SBORNÍK

prací účastníků vědeckého semináře
doktorského studia
Fakulta informatiky a statistiky
Vysoké školy ekonomické

Abstrakty

Vědecký seminář se uskutečnil dne 2. února 2023
pod záštitou děkana FIS
prof. Ing. Jakuba Fischera, Ph.D.

Sestavení sborníku
prof. Ing. Petr Doucek, CSc.
proděkan pro tvůrčí činnost a zahraniční vztahy

Nakladatelství Oeconomica – Praha 2023

ISBN 978-80-245-2485-6

OBSAH

STUDIJNÍ PROGRAM STATISTIKA

Předmluva

Jak se stalo dobrou tradicí, tak na přelomu ledna a února proběhla tradiční akce Fakulty informatiky a statistiky „Den doktorandů“. V tomto roce se jednalo již o dvacátý osmý ročník. Seminář se konal 2. února 2023 pod gescí děkana Fakulty informatiky a statistiky prof. Ing. Jakuba Fischera, Ph.D.

Přestože má fakulta tři programy doktorského studia, celkový počet doktorandů, kteří presentovali výsledky své vědecké práce byl poměrně skromný – celkem deset doktorandů. Do tohoto kroužku přispěl program „Aplikovaná informatika“ (AI) dvěma účastníky, stejně tak jako program „Statistika“ (ST). Největším počtem účastníků přispěl program „Ekonometrie a operační výzkum“ (EOV).

Hodnoticí komise posoudila vystoupení doktorandů a na jejich základě určila tři nejlepší vystoupení bez ohledu na doktorský studijní program FIS. V letošním roce tak získali prestižní „Cenu děkana FIS“, s níž je spojena i symbolická finanční odměna, následující studentky a studenti:

1. místo: Ing. Lukáš Veverka, Econometric analysis of the media investments (EOV)
2. místo: Ing. Lucie Dvořáčková, Automated detection of future Impactful scientific articles and entities (EOV)
3. místo: Ing. Lukáš Sýkora, Action Rules via APRIORI (AI)

Oceněným studentům doktorského studia upřímně blahopřeji a pevně věřím, že získané zkušenosti uplatní při své další práci, ať už vědecké nebo v praxi. Uznání také patří všem vědeckým a pedagogickým pracovníkům FIS – školitelům doktorandů, kteří se „Dne doktorandů“ zúčastnili a svým vedením a radami byli nápomocni při zpracování příspěvků.

Nedílnou součástí „Dne doktorandů“ je i práce hodnoticích komisí, jejichž členové pečlivě sledují jednotlivá vystoupení a potom

vybírají nejlepší práce k ocenění. Za práci v hodnoticí komisi děkuji prof. Ing. Vojtěchu Svátkovi, Dr. (Katedra informačního a znalostního inženýrství), prof. RNDr. Ing. Michalu Černému, Ph.D. (Katedra ekonometrie), prof. Ing. Haně Řezankové, CSc. (Katedra statistiky a pravděpodobnosti). Komise se zhostily své práce na výbornou.

Na závěr bych chtěl vyjádřit zvláštní poděkování studijní referentce doktorského studia paní Ing. Tereze Krajíčkové, DiS, díky níž byl seminář skvěle organizačně zajištěn, dále paní Petře Šarochové za administrativní podporu akce a Mgr. Lee Nedomové za práci při editaci a sestavení tohoto sborníku abstraktů.

prof. Ing. Petr Doucek, CSc.

proděkan pro tvůrčí činnost a zahraniční vztahy

STUDIJNÍ PROGRAM APLIKOVANÁ INFORMATIKA

Action Rules via APRIORI

Lukáš Sýkora

Lukas.sykora@vse.cz

Ph.D. student of Applied informatics

Supervisor: doc. Ing. Tomáš Kliegr, Ph.D. (tomas.kliegr@vse.cz)

Action rule mining is an extension of the widely used task of learning classification rules. In addition to information expressed in a standard classification rule, an action rule suggests a course of action. This text describes the possibilities of mining action rules using APRIORI. In particular, it focuses on the author's implementation in Python.

The mining of action rules is usually divided into two independent steps (The rule-based approach). In the first step, classification rules are mined by any suitable association rule mining algorithm (for example ARIORI modified for Classification Association Rules). In the second step, action rules are discovered from the classification rules (for example DEAR algorithm).

The text focuses on improving the first step in the action rules mining, which is the mining of classification rules. It explains the possibilities of pruning candidates for action rules in APRIORI iterations based on reduction trees. These Reduction trees in APRIORI work on the principle that candidates are divided into smaller groups so that each group contains only candidates with the same stable attribute. These groups of candidates can be excluded if they do not contain at least one unwanted target and at least one wanted target at the same time. They can also be excluded if the pre-change or post-change part does not reach the set minimum support. With these modifications, it is possible to achieve better performance of APRIORI for action rules. Another advantage is that fewer rules enter the second part, which brings additional savings in performance.

