyło znaczną popularność ze względu na zdolność do tworzenia atrakcyjnych wizualnie i artystycznych obrazów. Prezentacja dostarcza przeglądu kluczowych algorytmów NST, w tym klasycznej pracy Gatys-a i innych w dziedzinie przenoszenia stylu przy użyciu sieci neuronowych oraz kolejnych postępów w tej dziedzinie, skupiając się na wyjaśnieniu statystycznych podstaw zastosowanych metod. Prezentacja jest ilustrowana przykładami dostosowanymi do tematu.

Słowa kluczowe: Głębokie uczenie, sieci neuronowe, neuronowy transfer stylu

Title: Image and Audio Neural Style Transfer – Statistical Foundations

Abstract: The presentation explores the field of neural style transfer (NST), a technique that combines deep learning algorithms, computer vision, and audio processing to merge the artistic style of one image with the content of another or clone voice timbre. Neural style transfer has gained significant attention due to its ability to create visually appealing and artistic imagery. The presentation provides an overview of the key NST algorithms, including Gatys et al.'s seminal work on neural style transfer, and subsequent advancements in the field. The focus is on explaining the statistical background of applied methods. The presentation is illustrated with custom examples.

Keywords: Deep learning, neural networks, neural style transfer

Jerzy Korzeniewski, Adam Idczak, Uniwersytet Łódzki

Tytuł: Algorytm ustalania sentymentu tekstów polskojęzycznych z wykorzystaniem tłumaczenia na język angielski

Streszczenie: Ustalanie sentymentu tekstu pisanego jest istotnym zadaniem w badaniu tekstu. To zadanie może być rozumiane dwojako tj. w wersji nadzorowanej oraz nienadzorowanej. W tym referacie zajmujemy się wersją nienadzorowaną i przedstawiamy nowy algorytm dla dokumentów polskojęzycznych. Grupowanie tekstów polskojęzycznych ze względu na ich sentyment jest słabo rozwinięte w polskiej literaturze przedmiotu. Nowatorskość proponowanego algorytmu obejmuje rezygnację ze stoplisty, lematyzacji, tłumaczenie tekstu na język angielski oraz dwuetapowe grupowanie dokumentów. Lemmatyzację, która jak się wydaje jest nieunikniona dla tekstów polskojęzycznych, można pominąć, gdyż tekst jest najpierw przetłumaczony na język angielski. W pierwszym kroku algorytmu niektóre dokumenty są przypisane do klasy dokumentów pozytywnych albo negatywnych w oparciu o zbiór reguł leksykalnych, gramatycznych jak również zbiór terminów kluczowych. Terminy kluczowe nie muszą być wprowadzane przez użytkownika, odnajduje je algorytm. W drugim kroku pozostałe dokumenty są dołączane do jednej z klas w oparciu o reguły wykorzystujące słownictwo występujące w dokumentach pogrupowanych pierwszym krokiem.

Słowa kluczowe: text mining, sentyment dokumentu, grupowanie dokumentów

Title: An algorithm for determining the sentiment of Polish-language texts using translation into English

Abstract: Determining the sentiment of a written text is an important task in text research. This task can be understood in two ways, i.e. in a supervised and unsupervised version. In this paper, we deal with the unsupervised version of the paper and present a new algorithm for Polish-language documents. The grouping of Polish-language texts based on their sentiment is poorly developed in Polish literature on the subject. The novelty of the proposed algorithm includes the abandonment of stoplists, lemmatization, translation of text into English and two-stage grouping of documents. Lemmatization, which seems to be inevitable for Polish-language texts, can be omitted because the text is first translated into English. In the first step of the algorithm, some documents are assigned to a class of positive or negative documents based on a set of lexical and grammatical rules as well as a set of key terms. Key terms do not have to be entered by the user, the algorithm finds them. In the second step, the remaining documents are attached to one of the classes based on rules using the vocabulary found in the documents grouped in the first step.

Keywords: text mining, document sentiment, document grouping

Sesja IVa

Agnieszka Sompolska-Rzechuła, Iwona Bąk, West Pomeranian University of Technology in Szczecin

Tytuł: Realizacja celów zrównoważonego rozwoju – Polska na tle krajów UE

Streszczenie: Celem artykułu jest dokonanie oceny stopnia realizacji celów zrównoważonego rozwoju (SDG) w Polsce w porównaniu z innymi krajami UE. Badanie przeprowadzono dla 27 krajów UE bazując na danych z Eurostatu i Raportu o Rozwoju Społecznym z lat 2015 i 2021. Podstawę oceny stanowi wartość zmiennej syntetycznej określonej dla poszczególnych SDG na podstawie cząstkowych wskaźników. Do budowy miary syntetycznej wykorzystano metodę TOPSIS. Zbadano także kształtowanie się wszystkich wskaźników cząstkowych dla Polski w odniesieniu do wyników krajów UE. Badanie pokazuje, że w Polsce w przypadku realizacji większości SDG i wskaźników cząstkowych nie odnotowano poprawy. Wartością dodaną artykułu jest ocena spełnienia założeń Agendy 2030 na podstawie dogłębnej analizy stopnia realizacji zarówno wskaźników cząstkowych, jak i głównych SDG.

Słowa kluczowe: realizacja SDG, metoda TOPSIS, Polska, kraje UE

Title: Implementation of sustainable development goals - Poland in comparison with EU countries

Abstract: The purpose of the article is to assess the degree of implementation of the Sustainable Development Goals (SDGs) in Poland compared to other EU countries. The study was conducted for 27 EU countries based on data from Eurostat and the Human Development Report of 2015 and 2021. The basis of the assessment is the value of the synthetic variable determined for each SDG on the basis of partial indicators. The TOPSIS method was used to build the synthetic measure. The development of all sub-indicators for Poland in relation to the performance of EU countries was also examined. The study shows that Poland's performance on most SDGs and sub-indicators has not improved. The added value of the article is the assessment of the fulfillment of the 2030 Agenda on the basis of an in-depth analysis of the degree of realization of both sub-indicators and main SDGs.

Keywords: realization of SDGs, TOPSIS method, Poland, EU countries

Tytuł: Zielone miejsca pracy w Polsce – analiza przekrojowo-czasowa

Streszczenie: W niniejszym opracowaniu dokonaliśmy identyfikacji kluczowych sekcji i grup zawodowych dla oszacowania potencjału tworzenia zielonych miejsc pracy w Polsce. Wykorzystaliśmy dane z krajowego rejestru podmiotów gospodarki narodowej (REGON) oraz dane z Badania Aktywności Ekonomicznej Ludności (BAEL). W proponowanej metodzie założyliśmy, że zmiany w REGON odzwierciedlają przybliżenie zmiany zasobu GJ w sektorach działalności gospodarczej (PKD-2007) i grupach zawodowych. Oceniliśmy trendy zielonego zatrudnienia w latach 2015-2022. Większość sektorów i wiele grup zawodowych odnotowała wzrost zielonego zatrudnienia, z wyjątkiem. m.in. rolnictwa. Co więcej spadające zatrudnienie w tym sektorze ma znaczący wpływ na zielony rynek pracy.

Słowa kluczowe: zielone miejsca pracy, sektorowy i zawodowy potencjał tworzenia zielonych miejsc pracy, sekcje PKD, KZiS, BAEL

Title: Green jobs in Poland - cross-sectional and time-series analysis

Abstract: In this study, we employed a stepwise empirical approach to identify economic sectors and occupational groups for green job (GJ) creation in Poland. We used the operating register of economic entities (REGON) and Polish Labor Force Survey (BAEL) data for the period 2015-2022. The changes in REGON reflect a proxy of changes in GJ stock in sectors of economic activity (PKD-2007). We estimated trends of green employment. Most sectors and many occupational groups noted increases in GJ, except for agriculture, where the downward trend in employment has a significant influence on the green labor market. Our findings may be useful when formulating policy recommendations for educational institutions, employment institutions, local governments, government institutions, investors, and employers.

Keywords: green jobs, sectoral and occupational potential for green job creation, PKD sections, ISCO, LFS

Marta Kuc-Czarnecka, Iwona Markowicz, Agnieszka Sompolska-Rzechuła, Alina Stundziene, Politechnika Gdańska, Uniwersytet Szczeciński, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Kaunas University of Technology

Tytuł: Czynniki utrudniające i przyspieszające realizację SDG7 w krajach UE

Streszczenie: Czysta i przystępna cenowo energia jest jednym z celów zrównoważonego rozwoju (SDG7), który jest powiązany z większością pozostałych SDG i odgrywa kluczową rolę w rozwoju gospodarczym i dobrobycie człowieka. Celem artykułu jest wskazanie stymulatorów i przeszkód w realizacji SDG7. Badanie przeprowadzono dla 27 krajów UE bazując na danych z Eurostatu i Raportu o Rozwoju Społecznym z lat 2013-2020. Zidentyfikowano trzy kategorie potencjalnych czynników wpływających na stopień realizacji SDG7: ekologiczne, społeczne i ekonomiczne. Zaproponowane podejście oraz wykorzystanie analizy wrażliwości do przypisywania zmiennych wag i konstruowania rankingów krajów UE stanowi nowatorski aspekt pracy i wypełnia lukę w badaniach nad SDG7. Do realizacji celu badania wykorzystano także modele panelowe. Badanie pokazało, że dwa kraje: Dania i Szwecja zajmują czołowe miejsca w rankingach pod względem stopnia realizacji SDG7 natomiast Bułgaria, Cypr i Litwa zajmowały jedne z ostatnich pozycji. Nasze badania wykazały, że w każdym oszacowanym modelu wystąpiły te same zmienne znacząco wpływające na realizację SDG7. Spośród tych zmiennych jedynie stopa bezrobocia miała istotny ujemny wpływ na stopień realizacji SDG7.

Słowa kluczowe: SDG7; czynniki czystej energii, bariery czystej energii, analiza wrażliwości, model danych panelowych

Title: Factors hindering and boosting SDG7 implementation in EU countries

Abstract: Clean and affordable energy is one of the sustainable development goals (SDG7), which is related to most other SDGs and plays a crucial role in economic development and human well-being. The article aims to indicate the boosters and obstacles of SDG7 implementation. The study was conducted for 27 EU countries based on data from Eurostat and the Human Development Report from 2013 to 2020. Three groups of potential SDG7 determining factors have been identified: ecological, social and economic. The proposed approach and the use of sensitivity analysis to variables weighting and ranking constructions of EU countries is an innovative aspect of the work and fills the gap in research on SDG7. The study showed that Denmark and Sweden occupy leading positions in the rankings in terms of the degree of implementation of SDG7. In contrast, Bulgaria, Cyprus and Lithuania occupied one of the last positions. The results of panel-data model estimations showed that in each estimated model, the same 'indispensable variables' significantly affect the implementation of SDG7. Among these variables, only the unemployment rate significantly negatively impacted the SDG7 execution.

