АВТОМАТИЗИРАНИ АЛГОРИТМИ ЗА ИДЕНТИФИКАЦИЯ НА ARIMA МОДЕЛИ ПРИ ПРОГНОЗИРАНЕ НА ДИНАМИЧНИ РЕДОВЕ – ПРЕГЛЕД НА ЛИТЕРАТУРАТА

Доц. д-р Маргарита Шопова
Гл. ас. д-р Евгени Овчинников[1]

Резюме

Прогнозирането има ключово значение за ефективното планиране и управление и в икономическия, и в публичния сектор. Създаването на точни прогнози изисква значителна експертиза и е трудоемък процес, подлежащ на субективни грешки, особено при обработката на големи обеми данни. Това обуславя необходимостта от автоматизация на процесите на прогнозиране в съвременната практика.

Целта на изследването е да се изведат предимствата и ограниченията на различни автоматизирани алгоритми за идентификация на ARIMA модели и да се установи в каква степен тези алгоритми намаляват субективизма в процеса на моделиране.

Изследването включва три основни задачи: (1) да се обоснове необходимостта от автоматизация на процеса на прогнозиране и се разкрият предизвикателствата пред осъществяването й, както и да се проследи нейната еволюция; (2) да се аргументира автоматизирането на идентификацията на ARIMA модели въз основа разкриване на тяхната същност и ключовите етапи в процеса на моделиране; (3) да се представят и сравнят подходи за автоматизиране, които решават проблеми на идентификацията на ARIMA модели с цел последваща оценка на тяхната ефективност и приложимост.

Резултатите от изследването потвърждават тезата на авторите, че автоматизираните подходи за идентификация на ARIMA модели значително намаляват субективизма и времето, необходимо за анализ, но не могат напълно да елиминират намесата на изследователя. Научните приноси на разработката могат да се обособят в две направления: автоматизацията на процеса на идентификация преодолява ограниченията, характерни за класическото прилагане на процедурата; извършен е критичен анализ с извеждане на силните и слабите страни на съществуващи алгоритми за автоматизация.

Ключови думи: автоматизация на прогнозирането, ARIMA модели, автоматизирани алгоритми за идентификация на ARIMA модели.

JEL: C22, C53, C87.

[1] Разпределението на приноса на авторите при написване на студията е следното: доц. д-р М. Шопова – увод, т. 1.1, т. 2, заключение , гл. ас. д-р Е. Овчинников – резюме, т. 1.2, т. 2, т. 3, заключение.

AUTOMATIC ALGORITHMS FOR ARIMA MODEL IDENTIFICATION IN TIME SERIES FORECASTING: A LITERATURE REVIEW

Assoc. Prof. Margarita Shopova, PhD
Head Assist. Prof. Evgeni Ovchinnikov, PhD

Abstract

Forecasting plays a crucial role in effective planning and management within both the economic and public sectors, especially in a rapidly evolving environment. Despite its importance, producing accurate forecasts requires considerable expertise and is a labour-intensive process prone to subjective errors, especially when processing large volumes of data. These challenges make the role of automation of forecasting processes increasingly necessary in modern practice.

The objective of this study is to derive the advantages and shortcomings of various automatic algorithms for ARIMA model identification and to determine the extent to which these algorithms reduce subjectivism in the modelling process.

The research includes three main tasks: (1) to justify the necessity of automation of the forecasting process and reveal the challenges to its implementation, as well as to trace its evolution; (2) to motivate the automation of ARIMA model identification on the basis of revealing their nature and the key stages in the modelling process; (3) to present and compare automatic approaches that solve ARIMA model identification challenges in order to further evaluate their effectiveness and applicability.

The results of the study show that automated ARIMA algorithms provide high accuracy and reliability of forecasts, significantly reducing the time and resources required for the analysis. In addition, automation facilitates adaptation to changing data and conditions, which is key to making timely decisions in a dynamic environment. The findings of this study can assist organisations in selecting appropriate automated methods for time series forecasting.

Key words: Forecasting automation, ARIMA models, Automatic algorithms for identification of ARIMA models.

JEL: C22, C53, C87.

Увод

В съвременните условия на бързо променяща се икономическа и социална среда прогнозирането заема ключова роля за успешното планиране и управление както в икономиката, така и в публичния сектор. То намира приложение както във финансите, енергетиката и логистиката, така и за оптимизиране на обществените услуги в транспорта, здравеопазването, в управлението на кризи. Основната цел на прогнозирането е да предоставя точна и надеждна информация, необходима на органите на управление за вземане на информирани решения и разработване на устойчиви стратегии.