Keywords: Action Rules, Explainable Machine Learning, Rule Learning, APRIORI

JEL Classification: L86, D83

IIoT sensor data prediction by machine learning methods

Nikola Kuchtíková

nikola.kuchtikova@vse.cz

Ph.D. student of Applied informatics

Supervisor: doc. Ing. Miloš Maryška, Ph.D.
(milos.maryska@vse.cz)

The paper aims to provide initial insight into how, after collecting the data from the IIoT sensor, we can analyse it to make predictions using machine learning methods. From IIoT sensors continuously monitoring power, temperature, pressure, vibration, and other factors are typically collected time series data. There are three appropriate machine learning algorithms selected in the paper – recurrent neural network, autoregressive integrated moving average (ARIMA) and gradient boosting. ARIMA, the well-known model in statistics, was used for a use case solving a time series problem. For this paper, a model is based on a simple dataset of the date and number of gas supply disruptions. The time series problem is demonstrated through the following processes: (1) visualizing time series, (2) stationarising time series, (3) finding the best parameters for our model, (4) fitting the model (we can fit an ARIMA model to learn the pattern of the series) and (5) predictions. The prediction model seeks to use knowledge of the past to explain what should be expected in the future at various periods, which could be very beneficiary for any Industry that makes use of IIoT.

Keywords: Predictive modelling, IIoT, sensors, time series, machine learning, recurrent neural network, ARIMA, XGBoost.

JEL Classification: C45, M15, O13, O14, 031

STUDIJNÍ PROGRAM EKONOMETRIE A OPERAČNÍ VÝZKUM

Econometric analysis of the media investments

Lukáš Veverka

lukas.veverka@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: doc. Ing. Jan Zouhar, Ph.D. (zouharj@vse.cz)

Evaluation of whether an advertisement was profitable or not is an essential part of marketing. This research proposes a methodology for the estimation of the immediate response to a TV ad and for the analysis of the long-term effect of the ad. Both methods are applied to a dataset containing minute-by-minute organic website visits and detailed characteristics of TV ads for an e-commerce company in 2019. To estimate the immediate (within a few minutes) responses to a TV ad it is necessary to capture diurnal and seasonal patterns and estimate a gradual increase in website visits after an ad. The results can be tested for non-linear dependencies on the characteristics of the ads. The results show that people are willing to switch between screens and multitask. To calculate the long-term effect of TV ads, it is necessary to deal with the effect of seasonality and highly non-linear relationships (response curves, carry-over effect) between advertisements and KPIs. Based on the results it is possible to decide on common marketing questions (ROI, budget optimization, factors affecting KPIs).

Keywords: Data-driven marketing, Dynamic models, Ad Effectiveness, Website Traffic, TV Advertising

JEL Classification: C22, M37

Do KovaaK's players have hot hands?

Jan Rejthar

rejj02@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: doc. Ing. Jan Zouhar, Ph.D., (zouharj@vse.cz)

This paper study whether the hands of gamers can get hot. It identifies three major issues occurring in the sports literature – complexity of sports resulting in many sources of complex endogeneity, shortcomings of many sports with regards to what data they produce and lack of statistical tools powerful enough to be able to reliably detect the Hot Hand when present. The paper employs data from players of an aim trainer called KovaaK's and argues that these KovaaK's data are robust to the first two issues and suffer from them less than any data used previously. To resolve the third issue, the paper employs permutation tests, since Ritzwoller & Romano (2021) showed that these tests do not require a correction for finite-sample bias and proved that permutation tests are the only tests that control the type 1 error rate exactly in finite samples.

Keywords: Hot Hand, permutation tests, gaming, KovaaK's

JEL Classification: C12, C18, Z29

The Role of Precommitment in Consumption

Petr Krautwurm

petr.krautwurm@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: prof. RNDr. Ing. Michal Černý, Ph.D., (cernym@vse.cz)

This article illustrates that precommitment plays a key role in consumer decision making. Further, it explains multiple phenomena such as delayed consumption, the preference for flexibility or inflexibility, and mental accounting as the consequence of this omnipresence of precommitment in decision making. It then proposes the existence of two effects measuring the implications of consumer's decisions on his or her subsequent consumption possibilities. These effects could be used for decomposition of uncertainty. The essential part of the article is the model, predicting that uncertainty increases with a time delay between the planning of an action and its realization. In conclusion, the main proposal of the article is that precommitment and connected effects could potentially be used as a first step in deriving the metric for opportunity.