Keywords: SDG7; clean energy boosters, clean energy obstacles, sensitivity analysis, panel-data model

Sesja IVb

Artur Zaborski, Marcin Pełka, Uniwersytet Ekonomiczny we Wrocławiu

Tytuł: Analiza unfolding obiektów symbolicznych na przykładzie zewnętrznej oceny reklam samochodów

Streszczenie: Wielowymiarowa analiza unfolding pozwala na przedstawienie zarówno kolumn (np. produktów, usług), jak i wierszy (np. klientów) macierzy preferencji na tej samej mapie percepcyjnej (zwykle jest to mapa dwu- lub trójwymiarowa). W klasycznym podejściu wiersze i kolumny macierzy preferencji są przedstawiane na mapie percepcyjnej w postaci punktów. Jednakże taką analizę można również przeprowadzić dla symbolicznych danych o wartościach przedziałowych

Celem artykułu jest przedstawienie propozycji przeprowadzenia analizy unfolding dla obiektów symbolicznych. W części empirycznej wykorzystano hybrydowe podejście analizy unfolding dla obiektów symbolicznych do oceny wybranych reklam samochodów.

Słowa kluczowe: symboliczna analiza danych, analiza unfolding, pomiar preferencji, reklamy samochodów

Title: The unfolding analysis for symbolic objects on the example of the external car advertisement evaluation

Abstract: Multidimensional unfolding allows representing both columns (e.g. products, services) and rows (e.g. customers) of the preference matrix on the same low-dimensional map (usually it's a two- or three-dimensional map). In the classical multidimensional unfolding, rows and columns are represented by points. However, such unfolding analysis can be done for symbolic interval-valued data.

The aim of the article is to present suggestions how to perform unfolding analysis for symbolic objects. In the empirical part, a hybrid approach of unfolding analysis for symbolic objects was used to evaluate selected car advertisements.

Keywords: symbolic data analysis, unfolding analysis, preference measurement, car advertisements

Tytuł: Wiedza ekspercka w ekonometrycznych modelach wyceny nieruchomości

Streszczenie: W artykule zbadano czy wiedza ekspercka poprawia wyniki estymacji modeli wyceny nieruchomości. Porównano sześć modeli ekonometrycznych: KMNK, model estymacji mieszanej (mixed estimation model), model bayesowski, model z restrykcjami dla parametrów w postaci nierówności (IRLS), regresję grzbietową i LASSO (z regularyzacją). W trzech modelach wykorzystano wiedzę ekspercką (model estymacji mieszanej, bayesowski i z restrykcjami). W modelu estymacji mieszanej i bayesowskim wiedza a priori (eksperska) przyjęła postać przedziałów dla parametrów modelu. W modelu IRLS wiedza ekspercka warunkowała ograniczenia w postaci nierówności. W przykładzie empirycznym omawiane modele zastosowano przy wycenie gruntów niezabudowanych pod zabudowę mieszkaniową. Modele z wiedzą ekspercką okazały się lepsze z punktu widzenia zgodności z teorią. Cechowały się również lepszymi własnościami predyktywnymi. Reasumując, w przypadku danych o niskiej jakości wiedza ekspercka może poprawiać wyniki estymacji ekonometrycznych modeli wyceny nieruchomości.

Słowa kluczowe: ekonometryczne modele wyceny nieruchomości, wiedza ekspercka, modele z restrykcjami dla parametrów w postaci nierówności, estymacja mieszana, regresja bayesowska, regresja grzbietowa, LASSO

Title: Expert knowledge in econometric models of real estate appraisal

Abstract: The article examines whether expert knowledge improves the estimation results of real estate appraisal models. Six econometric models were compared: OLS, mixed, the Bayesian model, the Inequality Restricted Least Squares (IRLS) model, ridge and LASSO regression (with regularization). In three of the models (mixed, Bayesian, and IRLS) prior knowledge was applied. In mixed and Bayesian models priors took the form of intervals for model parameters. In IRLS, restrictions in the form of inequalities were applied. In the empirical example appraisal models were applied in the valuation of undeveloped land for residential purposes. Models with prior knowledge turned out to be the best with regard to the consistency of estimates with theory. Also, prediction accuracy was better for models with prior knowledge. In the case of low quality data expert knowledge might significantly improve estimation results of real estate appraisal econometric models.

Keywords: econometric models of real estate mass appraisal, expert knowledge, Inequality Restricted Least Squares, mixed estimation, Bayesian regression, ridge regression, LASSO

Dominik Krężolek, Szymon Stolarski, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Semiparametryczna optymalizacja portfela inwestycyjnego z wykorzystaniem funkcji kopula. Zastosowanie na rynku metali i surowców energetycznych

Streszczenie: Teoria portfelowa Harry'ego Markowitza zakłada, że stopy zwrotu mają rozkład eliptyczny, co obejmuje rozkład normalny jako szczególny przypadek. Założenie to wynika z użycia wariancji jako miary ryzyka, która jest łatwo interpretowalna w kontekście rozkładu normalnego. Jednak warto zauważyć, że w praktyce rzeczywiste stopy zwrotu często nie mają rozkładu normalnego. Mogą występować zjawiska takie jak skośność czy grube ogony, które wskazują, że rzeczywiste stopy zwrotu mogą być bardziej ekstremalne, niż przewidywałby model oparty na założeniu normalności. Ponadto, model Markowitza nie uwzględnia ryzyka ekstremalnego, tj. możliwości wystąpienia zdarzeń o niskim prawdopodobieństwie, ale znacznych konsekwencjach, które mogą istotnie wpłynąć na wyniki portfela. W pracy podjęto próbę wykorzystania podejścia semiparametrycznego do budowy portfela inwestycyjnego na rynku metali nieżelaznych i szlachetnych oraz surowców energetycznych. Oszacowano udziały portfeli optymalnych na podstawie funkcji kopula z jądrowymi rozkładami gęstości oraz charakterystyki portfeli, a następnie

porównano wyniki z podejściem klasycznym. Dokonano także oceny efektywności zbudowanych portfeli. Wyniki analizy pokazują, że optymalizacja semiparametryczna jest nie gorsza od klasycznej optymalizacji wg Markowitza, a dodatkowo uwzględnia ryzyko ekstremalne portfela.

Słowa kluczowe: optymalizacja portfela, metody semiparametryczne, estymacja jądrowa funkcje kopula

Title: Semiparametric portfolio optimization with copulas. Application on metals and energy commodity markets

Abstract: Harry Markowitz's portfolio theory assumes that returns have an elliptical distribution, which includes the normal distribution as a special case. This assumption stems from the use of variance as a measure of risk, which is easily interpretable in the context of a normal distribution. However, it is worth noting that in practice, real returns often do not have a normal distribution. There may be phenomena such as skewness or fat tails, which indicate that real returns may be more extreme than a model based on the assumption of normality would predict. In addition, the Markowitz model does not take into account extreme risk, i.e. the possibility of events with low probability but significant consequences that can significantly affect portfolio performance. In this paper, we attempted to use the semiparametric approach to construct investment portfolios in the non-ferrous and precious metals and energy commodities markets. The shares of optimal portfolios were estimated on the basis of copula functions with kernel density distributions and the characteristics of the portfolios were determined, and then the results were compared with the classical approach. The efficiency of the constructed portfolios was also evaluated. The results of the analysis show that semiparametric optimization is no worse than classical optimization according to Markowitz, and additionally takes into account the extreme risk of the portfolio.

Keywords: portfolio optimization, semiparametric methods, kernel estimation, copula functions

Mirosław Szreder, Uniwersytet Gdański

Tytuł: Precyzja i uniwersalność języka statystyki jako elementy zachwytu nad nią w pracach i działalności Profesora Czesława Domańskiego

Streszczenie: Statystyka, jako nauka wykorzystująca ściśle zdefiniowane pojęcia i kategorie, posługuje się precyzyjnym językiem, zbliżonym do języka matematyki. Jest to konieczne dla właściwego opisanie wykrywanych w zbiorach liczb prawidłowości, a także dla poprawnego zrozumienia zarówno wniosków, jak i elementów niepewności tkwiących we wnioskowaniu statystycznym. Przez 55 lat swojej pracy zawodowej prof. dr hab. Czesław Domański dał się poznać jako niestrudzony apologeta statystyki i języka, którym się ona posługuje.

W mniejszym stopniu do tej pory zwracało się uwagę na uniwersalność języka statystyki, obecnego nie tylko w nauce i codziennym życiu, ale także w utworach literackich, w tym w poezji. Język statystyki okazuje się niejednokrotnie najwłaściwszym do scharakteryzowania ludzkich przeżyć i kondycji współczesnego człowieka.

W referacie omówione zostaną oba powyższe atrybuty języka statystyki.

Słowa kluczowe: język statystyki, interpretacja i rozumienie, opis i wnioskowanie statystyczne.

Title: Precise and universal language of statistics as contributors to the admiration of statistics in works and activities of Professor Czesław Domański

Abstract: The science of statistics utilizes strictly defined notions and categories and uses the language similar to that of mathematics. This seems necessary not only for describing patterns and regularities identified in sets of numbers but also for correct understanding results of statistical inference accompanied by certain

amount of uncertainty. Professor Czesław Domański is one of those who have long praise both statistical methods and statistical language in his works and public activities.

Less attention has been paid to another attribute of statistical language, which is its universality. Statistical language is present not only in science and in common everyday life but also in literature, including poetry. It seems that statistics offers the language which in some circumstances is the most appropriate for describing human emotions and condition.

The paper discusses the above aspects of statistical language.

Keywords: language of statistics, interpretation and understanding, statistical description and inference.

Sesja Va

Tomasz Żądło, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Bootstrapowa ocena dokładności oszacowań dla prób nieprostych i nielosowych

Streszczenie: Rozważania dotyczą algorytmów bootstrapowych wykorzystywanych między innymi w dwóch podejściach w metodzie reprezentacyjnej – w podejściu randomizacyjnym i modelowym. W podejściu randomizacyjnym wnioskowanie dotyczące parametrów populacji lub podpopulacji wymaga, aby próba była próbą losową, niekoniecznie próbą prostą. W tym przypadku przyjmuje się, że wartości badanej zmiennej są nielosowe, a źródłem losowości oszacowań jest przyjęty plan losowania. Natomiast w podejściu modelowym próba nie musi być próbą losową, a źródłem losowości oszacowań charakterystyk populacji lub podpopulacji jest rozkład badanej zmiennej, co do którego przyjmuje się pewne założenia nazywane modelem nadpopulacji. W pracy zostaną przedstawione wybrane algorytmy bootstrapowe stosowane w tych podejściach, w tym autorskie propozycje, dostępne w pakietach programu R: pipsboot (<https://github.com/kucharsky/pipsboot>) oraz qape (<https://CRAN.R-project.org/package=qape>).

Słowa kluczowe: bootstrap, ocena dokładności estymacji, ocena dokładności predykcji

Title: Bootstrap accuracy estimation for complex and non-random samples

We consider the bootstrap algorithms utilized in, among other things, two approaches in survey sampling, namely the design-based and model-based approaches. In the design-based approach, inference regarding the parameters of a population or subpopulation requires that the sample be random, not necessarily a simple random sample. In this case, the values of the variable under study are assumed to be non-random, and the source of the randomness of the estimates is the sampling design adopted. In contrast, in the model-based approach, the sample does not have to be a random sample, and the source of the randomness of the estimates of the population or subpopulation characteristics is the distribution of the variable under study, about which certain assumptions are made, called the superpopulation model. Selected bootstrap algorithms used in these approaches are presented, including the author's proposals, available in R packages: pipsboot (<https://github.com/kucharsky/pipsboot>), and qape (<https://CRAN.R-project.org/package=qape>).