In this article it is shown that consumers would always have to precommit because they have no other option. Therefore, the article develops the idea that the whole decision process of consumers is solely the choice of optimal sequence of precommitments. According to this, life is nothing more than continuously restricting oneself in the most suitable way. Furthemore, if any decision a consumer can make is a precommitment, it means that his or her behaviour might be analyzed through such perspective. In this matter, the article shows the importance of inventory management in consumers' decision making because most economic activities behave as if they were inventory operations; therefore, this inventory-like nature of decision making is the most common precommitment consumers have to undergo. The article then builds on this idea and shows main implications such as: the possibility to decompose uncertainty into two components based on how much the

consumer precommits, the possibility to decompose preferences based on which precommitment the consumer chooses, the similar implication for sunk costs paradox, and the implication for paternalism. Thus, the main contribution of this article is the generalization of precommitment, which is achieved by showing that previous definition of precommitment was inconsistent with conclusions based on it. The article might be relevant for non-academic reader in a way that it emphasize the importance of quality of prediction of one's future preferences in their welfare. The superior knowledge of someone's future preferences is required to be proved by paternalists in order to proclaim their policy to be even welfare improving.

For these purposes the article constructs a mathematical optimization model which satisfies the core of the main argument. The argument consists of following factors: The first one is the assumption that there is only infinitesimal difference between dynamically consistent consumer and an inconsistent one, thus, the consistency is a parametrical question rather than cathegorical. This assumption is relevant because dynamic consistency means that ordering of preferences is not a function of time. Due to that, it is obvious that ordering of preferences might change either slightly or substantially; therefore, any deviation from the stable ordering means that the consumer would be inconsistent. The second factor is acknowledging that any choice changes the future prospects, meaning that possibilities are always path dependent. Under this setup, the rational, inconsistent, and altruistic-towards-future-selves consumer would always be willing to pay something solely for a change in his or her path. This implies that the most profound definition of precommitment up to date is inefficient and is not consistent with implications based on it. The problem resides in one of the necessary conditions required for an action to be called a precommitment: the "purpose" condition. According to this condition, an action could be called a precommitment only if the consumer would be willing to pay for the restriction possibility alone. However, this condition is always satisfied, which renders it blank. The article thus shows how the correct model of consumer behaviour should be constructed and offers a prototype model on which subsequent models might be build.

Keywords: Precommitment; Uncertainty; Preference for flexibility; Changing tastes; Sunk costs fallacy; Metric for opportunity

JEL Classification: D01; D11; D15; D84; D90

A state-space modeling approach to assessing monetary policy performance

Dominik Kavřík

kavd00@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: prof. RNDr. Ing. Michal Černý, Ph.D., (cernym@vse.cz)

This study examines the use of state space modeling in the evaluation of monetary policy. The empirical approach to monetary policy evaluation is carried out using the concept of the Taylor curve, named after John B. Taylor. The Taylor curve graphically illustrates the trade-off between the variance of the output gap and the difference between actual and target inflation (inflation gap). This trade off can be derived from second-order approximations in dynamic stochastic general equilibrium (DSGE) models, which are also an option for assessing the policy efficiency, although from a theoretical perspective. This study primarily focuses on the empirical approach. The output gap is estimated using the Hodrick-Prescott (HP) filter on the real GDP time series, and the conditional volatilities of inflation and the output gap are estimated using a multivariate volatility model. In time periods where the monetary policy effectively responds to endogenous shocks, this negative relationship should hold. A time-varying parameter model is then estimated using the Kalman filter to obtain time-varying parameters describing the relationship between the variance of the output gap and the variance of the inflation gap. These findings have important implications for central banks, as the time-varying parameter model output indicates the periods in which the negative relationship holds and the periods in which monetary policy might be considered inefficient under the specified assumptions of the Taylor curve.

Keywords: Monetary policy, State-space models, Efficiency

JEL Classification: E31, E32, E52, C13, C32

The Construction of a Composite Indicator Using the Two – Stage "Benefit of the Doubt" Approach

Jakub Hanousek

xhanj52@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: prof. Ing. Mgr. Martin Dlouhý, Dr., MSc

This article deals with the two-stage construction of a composite indicator based on the "Benefit of the Doubt" approach. The model of the "Benefit of the Doubt" is based on a non-parametric method called data envelopment analysis. An arrangement of the criteria by experts can be controversial in many ways. The biggest advantage of this approach is that does not need input information about the importance of the criteria.

The application of the method is shown in the data about municipalities with more than 3,000 inhabitants in the Czech Republic. The borderline of 3,000 inhabitants was chosen because data is rarely available for smaller municipalities. Municipalities with more than 3,000 inhabitants were divided into three groups according to the number of inhabitants. This division was made for greater homogeneity between municipalities. The assessment of municipalities was carried out in 8 different areas. Each of these areas contained several criteria. The results from these 8 areas are aggregated using the data envelopment model into the overall result. Municipalities located on the border of the data envelope are the municipalities with the best results. Municipalities that lie on the border of the data envelope are further analyzed through the super-efficiency model. Finally, the results are subjected to a robustness test using correlation analysis.

The total number of analyzed municipalities was 456 according to 63 criteria. For each municipality, we obtained the value of the composite indicator. Smaller municipalities have bigger differences than bigger municipalities in the results from the first stage. The value of the

aggregated composite indicator did not confirm the dependence on the size of the municipalities.

Keywords: Composite indicators, Benefit of the doubt, Municipalities

JEL Classification: C44

Automated detection of future Impactful scientific articles and entities

Ing. Lucie Dvořáčková

berl03@vse.cz

Ph.D. student of Econometrics and operations research

Supervisor: prof. RNDr. Ing. Michal Černý, Ph.D., (cernym@vse.cz)

We propose a unified procedure based on document classification, which enables simultaneous detection of both articles and entities that may have a high scientific impact in the future. To express the scientific impact, we use data on the marginal annual citations of an article, which express the relevance of a given research in a given year. The classification model works with aggregated word embeddings as input variables at the document level.

As the individual components (features) in word embeddings do not have any meaning, it is not possible to directly apply existing model-agnostic methods for interpreting black-box models through feature importance scores such as LIME and SHAP. We therefore propose a new feature importance algorithm that allows explaining a supervised model (logistic regression) trained on an embeddings-based representation by applying the model to documents corresponding to embedding word vectors for single words.

We demonstrate the whole approach on the CORD-19 corpus of biomedical articles showing the ability of the proposed method to search for a specific thematic area of entities with potential scientific impact--- such as drugs, vaccines or SARS-CoV-2 virus variants.

Keywords: text mining, embedding, scientometrics, feature importance

JEL Classification: C53: Forecasting and Prediction Methods, Simulation Methods, C52: Model Evaluation, Validation and Selection.

STUDIJNÍ PROGRAM STATISTIKA

Interest rates and their statistical relationship to other macroeconomic variables since 1960

Tomas Stastny
xstat07@vse.cz
Ph.D. student of Statistics
Supervisor: prof. RNDr. Luboš Marek, CSc. (marek@vse.cz)

An interest rate is an amount that a borrower pays to a lender as a price for getting access to borrowed money. Interest rates influence various areas of our lives, starting with obvious things like deposits on our bank accounts, to interests that we have to pay for our loans and finally with their hidden impact on our currency, wages and prices in shops. Their impact is tangible all around. Motivation for this contribution is the current sudden increase in interest rates that has occurred across the globe since the end of the Covid-19 pandemics. Interest rates are closely linked with various macroeconomic variables, in particular with inflation, output gap and unemployment rates. In the contribution, I focused on short-term and long-term interest rates and their developments since 1960. Their developments were modelled by using the above-mentioned macroeconomic variables. In the contribution, I primarily analyzed the US and Swiss data thanks to their availability. Various statistical approaches (e.g. dynamic correlations) were applied, summarized and presented. A comparison of the current situation with late seventies and early eighties was provided. In addition, the famous Taylor rule was implemented on the US and Swiss data to model developments of short-term interest, i.e. the rates announced and strongly influenced by central banks.

Keywords: interest rates, inflation, output gap, unemployment

JEL Classification: E43, E44, C58, C32

International comparison of total fertility rate

Petra Smolíková

smop02@vse.cz

Ph.D. student of Statistics

Supervisor: doc. Ing. Jitka Langhamrová, CSc.
(jitka.langhamrova@vse.cz)

The article deals with the analysis of total fertility rate of Word population, mainly in countries of European union. Women in the European Union have less children, which contributes to a slowdown in natural population growth or even a negative population movement (more deaths than births). For this analyse total fertility rate was compared to socio-economic factors that may influence fertility in 27 countries. It concerns the unemployement rate of women, the share of women with a tertiary education and mean age of the mother at childbirth. Countires of the European Union have different approaches to reproductive behavior. For that reason, the development of total fertility and the mean age of the mother at childbirth were observed in more detail. The countries were divided into 5 groups based on historical-geographic sections. The development of reproductive behavior in these groups was monitored in 2000, 2010 and 2020. The aim is to analyse and evaluate the basic trends in international and regional differentiation of fertility by my own analysis and try to explain the observed spatial differences.

Keywords: Fertility, fertility rate, reproductive behavior, european union

JEL Classification: J13