Keywords: bootstrap, design accuracy estimation, prediction accuracy estimation

Stanisław Jaworski, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Tytuł: Optymalny rozmiar próby w modelu trójkątnym dla pytań drażliwych

Streszczenie: Problem polega na oszacowaniu odsetka populacji o cechach stygmatyzujących. Do oszacowania zastosowano nierandomizowany model odpowiedzi zaproponowany przez Tian, G. L. i Yu, J. W. oraz Tang, M. L. i Geng, Z. (2007). Skonstruowano dokładny przedział ufności dla tego odsetka oraz wyznaczono optymalną wielkość próbki w celu uzyskania przedziału ufności o zadanej długości.

Słowa kluczowe: pytania drażliwe, model niezrandomizowanych odpowiedzi, dokładny przedział ufności

Title: Optimal sample size in triangular model for sensitive questions

Abstract: The problem is in the estimation of the fraction of population with a stigmatizing characteristic. In the paper the nonrandomized response model proposed by Tian, G. L. and Yu, J. W. and Tang, M. L. and Geng, Z. (2007) is considered. The exact confidence interval for this fraction is constructed. Also the optimal sample size for obtaining the confidence interval of a given length is derived.

Keywords: sensitive questions, nonrandomized response model, exact confidence interval

Małgorzata Krzciuk, University of Economics in Katowice

Tytuł: O estymatorach MSE predyktorów typu plug-in dla liniowych modeli mieszanych ze skorelowanymi efektami losowymi

Streszczenie: Rozważamy problem estymacji błędów średniokwadratowych predyktorów typu plug-in przy założeniu liniowego modelu mieszanego ze skorelowanymi wektorami efektów losowych. Celem badań jest porównanie, w badaniu symulacyjnym, własności wybranych estymatorów MSE w tym wykorzystujących parametryczną metodę bootstrap. W opracowaniu szerzej omówiono także rozważaną klasę predyktorów charakterystyk podpopulacji. Analiza Monte Carlo przeprowadzona została z wykorzystaniem języka R oraz danych wielookresowych o charakterze ekonomicznym.

Słowa kluczowe: statystyka małych obszarów, metoda reprezentacyjna, predyktory plug-in, Monte Carlo, analizy symulacyjne

Title: On MSE estimators of plug-in predictors for linear mixed models with correlated random effects

Abstract: We consider the problem of estimating the mean squared errors of plug-in predictors under the assumption of a linear mixed model with correlated vectors of random effects. The aim of the study is to compare, in a simulation study, the properties of selected MSE estimators, including those using the parametric bootstrap method. The paper also discusses the class of predictors of subpopulation characteristics considered in more detail. Monte Carlo analysis was conducted using the R language and longitudinal economic data.

Keywords: small area estimation, survey sampling, plug-in predictors, Monte Carlo, simulation analyses

Justyna Majewska, Uniwersytet Ekonomiczny w Katowicach

Tytuł: O odpornym na obserwacje odstające prognozowaniu umieralności

Streszczenie: Istotą wielu stochastycznych modeli umieralności jest ekstrapolacja liniowa zmiennego w czasie wskaźnika umieralności, w związku z czym na jakość estymacji i prognozowania mogą istotnie wpływać wartości odstające. Wartości odstające występują naturalnie w kontekście modelowania umieralności, w szczególności pandemii i wojen, które mogą powodować nieoczekiwane zmiany trendu umieralności.

Anomalie umieralności mają wpływ na dokładność prognoz poprzez estymację parametrów modeli, wybór ostatniego obserwowanego roku w okresie kalibracji modelu i zwiększenie wariancji. Stochastyczne modele śmiertelności wymagają zatem metod prognozowania odpornych na wartości odstające, przy czym niewiele uwagi poświęcono w dotychczasowych badaniach odpornej estymacji parametrów modeli.

W pracy rozważamy wybrane modele prognozowania umieralności, w każdym przypadku identyfikujemy wartości odstające, a następnie estymujemy efekty obserwacji odstających wraz z innymi parametrami modeli. W ten sposób usuwamy odchylenia i zniekształcenia prognozy spowodowane anomaliami, zapewniając jednocześnie odporny punkt wyjścia dla prognoz.

Słowa kluczowe: umieralność, odporna estymacja, obserwacje odstające

Sesja Vb

Katarzyna Frodyma, Uniwersytet Ekonomiczny w Krakowie

Tytuł: Metoda DEA w analizie efektywności energetycznej

Streszczenie: Badania koncentrujące się na efektywności energetycznej mają kluczowe znaczenie dla osiągnięcia zarówno redukcji emisji dwutlenku węgla, jak i zrównoważonego wzrostu gospodarczego. Jedno z podejść do badania efektywności energetycznej opiera się na koncepcji energochłonności, która jest definiowana jako zużycie energii na jednostkę produkcji (zwykle produkt krajowy brutto). Podejście to jest krytykowane za ignorowanie struktury gospodarki lub klimatu, zakładając tym samym, że produkcja gospodarcza jest jedynym źródłem produktywności. Inne podejście opiera się na DEA (Data Envelopment Analysis), która jest nieparametryczną metodą programowania liniowego przeznaczoną do pomiaru wydajności wielu jednostek decyzyjnych.

Celem tego badania jest śledzenie ewolucji efektywności środowiskowej w dużych zbiorach krajów w celu uzyskania dostępu do potencjału poszczególnych krajów w zakresie redukcji emisji. Badanie będzie poświęcone ocenie zakresu, w jakim dany kraj może zmniejszyć swoje emisje bez wywierania negatywnego wpływu na swoją gospodarkę.

W przypadku modelowania nieefektywności węglowej (lub ekoefektywności), porównania między krajami są rzadkością. Bardzo niewiele artykułów porusza tę kwestię i tylko jeden z nich obejmuje dużą próbę krajów. Większość artykułów poświęcona jest podstawowej koncepcji pomiaru emisji CO₂. Istnieją jednak dwa podejścia do pomiaru emisji CO₂: rachunkowość oparta na produkcji (PBA) i rachunkowość oparta na konsumpcji (CBA). Rachunkowość oparta na produkcji mierzy emisje CO₂ spowodowane produkcją krajową, w tym eksportem. Podejście to jest szeroko stosowane w globalnych porozumieniach w sprawie zmian klimatu, w tym w Ramowej konwencji Narodów Zjednoczonych w sprawie zmian klimatu (UNFCCC) i Protokole z Kioto. Rachunkowość oparta na konsumpcji mierzy emisje CO₂ spowodowane przez towary i usługi na poziomie ich konsumpcji, przypisując w ten sposób emisje do konsumenta końcowego. Tak więc w tym podejściu uwzględnia się emisje spowodowane handlem międzynarodowym, głównie importem z rynków wschodzących (zwłaszcza z Chin). Oba podejścia rozważane w tym badaniu znacznie się różnią, ponieważ import i eksport towarów i usług wiążących się z emisjami CO₂ wykazują potencjał do międzynarodowej ucieczki emisji.

Słowa kluczowe: efektywność energetyczna; DEA

Title: DEA method in energy efficiency analysis

Abstract: Research focusing on energy efficiency, is crucial for achieving both the reductions in carbon emissions and sustainable economic growth. One approach to studying energy efficiency is based on the concept of energy intensity, which is defined as energy consumption per output unit (usually gross domestic product). This approach is criticized for ignoring the economy structure or climate, thus assuming that the economic output is the only source of productivity. Another approach is based on the DEA (Data Envelopment Analysis), which is a non-parametric linear programming method dedicated to measuring the efficiency of multiple decision-making units.

The aim of this study is to follow the evolution of environmental efficiency in large sets of countries with a view to accessing the individual countries' potential to reduce the amount of emissions. The research will be dedicated to assessing the extent to which a particular country might reduce its emissions without exerting a negative effect on its economy.

When carbon inefficiency (or eco-inefficiency) is modelled, cross-country comparisons are scarce. Very few papers tackle this issue and only one of them covers a large sample of countries.

Most articles are devoted to the fundamental concept of measuring CO₂ emissions. There are two approaches, however, to measuring CO₂ emissions: production-based accounting (PBA) and consumption-based accounting (CBA). Production-based accounting measures CO₂ emissions caused by domestic production, including exports. This approach is widely used in global climate change agreements, including the United Nations Framework Convention on Climate Change (UNFCCC) and the Kyoto Protocol. Consumption-based accounting measures CO₂ emissions caused by goods and services at the level of their consumption, thus allocating the emissions to the final consumer. So, in this approach emissions caused by international trade, primarily by import from emerging markets (especially China) are considered. The two approaches considered in this research differ significantly because imports and exports of goods and services involving CO₂ emissions demonstrate the potential for international carbon leakage.

Keywords: energy efficiency; DEA

Monika Zielińska-Sitkiewicz, Mariola Chrzanowska, SGGW w Warszawie

Tytuł: Analiza konsumpcji energii elektrycznej w Polsce z wykorzystaniem modeli predykcji i sieci neuronowych

Streszczenie: Zmiany w uwarunkowaniach ekonomicznych, regulacyjnych i ekologicznych oraz postęp technologiczny doprowadziły do ukształtowania się nowoczesnego modelu funkcjonowania sektora energii elektrycznej. Jednakże pilna konieczność modernizacji polskich i europejskich sieci energetycznych stale obciąża innowacyjność i elastyczność rozwiązań w tej kluczowej dla konkurencyjnej gospodarki branży. Dla efektywnego funkcjonowania rynku energii niezbędny jest odpowiedni zasób informacji o energii elektrycznej jako towarze. Celem artykułu jest analiza dziennej konsumpcji energii elektrycznej w Polsce przy użyciu wybranych metod predykcji - metody naiwnej, modeli SARIMA, sieci neuronowych oraz porównanie jakości otrzymanych prognoz. W badaniu szeregu czasowego wykorzystano dane dziennego zużycia energii elektrycznej dla lat 2019 – 2022 pozyskane z portalu www.pse.pl. Metodą prognostyczną najczęściej wykazującą się najniższym błędem MAPE, były modele SARIMA. Uzyskane wyniki potwierdzają możliwość stosowania wybranych modeli predykcji oraz sieci neuronowych do prognozowania zapotrzebowania na energię elektryczną w krótkim horyzoncie czasowym.

Słowa kluczowe: energia elektryczna, modele SARIMA, sieci neuronowe

Title: Analysis of electricity consumption in Poland using prediction models and neural networks

Abstract: Changes in economic, regulatory and environmental conditions, as well as technological progress, have led to the emergence of a modern operating model for the electricity sector. However, the urgent need to modernise the Polish and European energy networks requires constant innovation and flexibility of solutions in this key sector for a competitive economy. An adequate supply of information on electricity as a commodity is essential for the effective functioning of the energy market. The purpose of this paper is to analyse the daily electricity consumption in Poland using selected forecasting methods - naive method, SARIMA models, neural networks - and to compare the quality of the obtained forecasts. The time series study used daily electricity consumption data for the years 2019 - 2022 obtained from www.pse.pl. The forecasting method with the lowest MAPE error was SARIMA models. The results obtained confirm the applicability of selected forecasting models and neural networks for short-term electricity demand forecasting.

Keywords: electricity, SARIMA models, neural networks

Klaudia Lenart, Szkoła Doktorska Uniwersytetu Ekonomicznego w Katowicach

Tytuł: Zastosowanie Google Trends jako źródła danych w modelach statystycznych

Streszczenie: Wraz z postępowaniem technologicznym rośnie liczba potencjalnych źródeł danych, które mogą stanowić alternatywę dla tradycyjnych badań ankietowych. Przykładem tego mogą być dane o popularności wyszukiwań udostępniane w czasie rzeczywistym za pośrednictwem Google Trends. Dane tego typu pozwalają na badanie zachowań i postaw w społeczeństwie, opinii publicznej czy prognozowanie zjawisk ekonomicznych.

Zdecydowaną zaletą wykorzystania danych o popularności wyszukiwań jest natychmiastowy czas i niski koszt ich pozyskania. Nie bez znaczenia jest też fakt, że Google Trends pozwala na bezpośrednie badanie zachowań użytkowników Internetu, a nie jedynie ich deklaracji jak w przypadku ankiety. Może to mieć znaczenie jeżeli ankietowani uważają którąś z odpowiedzi za bardziej moralnie słuszną. Korzystanie z Google Trends wymaga jednak trafnego dobrania uwzględnianych w badaniu wyszukiwań oraz świadomości ograniczenia próby badawczej do użytkowników wyszukiwarki Google.

W referacie zaprezentowane zostaną wady i zalety Google Trends oraz przykładowe sposoby wykorzystania udostępnianych za pośrednictwem niego danych.

Słowa kluczowe: Google Trends, analizy statystyczne, prognozowanie

Title: Application of Google Trends as a data source for statistical models

Abstract: As technology advances, there is a growing number of potential data sources that can provide an alternative to traditional surveys. An example of this is the real-time search popularity data made available through Google Trends. This type of data makes it possible to study behaviour and attitudes in society, public opinion or forecast economic phenomena.

A definite advantage of using search popularity data is the immediate time and low cost of obtaining it. Also of significance is the fact that Google Trends allows for direct research into the behaviour of internet users, and not just their declarations as it is in the case of a survey. This can make a difference if respondents consider one of the answers to be more morally correct. Nevertheless the use of Google Trends requires selecting correct search queries to be included in the study and an awareness of the fact that the research sample is limited to Google search engine users.

This paper will present the advantages and disadvantages of Google Trends and examples of how the data it provides can be used.

Keywords: Google Trends, statistical analysis prognosis

Magdalena Kawecka, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Kwalifikacje i umiejętności a kompetencje przyszłości przez dorosłych Polaków;

Streszczenie: Kompetencje odgrywają kluczową rolę w gospodarce, wpływając na jakość kapitału ludzkiego, który jest jednym z głównych motorów rozwoju społeczno-gospodarczego. Ich znaczenie można analizować zarówno w kontekście funkcjonowania organizacji, jak i w aspekcie indywidualnym, związanym z rozwojem i ścieżką kariery jednostek. Ze względu na istotność kompetencji przyszłości, istotne jest badanie kwalifikacji i umiejętności dorosłych Polaków. Uwzględniając złożoność relacji między posiadanymi umiejętnościami a przyszłymi kompetencjami, zastosowano metody statystyczne do oceny zasobów kompetencyjnych. Analizy oparto na szerokim spektrum danych z badania Bilansu Kapitału Ludzkiego.

Sesja VIa

Eugeniusz Gatnar, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Jak włączyć koszty użytkowania mieszkań przez właścicieli do koszyka inflacyjnego?

Streszczenie: Europejski Bank Centralny, który w 2021 roku opublikował nową strategię polityki pieniężnej, zdecydował o włączeniu do wskaźnika inflacji HICP kosztów gospodarstw domowych związanych z posiadaniem i utrzymaniem własnego mieszkania lub domu (owner-occupied housing, OOH). Obecnie te koszty nie są ujmowane, chociaż prawie 70 proc. gospodarstw domowych w Unii Europejskiej mieszka we własnych domach. To oznacza, że obraz struktury konsumpcji nie jest pełny.

Również w Polsce, tego typu wydatki nie są ujmowane we wskaźniku inflacji CPI, chociaż wskaźnik zajmowania mieszkań przez właścicieli jest jeszcze wyższy i wynosił w roku 2022 przeszło 87 proc.

W referacie zostaną przedstawione metody pomiaru kosztów ponoszonych przez właścicieli w związku z nabyciem, utrzymaniem i zamieszkaniem we własnym domu. Zostaną omówione także przykłady skutecznej implementacji tego rozwiązania przez niektóre kraje, takie jak Czechy, czy Wielka Brytania.

Słowa kluczowe: Inflacja, koszyk inflacyjny, CPI, HICP, OOH

Title: How to add owner occupiers' housing costs to the inflation basket?

Abstract: The European Central Bank published in 2021 its new monetary policy strategy, decided to add to the HICP inflation index, the household expenses related to owning and maintaining own apartment or home (owner-occupied housing, OOH). Currently, it is not included, although 70 percent of households in the EU live in their own homes. This means that the picture of the consumption structure is not complete.

Also in Poland, this type of expenditure is not included in the CPI inflation index, although the owner occupancy rate is even higher and amounted to over 87% in the year 2022.

The paper will discuss methods of measuring the costs of homeowners associated with owning, maintaining and living in their own home. Examples of successful implementation in some countries, such as the Czech Republic, and Great Britain will also be presented.

Keywords: Inflation, inflation basket, CPI, HICP, OOH

Jacek Białek, Natalia Pawelec, Uniwersytet Łódzki i Główny Urząd Statystyczny,

Tytuł: Zastosowanie tranzytywnych indykatorów Montgomery'ego do analizy danych skanowanych

Sterszczenie: Chociaż współczesna teoria indeksów cen opiera się na analizie ilorazów cen i ilości, często możemy być bardziej zainteresowani pracą z różnicą tych wartości. Ma to miejsce na przykład przy dekompozycji zmian przychodów, dekompozycji zmian zysków i kosztów lub w analizie zmian wydatków konsumenta. Zaletą korzystania z indykatorów jest fakt, iż nie ma wówczas problemu związanego z występowaniem zerowych cen i ilości, które pojawiają się, gdy pracujemy z ilorazami. W praktyce najczęściej zależy nam na dekompozycji różnicy wartości pewnego koszyka na indykatory wkładu różnic cenowych i ilościowych (efekt cenowy i ilościowy). Dobrze znanymi indykatorami cen i ilości są indykatory Benneta i Montgomery'ego, które jednak nie posiadają własności tzw. tranzytywności (przechodniości).

Niniejsza praca rewiduje cenowe i ilościowe indykatory Montgomery'ego oraz ich wielostronne wersje w ramach analizy danych skanowanych. W szczególności, zamiast rozważać "klasyczne" porównania między firmami, krajami lub regionami, tranzytywne wersje indykatorów Montgomery'ego są przystosowane do pracy na zestawach danych skanowanych obserwowanych w ustalonym oknie czasowym. Jednym z celów badania jest porównanie dwustronnych (bilateralnych) i wielostronnych (multilateralnych) indykatorów Montgomery'ego dla różnych poziomów agregacji danych i trzech typowych filtrów danych. Zgodnie

z najlepszą wiedzą autorów, niniejsze badanie jest pionierskie na gruncie implementacji multilateralnych wskaźników Montgomery'ego w analizie danych skanerowych.

Słowa kluczowe: dane skanowane, indykatory Montgomery'ego, tranzytywność, indykatory multilateralne

Title: The use of the transitive Montgomery indicators for scanner data analysis

Abstract: Although the modern price index theory is based on an analysis of ratios of prices and quantities, we often may be more interested in working with differences in these values in many economic areas, e.g.: revenue change decompositions, profit and cost change decompositions, or an analysis of changes in consumer surplus. The benefit of using these differences is that there is no problem associated with the occurrence of zero prices and quantities, a problem that arises when we work with ratios. In practice, we most often care about decomposing value difference into indicators of contributions from price and quantity differences. The well-known price and quantity indicators are the Bennet or Montgomery indicators, which are not transitive.

This paper revises the price and quantity Montgomery indicators and their multilateral versions for the analysis of scanner data. Specifically, instead of considering "classical" comparisons across firms, countries or regions, the transitive versions of the Montgomery indicators are adapted to work on scanner data sets observed over a fixed time window. One of the objectives of the study is to compare bilateral and multilateral Montgomery indicator values for different data aggregation levels and three main types of data filters. According to the best of the authors' knowledge, this study is pioneering on the grounds of implementing the multilateral Montgomery indicators in scanner data analysis.

Keywords: scanner data, Montgomery indicators, transitivity, multilateral indicators

Adam Juszcak, Uniwersytet Łódzki

Tytuł: Zastosowanie danych skanowanych w pomiarze dynamiki cen

Streszczenie: Dane skanowane to szczegółowe dane o dobrach konsumpcyjnych uzyskane dzięki skanowaniu ich kodów kreskowych w punktach sprzedaży. Mimo, że technologia ta jest znana od lat 70 dopiero w ciągu ostatnich 20 lat rozwijają się metody użycia danych z kodów kreskowych do liczenia inflacji. W pracy autor omawia wyniki analiz dla indeksów bilateralnych i multilateralnych używanych do szacowania cen produktów skanowanych. Analizie poddano także wpływ metod rozszerzania okna obliczeń, metod filtracji danych i metod agregacji danych.

Słowa kluczowe: CPI, Inflacja, dane skanowane, GEKS-J, SPQ, Geary-Khamis

Title: Usage of scanner data in price dynamic measurement

Abstract: Scanned data is detailed data about consumer goods obtained by scanning their barcodes at points of sale. Although this technology has been known since the 1970s, methods of using barcode data to calculate inflation have only been developed in the last 20 years. In the work, the author discusses the results of analyzes for bilateral and multilateral indices used to estimate prices of scanned products. The impact of methods for expanding the calculation window, data filtration methods and data aggregation methods was also analyzed.

Keywords: CPI, Inflation, scanner data, GEKS-J, SPQ, Geary-Khamis

Tytuł: Porównanie indyktorów Benneta i Montgomery'ego z wykorzystaniem danych skanowanych

Streszczenie: Indykatory cen i ilości, zaproponowane w I połowie XX wieku, przeżywają ostatnio swoiste odrodzenie w literaturze ekonomicznej poświęconej analizie dynamiki cen i wartości sprzedaży. Wynika to z faktu, iż w przeciwieństwie do indeksów, indykatory te są mniej wrażliwe na występowanie zerowych cen i ilości. Ma to szczególne znaczenie przy analizie segmentów rynku charakteryzujących się dużą rotacją produktów oferowanych w sprzedaży. W szczególności, indykatory cen i ilości mogą znaleźć zastosowanie w analizie danych skanowanych, czyli danych pochodzących bezpośrednio z sieci handlowych. Analityk stoi wówczas przed dylematem wyboru odpowiedniej formuły indykatora, przy czym dwie najpopularniejsze z nich to formuła Benneta i Montgomery'ego. W pracy dokonano zatem porównania zarówno własności aksjomatycznych tych indyktorów, jak i potencjalnych różnic wartości, jakich możemy oczekiwać analizując dane skanowane.

Słowa kluczowe: dane skanowane, indykator Benneta, indykator Montgomery'ego

Title: Comparison of Bennet and Montgomery indicators using scanner data

Abstract: Price and quantity indicators, offered in the first half of the 20th century, have recently experienced a kind of revival in the economic literature devoted to the analysis of price and sales value dynamics. This is due to the fact that, contrary to indices, these indicators are less sensitive to the occurrence of zero prices and quantities. This is particularly important when analyzing market segments characterized by high turnover of products offered for sale. In particular, price and quantity indicators can find application in the analysis of scanned data, i.e. data coming directly from retail chains. The analyst is then faced with the dilemma of choosing an appropriate indicator formula, the two most popular being the Bennet and Montgomery formulas. The paper therefore compares both the axiomatic properties of these indicators and the potential value differences which we can expect when analyzing scanner data.

Keywords: scanner data, Bennet indicator, Montgomery indicator

Sesja VIb

Stanisław Maciej Kot, Gdansk University of Technology

Title: Aversion to inequality and to rank inequality in European Union countries

Abstract: Welfare economics seeks an answer to whether a given policy provides a higher (lower) economic welfare of society than an alternative policy. A social welfare function (SWF) is the fundamental instrument for such comparisons. However, various competing SWFs exist, so answering the question could be problematic. This paper uses two families of SWFs, namely $SWF_{\epsilon} = \mu(1 - A_{\epsilon})$ and $SWF_{\nu} = \mu(1 - G(\nu))$, where μ is the mean income, A_{ϵ} is Atkinson's inequality index, and $G(\nu)$ is the generalised Gini index. Parameter $\epsilon \geq 0$ reflects aversion to income inequality, parameter $\nu \geq 1$ reflects aversion to rank inequality. These parameters characterise the individual members of the families in question. This paper asks the following question: What should ϵ and ν be to guarantee the equality $SWF_{\epsilon} = SWF_{\nu}$ when assessing social welfare in a given income distribution? The paper offers the solution to this problem and estimates of ϵ and ν for The European Union countries in 2004-2021.

Keywords: income, social welfare, inequality aversion, European Union

Title: Multiple faces of poverty. Exploring housing-costs-induced energy poverty in Central and Eastern Europe

Abstract: The series of crises, first pandemic, then energy, and finally war threaten the most vulnerable in all countries. This study attempts to unveil the faces of the energy poor to shape conscious policies concerning these households. We focus on the profile of energy-poor households in 11 CEE countries, such as Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia.

We identify nine distinct profiles of energy poor households and key vulnerability characteristics. The first one is low income, which, especially for retirees, does not allow them to meet their energy needs. The second one is living in houses. The third one is the mismatch between housing size and household needs. The fourth is a large number of children in the household. Our source of micro-data is the EU Survey on Income and Living Conditions collected in 2020. We perform the analysis using the partitioning around medoids algorithm on the sample of 18331 energy poor households, which represent about 22.6% of all observations.

We highlight several policy issues in this study. It is important to make the housing market more flexible to adjust accommodation to the size of households. Pension systems should account for population aging and improve the financial situation of single elderly persons. Finally, large families with dependent children should be supported more effectively.

Keywords: energy poverty profile, housing costs, cluster analysis, Central and Eastern Europe

Svitlana Chugaievskaya, Andrzej Frycz Modrzewski Krakow University, Poland

Tytuł: Integracja przymusowych migrantów ukraińskich w nowych przestrzeniach europejskich: wyniki przeprowadzonego ankietowania

Streszczenie: Nowa fala migracji ukraińskiej rozpoczęła się na tle inwazji wojskowej RF na Ukrainę na pełną skalę w dniu 24 lutego 2022 r., kiedy około 8 mln Ukraińców opuściło kraj w poszukiwaniu bezpieczeństwa dla siebie i swoich dzieci, a około 6 mln milion ludzi zostało zmuszonych do przeniesienia się do innych regionów kraju. W rzeczywistości około 20% stałej populacji państwa, czyli mniej więcej co piąty Ukrainiec, został imigrantem przymusowym od początku pełnoskalowej wojny.

Badanie wyzwań gospodarczych stojących przed krajami UE na tle największej fali migracyjnej w Europie od 80 lat zawiera analizę statystyczną wyników badania ankietowego przeprowadzonego wśród 218 migrantów z Ukrainy. Zbadano specyfikę wsparcia społecznego, psychologicznego i finansowego migrantów w nowych miejscach zamieszkania. Szczególną uwagę poświęcono badaniu szeregu czynników stymulujących osoby migrujące do opuszczenia swoich domów i możliwości ich powrotu. Jednocześnie podkreślano rosnące obciążenie instytucji edukacyjnych, medycznych i społecznych krajów goszczących.

Słowa kluczowe: Fala migracyjna, przymusowi migranci z Ukrainy, integracja, wyniki ankietowania

Title: Integration of forced Ukrainian migrants into new European spaces: results of the conducted survey

Abstract: The new wave of Ukrainian migration began against the background of the full-scale military invasion of the Russian Federation into Ukraine on February 24, 2022, when about 8 million Ukrainians left the country in search of safety for themselves and their children, and about 6 million people were forced to move to other regions of the country. In fact, around 20% of the permanent population of the state, or approximately every fifth Ukrainian, has become a forced migrant since the beginning of the full-scale war. The study of economic challenges for EU countries against the background of the largest migration wave in Europe in the last 80 years includes a statistical analysis of the results of a survey of 218 Ukrainian migrants. Peculiarities of social, psychological and financial support of migrants in new places of residence were studied. Special attention was paid to the study of a number of factors that stimulated migrating people to

leave their homes and the possibility of returning home. Along with this, the increasing burden on educational, medical institutions and social services of the host countries was emphasized.

Keywords: Migration wave, forced Ukrainian migrants, integration, survey results

Krzysztof Dmytrów, University of Szczecin

Title: Comparison of results for selected multi-criteria methods of supporting location selection during order picking

Abstract: In warehouse management, one of the most fundamental processes is the order picking. The costs associated with it represent up to 65% of the total warehouse operating costs. Moreover, about 55% of the total order picking time is consumed by the movement of the picker. Therefore, the greatest effects associated with optimising the order picking process can be achieved by reducing the picker's distance as much as possible. This can be achieved in several ways:

- by using an appropriate storage order,
- by using an appropriate warehouse layout,
- by an appropriate choice of locations in which the items are stored,
- by an appropriate determination of the picker's route.

The aim of the study is selection of the best multi-criteria decision-making method for location selection in a rectangular, one-block warehouse with two main aisles. The problem of location selection is applicable to shared storage (where each item can be stored in multiple, often very distant locations, and many different items can be stored in every location). Locations are described using several criteria, and they are selected using selected multi-criteria decision-making methods: the patternless COPRAS method, the TOPSIS method (which assumes weighted distances from pattern and anti-pattern) and the TMAL method (which is a variation of the Z. Hellwig's synthetic measure of development in two variations – one assuming weighted distances from the pattern and the other from the anti-pattern). Six strategies of locations selection were applied, by means of which criteria were weighted. Goods were stored in random order and by ABC class storage with within aisle, across aisle, diagonal and perimeter layouts. The picker's route was determined using the s-shape, return and midpoint heuristics. Each item was stored in a variable number of locations, ranging from 1 to 10. It was considered that the order consisted of 5, 10 and 20 items. For every multi-criteria decision-making method, storage order, heuristics, number of items in the order and strategy of location selection, 1,000 simulation runs were performed. The shortest route (and order picking time) can be obtained by using a quantity adjustment strategy and a strategy that groups locations along the fewest number of picking aisles. In most cases, the best results were obtained using the COPRAS method.

Keywords: warehouse management, order picking, multi-criteria decision-making, simulation

Sesja VIIIA

Andrzej Sokołowski, Małgorzata Markowska, Katarzyna Budny, Collegium Humanum – Szkoła Główna Menedżerska, Uniwersytet Ekonomiczny we Wrocławiu, Uniwersytet Ekonomiczny w Krakowie

Tytuł: MODALNA MEDIAN

Streszczenie: Celem pracy jest zaproponowanie metody szacowania mediany wielowymiarowej zmiennej losowej. Załóżmy, że mamy n -elementową próbę. Dla każdej pary punktów wyznaczamy prostą przechodzącą przez te punkty, a następnie na tę prostą rzutujemy pozostałe punkty empiryczne. Na tej prostej znajdujemy medianę rzutów. Takich par punktów, a w konsekwencji prostych i median jest $(n/2)$. Jeżeli jest ich bardzo dużo to można ograniczyć się do zadanej liczby par punktów (prostych, median) losowanych ze zwracaniem z próby. Pojedyncza mediana - jakkolwiek wyznaczana jest na prostej –

definiowana jest w przestrzeni wielowymiarowej. W zbiorze takich median wyznaczamy modalną. Jedną z metod możliwych tu do zastosowania jest stopniowe eliminowanie obiektów najbardziej odległych od średniej arytmetycznej pozostałych punktów, aż do momentu gdy w analizie pozostaną dwa obiekty. Średnie arytmetyczne ich współrzędnych wyznaczają ocenę modalnej wielowymiarowej, która jednocześnie jest proponowanym przez nas estymatorem mediany wielowymiarowej.

Ilustracją proponowanej procedury będą przykłady symulowane oraz wykorzystujące dane rzeczywiste.

Słowa kluczowe: mediana wielowymiarowa, modalna wielowymiarowa

Title: THE MODE OF MEDIANS

Abstract: The aim of the paper is to propose a new method for estimating the median of multidimensional random variable. In n -element sample, a straight line is defined for each pair of points. Then the rest of the sample is projected on this line, and the median is being found. There are $(n/2)$ such pairs of points, straight lines and individual medians. If this number is very large we can limit it to a given number of pairs of points selected randomly with replacement from the sample. A single median, despite from being found on the line, is finally defined in the original multidimensional space. The next step is to find the mode in the set of individual medians. So called data peeling is one of the possible methods used for this purpose. Distances between each object and the average (centroid) of the rest of objects are calculated. The most distance object is deleted from the lot, step by step. When only two objects are left, the average of them serves as the estimate of the multidimensional mode of medians, being also the estimate of the overall multidimensional median.

The procedure is illustrated by simulated and real data.

Keywords: multidimensional median, multidimensional mode

Piotr Sulewski, Uniwersytet Pomorski w Słupsku

Streszczenie : Celem tego artykułu jest przetestowanie zdolności testów zgodności do wykrywania odchyień od normalności. Rozważany jest bardzo szczególny przypadek, a mianowicie odchylenie od normalności uzyskuje się za pomocą rozkładów asymetrycznych z małymi wartościami skośności γ_1 . Pierwszym krokiem w realizacji powyższego celu jest zebranie testów zgodności zorientowanych na normalność, które są zalecane do stosowania, głównie w niedawno opublikowanej literaturze źródłowej. Drugim krokiem jest utworzenie rodziny rozkładów asymetrycznych ze zmienną skośnością, zwanych dalej alternatywami. Dla każdego rozkładu podano wzory do obliczania wartości γ_1 . Aby porównać alternatywy z rozkładem normalnym, stosuje się odpowiednią miarę podobieństwa. Trzecim krokiem są symulacje Monte Carlo. Badanie przeprowadzono z udziałem dwudziestu jeden testów zgodności i trzynastu alternatyw. Uzyskane wyniki wskazują, że testy zgodności $LF_{\alpha, \beta}$ i H_n znacznie lepiej wykrywają rozkłady asymetryczne z małymi wartościami skośności, nawet 0.05.

Słowa kluczowe: rozkład normalny test zgodności, modelowanie eksces kurtozy

Abstract: The aim of this article is to seamlessly test GoFT's ability to detect deviations from normality. The very specific case is considered, namely deviation from normality consists in coincidence of asymmetry and small skewness γ_1 . The first step in fulfilling the above aim is to collect the set of normality-oriented GoFTs being recommended for use, mainly in recently published source literature. The second step is to create a family of asymmetric distributions with non-constant γ_1 , further called alternatives. Formulas for calculating γ_1 are provided for each alternative. To compare alternatives with the Normal distribution, the appropriate similarity measure is applied. The third step is the Monte Carlo simulation. The study is carried out involving twenty one GoFTs and thirteen alternatives. The obtained results show that the $LF_{\alpha, \beta}$ and H_n GoFTs best detect asymmetric distributions that deviate from normality due to small skewness, even 0.05.

Keywords: normal distribution, goodness-of-fit testing, excess kurtosis modeling

Grzegorz Sitek, Uniwersytet Ekonomiczny w Katowicach

Tytuł: Rozkład sumy rozkładu normalnego i rozkładu gamma.

Streszczenie: Audyt finansowy to kompleksowe badanie sprawozdania finansowego przez biegłego rewidenta. Audyt jest niezależną weryfikacją rzetelności tych sprawozdań finansowych. Różnica między wartością księgową i wartością prawidłową nazywana jest błędem audytowym. Wartości księgowe traktowane są jako wartości zmiennej losowej, której rozkład jest mieszanką rozkładów wartości prawidłowej i wartości obciążonej błędem. Współczynnik mieszający jest równy proporcji pozycji o niezerowych wartościach błędów. Suma rozkładu normalnego i gamma może być przydatna do modelowania błędów audytowych. W pracy wyprowadzamy wzór funkcji gęstości sumy zmiennej losowej o rozkładzie normalnym i zmiennej losowej o rozkładzie gamma i przedstawiono problem estymacji parametrów otrzymanej funkcji gęstości. Badanie to może być przydatne w rewizji finansowej.

Słowa kluczowe: mieszanka rozkładów prawdopodobieństwa, suma rozkładu normalnego i rozkładu gamma, błąd audytowy.

Title: Sum of gamma and normal distribution.

Abstract: In financial accounting, data about the economic activities of a given firm is collected and then summarized and reported in the form of financial statements. Auditing, on the other hand, is the independent verification of the fairness of these financial statements. An item in an audit sample produces two pieces of information: the book (recorded) amount and the audited (correct) amount. The difference between the two is called the error amount. The book amounts are treated as values of a random variable whose distribution is a mixture of the distributions of the correct amount and the true amount contaminated by error. The mixing coefficient is equal to the proportion of the items with non-zero errors amounts. The sum of normal and gamma distribution can be useful for modeling audit errors. In this we derive a formulation of the probability distribution of the sum of a normally distributed random variable and one with gamma distribution and then the problem of estimating the parameters of the obtained final function was considered. This research could be useful in financial auditing.

Keywords: mixture of probability of distribution, sum of gamma and normal distribution, accounting error

Kamila Hasilova, Ivana Horova, David Valis, Stanislav Zamecnik, University of Defence, Masaryk University, University of Defence, Masaryk University

Title: Kompleksowa eksploracja zupełnej metody krzyżowego uwiarygodnienia dla danych okrągłych

Streszczenie: Estymatory jądrowe gęstości danych okrągłych zyskały ostatnio znaczną uwagę ze względu na ich zdolność do modelowania i analizowania rozkładów prawdopodobieństwa na okręgach jednostkowych i innych domenach okresowych. Naszym celem jest wnieść wkład w literaturę na temat selektorów wartości parametru wygładzania w estymacji jądrowej gęstości danych okrągłych. Proponujemy nowatorską metodę specyficzną dla danych okrągłych, która opiera się na metodę krzyżowego uwiarygodnienia. Korzystając z symulowanych danych, a także rzeczywistych zbiorów danych okrągłych, oceniamy i weryfikujemy proponowaną metodę oraz porównujemy ją z istniejącymi metodami.

Słowa kluczowe: dane okrągłe, estymator jądrowy gęstości, gęstość von Misesa, metoda krzyżowego uwiarygodnienia

Title: A comprehensive exploration of complete cross-validation for circular data

Abstract: Kernel density estimation of circular data has recently received considerable attention for its ability to model and analyse distributions on unit circles and other periodic domains. Our aim is to contribute to the literature on data-driven bandwidth selectors in circular kernel density estimation. We propose a novel circular-specific method that is based on a cross-validation procedure. Using simulated data as well as real-world circular datasets, we evaluate and validate the proposed method and compare it with existing methods.

Keywords: circular data, kernel density estimation, von Mises density, cross-validation method

Sesja VIIIb

Paweł Lula, Józef Pocięcha, Krakow University of Economics

Tytuł: Problemy modelowania regresyjnego narzędziami wnioskowania statystycznego oraz narzędziami uczenia maszynowego

Streszczenie: Podstawową funkcją nauki jest jej funkcja poznawcza, a najbardziej ogólnym narzędziem poznawania rzeczywistości społeczno-ekonomicznej jest jej aproksymacja. W sytuacji, gdy rozpatrywane zmienne mają charakter losowy, aproksymacja ta jest aproksymacją stochastyczną. Podstawowym narzędziem badań ekonomicznych jest model ekonometryczny, mający najczęściej postać modelu regresyjnego. W pracy przypomniano ogólne ujęcie modelu regresyjnego jako narzędzia aproksymacji stochastycznej, zaprezentowane w przeszłości przez Prof. Zdzisława Hellwiga. Alternatywnym, w stosunku do klasycznego ujęcia aproksymacji stochastycznej jest modelowanie regresyjne oparte na koncepcji statystycznego uczenia maszynowego. W szczególności odniesiono się do koncepcji prawdopodobnie w przybliżeniu poprawnego uczenia się modelu oraz do koncepcji wymiaru Vapnika-Chervonenkisa. W swoim wystąpieniu autorzy odnoszą się do wymienionych powyżej kwestii i omawiają ich aspekty teoretyczne i praktyczne w badaniach społeczno-ekonomicznych.

Keywords: aproksymacja stochastyczna, model regresyjny, uczenie maszynowe

Grzegorz Kończak, Martyna Kosińska, Uniwersytet Ekonomiczny w Katowicach

Tytuł: O wnioskowaniu statystycznym dla danych w wielowymiarowych tablicach wielodzielczych

W badaniach ekonomicznych i społecznych często identyfikuje się wiele kategorii zmiennych jakościowych. W takich przypadkach do raportowania wyników badania są często wykorzystywane tabele wielowymiarowe. Analizy opisowe wykorzystują współczynniki kontyngencji C Pearsona, V Cramera, T Czuprowa do opisu siły związku. Wnioskowanie statystyczne na podstawie danych zawartych w tablicach kontyngencji przeprowadza się najczęściej z wykorzystaniem testu niezależności chi-kwadrat lub testu jednorodności chi-kwadrat. Postać statystyki w obu tych testach jest taka sama, ale mechanizm powstawania danych jest inny. W pierwszym przypadku próba o rozmiarze n jest losowana z pojedynczej populacji z dwiema zmiennymi klasyfikacyjnymi. W drugim przypadku próbki z k populacji o rozmiarach n_1, n_2, \dots, n_k są losowane niezależnie, biorąc pod uwagę tylko jedną zmienną klasyfikującą. Podczas próbkowania k populacji przy użyciu odpowiednich miar można opisać podobieństwo struktur w tych populacjach. Struktury różnych populacji w tym samym okresie i dla tej samej populacji w różnych okresach mogą być porównywane. Znacznie rzadziej dokonuje się porównań struktur dla więcej niż dwóch zmiennych jakościowych. W takich

przypadkach wyniki badania prezentowane są w wielowymiarowych tabelach. Celem niniejszego artykułu jest zaproponowanie testu statystycznego potwierdzającego występowanie istotnych różnic w strukturach na podstawie danych zawartych w tablicach wielowymiarowych. Zastosowanie proponowanego testu przedstawiono na przykładzie wyników badań własnych dotyczących uczestnictwa w kulturze aktywnych uczestników portali internetowych bezpośrednio przed i w trakcie pandemii COVID-19.

Słowa kluczowe: wnioskowanie statystyczne, tablice wielodzielcze, testy permutacyjne

Title: On the statistical inference for data in multidimensional contingency tables

Abstract: In economic and social scientific research, multiple categories of qualitative variables are often identified. In these cases, multivariate tables are most often used to report the results of the study. Descriptive analyses use Pearson's C, Cramer's V, Czuprow's T contingency coefficients to describe the strength of the relationship. Statistical inference on the basis of the data contained in contingency tables is most often carried out using the chi-square test of independence or the chi-square test of homogeneity. The form of the statistic in both of these tests is the same, but the method of obtaining the data is different. In the first case, a sample of size n is drawn from a single population with two classification variables. In the second case, samples from k populations of sizes n_1, n_2, \dots, n_k are drawn independently, considering a single classifier variable only. The structures of different populations in the same period and for the same population in different periods can be compared.

It is much rarer to conduct an analysis simultaneously for more than two qualitative variables. In such cases, the results of the study are presented in multidimensional multivariate tables. The purpose of this article is to propose a permutation statistical test to confirm the presence of significant differences in structures based on data contained in multidimensional contingency tables. The application of the proposed test is presented using the example of the results of our own research on the cultural participation of active participants in Internet portals immediately before and during the COVID-19 pandemic.

Keywords: inferential statistics, contingency tables, permutation tests

Piotr Bolesław Nowak, University of Wrocław

Tytuł: Estymacja w modelu Coxa z pogrupowanymi czasami życia

W praktyce nie da się zmierzyć czasu przeżycia z nieskończoną precyzją. Na przykład w konstrukcji tabel przeżycia czas jest zaokrąglony do najbliższego roku. Co więcej, wartości czasu często dzieli się na rozłączne podzbiory, np. przedziały, prostokąty. Oznacza to, że nie jest możliwe podanie dokładnych wartości czasu przeżycia, a jedynie liczby zdarzeń w każdej określonej klasie. W niniejszej pracy rozważamy problem estymacji parametrów w modelu Coxa, kiedy wartości czasów przeżycia są zgrupowane, tzn. dla każdej wartości regresorów nie obserwujemy dokładnej wartości czasu przeżycia a jedynie wiemy w którym przedziale czasu wystąpiło zdarzenie. Wyjaśnimy, jak można wykorzystać liczby losowe do przekształcenia zgrupowanych czasów życia w próbę pseudo-kompletną, co w konsekwencji pozwala na użycie klasycznego estymatora w modelu Coxa. Udowodnimy, że dla założenia o rozkładzie eliptycznie sferycznym dla zmiennych objaśniających, estymatory regresji parametrów w modelu Coxa opartym na próbie pseudo-kompletnej są zgodne z dokładnością do parametru skali. Badanie symulacyjne ilustrują asymptotyczne właściwości estymatorów. Przedstawimy także rzeczywistą analizę przypadku.

Słowa kluczowe: Model Coxa, dane zgrupowane, Fisherowska zgodność, rozkład eliptyczny

Title: Estimation in the Cox model with grouped lifetimes

Abstract: In practice it is impossible to measure time with infinite precision. For instance, in constructing life tables age is rounded to the nearest year. Moreover, sample elements are often classified into disjoint

subsets, like intervals, rectangles etc. It means that it is not possible to give individual sample values but only the numbers of observations in each specified class. In this paper we consider the problem of estimation in the Cox model when survival data are grouped. We show how random numbers can be used to transform grouped lifetimes into pseudo-complete sample. The Fisher consistency of partial likelihood estimator of the regression parameters in the Cox model based on the restored samples is investigated. It is proved that for elliptical type distributional assumptions about explanatory variables the estimators of regression parameters in the Cox model based on the pseudo-complete sample are consistent up to a scaling factor. A simulation study illustrates asymptotic properties of the estimates as well as real data case analysis is presented.

Keywords: Cox model, grouped data, Fisher consistency, elliptical distribution

Dominik Mielczarek, Krzysztof Najman, Uniwersytet Gdański

Title: Convolutional Neural Networks for Estimating Timber Volume from Wind Damage in Stare Jabłonki

Abstract: Over the past decade, the number of recorded tornadoes in Poland has been increasing. According to statistics from the European Severe Weather Database (ESWD), between 2012 and 2023, the country experienced 184 tornadoes, with 103 occurring from 2019 to 2023. Catastrophic phenomena are becoming cyclical in nature. The damages caused by hurricanes pose a challenge for forest managers, as uprooted trees block forest access roads, and unlocalized ones lose their industrial usefulness. Rapid acquisition of information about the scale of damage allows for precise planning of work to eliminate the effects of hurricanes. The aim of the study is an attempt to estimate the volume of timber based on selected artificial intelligence methods. The subject of estimating the amount of harvested timber is motivated by the increasing number of hurricane disasters in Poland. In the study, a method was proposed that is based on conducting a UAV aerial imaging campaign. It requires the use of data obtained after wind damage has occurred. Due to its low complexity, the proposed method allows for the rapid provision of a complete set of information.

Sesja IXa

Radosław Pietrzyk, Paweł Rokita, Wrocław University of Economics and Business, Empirica Sp. z o. o.

Tytuł: Zależność miar płynności od poziomu cen i stóp zwrotu na rynku kryptowalut

W pracy badane są właściwości wybranych miar płynności należące do grupy miar naprężenia i głębokości dla rynku kryptowalut. Grupy te reprezentują różne wymiary płynności, a celem badań jest rozpoznanie zależności między nimi nawzajem oraz między nimi a poziomami cen i stopami zwrotu. Wynik badań może mieć znaczenie dla prognozowania płynności lub zmian cen. Może być wykorzystany na przykład przez animatorów rynku przy opracowywaniu strategii inwestycyjnych.

Słowa kluczowe: kryptowaluty, miary płynności

Title: Dependence of liquidity measures on price levels and returns in the cryptocurrency market

This paper studies the properties of selected liquidity measures in the cryptocurrency market. The measures belong to the groups of tightness and depth measures, which represent different dimensions of liquidity. The research aims to recognize the relationship between the measures themselves and between them and price levels and rates of return. The result of the research can be relevant for forecasting liquidity or price changes. It can be used, for example, by market makers in developing investment strategies.

Keywords: cryptocurrencies, liquidity measures

Title: Climate risk impact on credit risk

Abstract: Climate risk is one of the type of risks in the banks' portfolio which is not fully recognized and its impact on the future overall risk changes is hidden due to lack of correct knowledge at the moment. One of the most common data used for predicting future climate scenarios are included in Network for Greening the Financial System data (NGFS). NGFS is a group of 65 central banks and supervisors committed to sharing best practices, contributing to the development of climate and environment related risk management in the financial sector. Scenarios included in NGFS data are related to the climate change (physical risk) and climate policy and technology trends (transition risk). In the paper we are focused on the transition risk scenarios and their impact on economy and in particular on credit risk. Scenarios related to the transition risk can be grouped into the four segments: (i) orderly scenarios which assume that climate policies are introduced early and it's trend it to become more rigorous; (ii) disorderly scenarios assume that climate policies are introduced with delay or they are divergent across countries and sectors; (iii) hot house world scenarios assume that some climate policies are introduced in a restricted area which is not enough to stop climate changes; (iv) too little, too late scenarios is the scenario which assume no effort to counteract the climate change. The basic population we used to build the climate risk model is data related to credit risk observed in Southern Europe banks for mortgage products. We decided to build the probability of default model where the target variable is the binary default flag which means the entering into the default state in the next 12 months after data assessment. We took into account many default triggers used in banks but not only the basic one which is the delay in payment over 90 days. The average observed default rate in the analysed portfolio is 0.534%. Various data sources were used as the input for potential explanatory variables. Modelling data includes different information such as customer-oriented data, macroeconomic data, basic financial ratios, transactional data, behavioral data and collateral features. Data includes around 5,000,000 observations and was divided into the train and test data sets where the train data set includes 70% of the entire sample. Data covers the years from 2003-12-31 to 2019-09-30. Finally we decided to build logistic model on the described sample. Taking into account the high correlation between macroeconomic variables we decided to select the following two alternative models based on macroeconomic variables like: unemployment rate, crude oil and interest rates. Additionally, the following customer variables occurred to be predictive: average income in the last 6 months, current LTV and average savings from the last 3 months. The AUC of models is 0.73 and 0.61 respectively. The second model is on the border of its acceptance but is used as an alternative in our research.

Our main goal is to check the tendency in the default prediction in relation to climate risk potential future scenarios. Transition risk variables mainly focus on transition from fossil fuels, carbon-intensive production and consumption towards emission-neutral alternatives. They are estimated for different scenarios and cover transition risk drivers including final energy consumption, energy prices per sector and energy source, carbon taxes and prices, energy consumption, carbon sequestration, emissions, GDP, crops prices. We can additionally compare different customers' subsegments to draw conclusions on the impact of transition risk for different groups by industry and other key portfolio characteristics. This is a significant result on upcoming climate risk policy changes on Southern Europe society.

Keywords: climate risk, transition risk, PD model, NGFS data

Wioletta Grzenda, Olga Momot, Szkoła Główna Handlowa w Warszawie

Tytuł: Zalety metody Shapley additive explanations w analizie czynników determinujących spłatę stałych zobowiązań finansowych wobec nieplanowanych zdarzeń

Streszczenie: Zdarzenia nieplanowane takie jak choroby naturalne czy pandemie mogą znacząco wpłynąć na zdolność gospodarstw domowych do spłaty stałych zobowiązań finansowych. Powszechnym narzędziem

badawczym wykorzystywanym w analizie zdolności do spłaty zobowiązań finansowych jest model regresji logistycznej. Model ten jest coraz częściej wspomagany innymi modelami uczenia maszynowego, w tym drzewami decyzyjnymi. Modele drzew decyzyjnych oraz regresji spełniają podstawowy warunek oceny zdolności spłaty zobowiązań finansowych, jakim jest interpretowalność otrzymanych wyników. Jednakże na podstawie wartości oszacowań parametrów modeli regresji logistycznej nie można ocenić, która ze zmiennych wykorzystanych w modelu jest ważniejsza. Natomiast ważność zmiennych wyznaczana dla modeli drzew decyzyjnych, nie dostarcza informacji o kierunku ich wpływu na badane zjawisko. Celem tej pracy jest wskazanie na użyteczność metody Shapley additive explanations w ocenie wyników otrzymanych z wykorzystaniem regresji logistycznej oraz drzew decyzyjnych na poziomie globalnym. W badaniu wykorzystano dane z Badania Budżetów Gospodarstw Domowych w Polsce z lat 2020-2021. Analizowano, czy gospodarstwo domowe było zmuszone zrezygnować ze spłaty stałych zobowiązań finansowych w związku z pandemią COVID-19.

Słowa kluczowe: Spłata zobowiązań finansowych, Badanie Budżetów Gospodarstw Domowych, regresja logistyczna, drzewa decyzyjne, metoda Shapley additive explanations.

Title: Advantages of the Shapley Additive Explanations method in analyzing the determinants of repayment of financial liabilities in the context of unscheduled events

Abstract: Unscheduled events such as natural diseases or pandemics can significantly impact the ability of households to meet their financial liabilities. A common research tool used in the analysis of meeting financial liabilities is the logistic regression model. This model is increasingly complemented by other machine learning models, including decision trees. Both decision tree and logistic regression models satisfy the fundamental requirement for assessing the ability to repay financial liabilities, which is the interpretability of the obtained results. However, based on the estimated parameter values of logistic regression models, it is not possible to assess the importance of the variables used in the model. On the other hand, the importance of variables determined for decision tree models does not provide information about the direction of their impact on the studied phenomenon. The aim of this study is to demonstrate the utility of the Shapley Additive Explanations method in evaluating the results obtained using logistic regression and decision trees at a global level. Data from Household Budget Surveys in Poland from 2020-2021 were used in the analysis. It was investigated whether households were forced to defer repayment of fixed financial obligations due to the COVID-19 pandemic.

Keywords: Repayment of financial liabilities, Household Budget Survey, logistic regression, decision trees, Shapley Additive Explanations method.

Sesja IXb

Michał Bernard Pietrzak, Politechnika Gdańska

Tytuł: Zastosowanie modelu ilorazu potencjałów na poziomie regionalnym z wykorzystaniem odległości ekonomicznej

Streszczenie: W badaniach ekonomicznych stosowany jest powszechnie model potencjału, który pozwala na identyfikację makroregionów ze względu na badane zjawisko społeczno-gospodarcze. Pierwszym wykorzystywanym w literaturze modelem był model potencjału ludności, którego zastąpienie modelem potencjału dochodu zaproponowano w kolejnych badaniach. Najnowsza propozycja w literaturze przedmiotu dotyczy zastosowania modelu ilorazu potencjału, który pozwala na analizę na poziomie regionalnym. Model ilorazu potencjału wyznaczany jest jako stosunek wartości wynikających z modelu potencjału dochodu do wartości modelu potencjału ludności. W związku z tym w artykule przedstawiono

próbę zastosowania modelu ilorazu potencjału w badaniach regionalnych na poziomie gmin. Zaprezentowane zostaną wyniki empirycznego badania rozwoju regionalnego gmin województwa Kujawsko-Pomorskiego, gdzie w procedurze wyznaczania ilorazu potencjału zastosowano oprócz odległości fizycznej, również odległość ekonomiczną. Zastosowanie odległości ekonomicznej wynika z faktu, że odległość fizyczna nie jest w stanie w pełni opisać stopnia dostępności wybranego regionu dla pozostałych, opierając się wyłącznie na dystansie między regionami. Dopiero wykorzystanie odległości ekonomicznej powinno pozwolić na znacznie głębsze uwzględnienie zależności między regionami, które wynikają z ich wzajemnych powiązań społeczno-ekonomicznych.

Słowa kluczowe: model ilorazu potencjału, analiza regionalna, odległość ekonomiczna, odległość fizyczna

Title: Application of the potential quotient model at the regional level using economic distance

Abstract: A potential model is commonly used in economic research to identify macro-regions due to the socio-economic phenomenon under study. The first model used in the literature was the population potential model, the replacement of which by the income potential model was proposed in subsequent studies. The most recent proposal in the literature is the use of the potential quotient model, which allows analysis at the regional level. The potential quotient model is determined as the ratio of the values derived from the income potential model to the values of the population potential model. Accordingly, the article attempts to apply the potential quotient model to regional studies at the municipal level. The results of an empirical analysis of the regional development of the municipalities of the Kuyavian-Pomeranian Voivodeship will be presented, where, in addition to physical distance, economic distance was used in the procedure for determining the potential quotient model. Economic distance is used because physical distance cannot fully describe the degree of accessibility of a selected region to others, based solely on the distance between regions. Economic distance should allow for a much deeper consideration of the dependencies between regions resulting from their socio-economic interconnections.

Keywords: potential quotient model, regional analysis, economic distance, physical distance

Ewa Feder-Sempach, Piotr Szczepocki, Uniwersytet Łódzki

Tytuł: Zmiany strukturalne CAPM w przekroju krajów oraz branż. Wpływ pandemii Covid-19 na firmy europejskie

Streszczenie: Głównym celem artykułu jest analiza zmian strukturalnych parametru beta w modelu CAPM na europejskim rynku akcji przed i po globalnej pandemii Covid-19 w oparciu o indeks STOXX Europe 600. Próba badawcza składa się z 600 europejskich emitentów z różnych krajów i branż. Indeks oferuje zdywersyfikowaną alokację, która odzwierciedla prawie 90% europejskiego rynku akcji. Metodologia badania opiera się na modelu wyceny aktywów kapitałowych (CAPM). Zastosowano kilka testów niestabilności parametrów i zmian strukturalnych o nieznanym punkcie zmiany. W badaniu wykorzystano miesięczne ceny akcji 600 spółek notowanych na głównych giełdach europejskich (tj. Euronext, London Stock Exchange, NASDAQ, Six Swiss Exchange i XETRA). Te miesięczne ceny akcji, indeksy krajowe i stopy wolne od ryzyka (3-miesięczne bony skarbowe) zebrano dla lat 2014–2022. Obserwacje uzyskano z bazy danych Refinitiv EIKON.

Słowa kluczowe: parametr beta, CAPM, Zmiana strukturalna

Title: CAPM Structural Change across Countries and Industries. The Impact of the COVID-19 Pandemic on European Companies

Abstract: The main objective of the paper is to analyse the structural change of CAPM beta in the European stock market before and after the global pandemic of COVID-19 based on the STOXX Europe 600 index. The

research sample consists of 600 European issuers from different countries and industries. The index offers a diversified allocation that replicates almost 90% of the European stock market. The research methodology is based on Capital Assets Pricing Model (CAPM). Several tests for parameter instability and structural change with un-known change point were applied.

In this study, monthly stock prices of 600 companies listed on main European stock exchanges (i.e., Euronext, London Stock Exchange, NASDAQ, Six Swiss Exchange and XETRA) were used. These monthly stock prices, home country indices, and risk-free rates (3 months T-Bills) were collected between 2014 and 2022. The observations were obtained from the Refinitiv EIKON database.

Keywords: beta parameter , CAPM , structural change

Maciej Oesterreich, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Wykluczenie cyfrowe a realizacja wybranych celów zrównoważonego rozwoju w krajach UE w 2021 roku

Streszczenie: Dynamiczny rozwój technologii cyfrowych, a także Internetu w latach 90 XX. oraz początku XXI w sposób istotny wpłynął na kierunki rozwoju społecznego, gospodarczego oraz kulturalnego społeczeństw wszystkich państw Świata. Negatywną konsekwencją tego procesu jest zjawisko wykluczenia cyfrowego. Celem niniejszej pracy jest pomiar tego zjawiska w krajach UE w 2021 roku oraz jego ocena w kontekście celów zrównoważonego rozwoju. W tym celu skonstruowano taksonomiczne mierniki syntetyczne wykorzystując metodę TOPSIS. Za źródło danych posłużyła baza Eurostat, a dotyczyły one wybranych zagadnień pochodzących z badania ICT (Information and Communication Technologies) oraz wskaźników opisujących poziom zrównoważonego rozwoju. Z wyników przeprowadzonych badań wynika, że kraje UE były wyraźnie zróżnicowane pod względem poziomu wykluczenia cyfrowego. Miało to również wpływ na realizację przez nie, wybranych celów zrównoważonego rozwoju.

Słowa kluczowe: Wykluczenie cyfrowe, zrównoważony rozwój, UE

Title: Digital exclusion and the implementation of selected sustainable development goals in EU countries in 2021

Abstract: Dynamic development of digital technologies and the Internet in the 1990s. and the beginning of the 21st century significantly influenced the directions of social, economic and cultural development of societies in all countries of the world. A negative consequence of this process is the phenomenon of digital exclusion. The aim of this work is to measure this phenomenon in EU countries in 2021 and assess it in the context of the sustainable development goals. For this purpose, synthetic taxonomic measures were constructed using the TOPSIS method. The data source was the Eurostat database, and they concerned selected issues from the ICT study (Information and Communication Technologies) and indicators describing the level of sustainable development. The research results show that EU countries were clearly differentiated in terms of the level of digital exclusion. This also had an impact on their implementation of selected sustainable development goals.

Keywords: Digital exclusion, sustainable development, EU

Anna Majdzińska, Uniwersytet Łódzki

Tytuł: Ocena potencjału demograficznego obszarów Polsk

Streszczenie: W referacie zaprezentowane zostaną wyniki klasyfikacji obszarów Polski (tj. województw, powiatów i gmin) z punktu widzenia potencjału demograficznego, przy wykorzystaniu wybranych miar analizy skupień oraz syntetycznych mierników rozwoju (klasycznych i ich przestrzennych modyfikacji).

W niniejszym opracowaniu, termin ‘potencjał demograficzny’ traktowany jest dość szeroko, tzn. uwzględnia zarówno stan i strukturę wieku populacji, jak i procesy reprodukcji oraz migracji ludności badanych obszarów. Źródłem danych wykorzystanych w analizach są publikacje Głównego Urzędu Statystycznego, dotyczące struktury wieku populacji oraz zdarzeń z zakresu ruchu naturalnego i migracyjnego ludności w wybranych latach okresu 2002-2022.

Session IXC

Józef Dziechciarz, Marta Dziechciarz, Uniwersytet Ekonomiczny we Wrocławiu

Title: Peoples’ with Disabilities Households’ Material Situation. Subjective Assessment

Abstract: The study aims to determine the mechanism of time variability in formulating income expectations. The analysis is devoted to the subjective assessment of the material well-being of households in which there are people with disabilities. Knowledge of the described mechanisms is crucial in designing social and income policy at the country level. Regression methods recommended for research based on panel data were used. The results indicate a great diversity in formulating income expectations of the surveyed households in particular types of families.

Second Bwanakare, UKSW; USR

Title: Energy efficiency Forecast based on sparse and conflicting Information

This paper proposes an entropy econometric technique to predict energy efficiency at region (NUT-2) level based on imperfect knowledge of the national overall efficiency in the sectors of industry, transport, households and services. The solution of this stochastic inverse problem is applied to Poland. As acknowledged in recent literature, non-extensive entropy model should remain a valuable device for econometric modelling even in the case of low frequency series since outputs provided by the Gibbs-Shannon entropy approach correspond to the Tsallis entropy limiting case of the Gaussian law when the Tsallis q -parameter converges to unity. We, therefore, set up a q -Tsallis-Kullback-Leibler entropy criterion function with a priori consistency constraints, including province energy intensity (known with uncertainty). Fortunately, adding more pertinent data to the model priors will enhance parameter precision and then allow for the recovery of the real influence of smaller events. The model outputs continue to conform to empirical expectations. In spite of the close to unity q -Tsallis parameter, this Tsallis related approach reflects higher stability for parameter computation in comparison with the Shannon-Gibbs entropy econometrics technique. The proposed technique can be applied in different countries for example in the context of experimental official statistics.

Keywords: overall energy efficiency score, energetic intensity, non-extensive cross-entropy econometrics, stochastic inverse problems, regional innovation

Błażej Kocharński, Politechnika Gdańska

Tytuł: The shape of an ROC curve in the evaluation of credit scoring models

Streszczenie: AUC – pole pod krzywą charakterystyki operacyjnej odbiornika (ROC) – lub jego przeskalowana wersja, współczynnik Giniego, są standardowymi miarami mocy dyskryminacyjnej scoringu kredytowego. Za pomocą modeli binormalnych krzywych ROC, pokazujemy, jak kształt krzywych wpływa na korzyści ekonomiczne wynikające z zastosowania modeli scoringowych o tej samej wartości AUC. W oparciu o wyniki proponujemy, aby parametr kształtu dopasowanej krzywej ROC był podawany wraz z jej AUC / Gini za każdym razem, gdy omawiana jest jakość karty scoringowej.

Słowa kluczowe: scoring kredytowy, krzywa ROC, AUC

Abstract: AUC – the area under the receiver operating characteristic (ROC) curve – or its scaled version, the Gini coefficient, are the standard measures of the discriminatory power of credit scoring. Using binormal ROC curve models, we show how the shape of the curves affects the economic benefits of using scoring models with the same AUC. Based on the results, we propose that the shape parameter of the fitted ROC curve should be reported alongside its AUC / Gini whenever the quality of a scorecard is discussed

Keywords: credit scoring, receiver operational characteristic, AUC

Ryszard Szupiluk, Paweł Rubach, Szkoła Główna Handlowa w Warszawie

Tytuł: Grupowanie finansowych szeregów czasowych z wykorzystaniem analizy składowych ukrytych

W niniejszym artykule przedstawimy zastosowanie metod analizy składowych ukrytych do analizy podobieństwa finansowych szeregów czasowych, co może być podstawą dla nowej koncepcji grupowania instrumentów finansowych. Cała procedura składać się będzie z etapu identyfikacji ukrytych komponentów o określonej charakterystyce statystycznej a następnie analizie relacji między nimi.

Słowa kluczowe: analiza składowych ukrytych, segmentacja, analiza szeregów czasowych

Title: Clustering Financial Time Series using Latent Component Analysis

This paper presents the use of Latent Component Analysis methods to analyze the similarity of financial time series, which may be the basis for a new concept of grouping financial instruments. The entire procedure will consist of the identification of latent components with specific statistical characteristics and then the analysis of the relationships between them.

Keywords: Latent Component Analysis, clustering, time-series analysis