Изготвянето на точни прогнози е комплексен процес, който изисква задълбочена експертиза, особено в областта на статистическия анализ. За

постигане на висока точност е необходимо, коректно да се моделират сложните зависимости в наличните данни. С развитието на цифровите технологии обемът от данни за прогнозиране нараства, което създава нови предизвикателства пред специалистите по прогнозиране. Те са изправени пред трудоемки и времеемки задачи, което прави процеса продължителен и предразположен към субективни грешки.

Тези предизвикателства правят автоматизацията на прогнозните процеси необходимост в съвременната практика. Автоматизираните методи позволяват създаването на прогнози с по-висока бързина и устойчивост, като същевременно намаляват времето и ресурсите, необходими за анализа. Освен това автоматизацията улеснява адаптирането към нови данни и променящи се условия, което прави прогнозите по-надеждни и гъвкави. С автоматизирани прогностични системи могат бързо и ефективно да се създават прогнози, което е от ключово значение за управлението и стратегическото планиране в динамична среда.

Обект на настоящото изследване са процесите на автоматизация при моделирането и анализа на динамични редове като част от общата тенденция за автоматизиране на прогнозирането, а **предмет** – приложимостта на различни подходи за автоматизацията на процеса на идентификация на ARIMA модели.

Целта на научното изследване е да се изведат предимствата и ограниченията на различни автоматизирани алгоритми за идентификация на ARIMA модели и да се установи в каква степен тези алгоритми намаляват субективизма в процеса на моделиране.

За реализирането на поставената цел са дефинирани следните **задачи**:

- да се обоснове необходимостта от автоматизация на процеса на прогнозиране и се разкрият предизвикателствата пред осъществяването й, както и да се проследи нейната еволюция;
- да се аргументира автоматизирането на идентификацията на ARIMA модели въз основа на разкриване на тяхната същност и ключовите етапи в процеса на моделиране;
- да се представят и сравнят подходи за автоматизиране, които решават проблеми на идентификацията на ARIMA модели с цел последваща оценка на тяхната ефективност и приложимост.

Изследователската теза на авторите в настоящото изследване е, че автоматизираните подходи за идентификация на ARIMA модели значително намаляват субективизма и времето, необходимо за анализ, но не могат напълно да елиминират намесата на изследователя.

Методологията на изследването включва използването на индуктивния и дедуктивния подход, метода на анализа и синтеза, сравнителния анализ и анализ на съдържанието. Основният метод е библиографски преглед на теоретични и емпирични изследвания в тематичната област на ARIMA моделирането.

1. Прогнозиране и ролята на автоматизацията

1.1. Значение и приложения на прогнозирането

Дългосрочното стратегическо планиране в бизнеса и информираното вземане на решения относно текущото планиране на производствената дейност, транспорта, персонала и други са неразривно свързани с прогнозирането. То се състои в предвиждане на бъдещето с максимална точност въз основа на наличната информация, включително данни за минали периоди и знания за бъдещи събития, които биха могли да влияят на прогнозите.

Бизнес прогнозите могат да се построят неефективно и понякога се заменят с целеполагане и планиране. Поставените цели се отъждествяват с желаното развитие, но е възможно да липсва план, как да се постигнат. Същевременно, ако бъдат обвързани с прогнози, може да се установи дали са реалистични. Планирането, от своя страна, се състои в определянето на действия, необходими за постигане на съответствие между прогнозите и целите.

Прогнозирането трябва да бъде неразделна част от управленските решения, тъй като може да играе важна роля в многообразието от дейности на различни компании или институции. Прогнозите могат да бъдат изисквани както за години напред (в случай на капиталови инвестиции), така и за кратки периоди от няколко минути (например за маршрутизация в телекомуникациите). Без значение от конкретните обстоятелства или времеви хоризонти, прогнозите представляват ценен инструмент за ефективно и ефикасно планиране на ресурси и дейности.

Модерните организации се нуждаят от краткосрочни, средносрочни и дългосрочни прогнози в зависимост от конкретните управленски задачи. Краткосрочните прогнози са необходими, например, за съставяне на графика на персонала, планиране на обема на производството, необходимостта от транспортиране на готовата продукция. Чрез средносрочните прогнози се определят бъдещите потребности от ресурси с цел осигуряване на суровини, наемане на персонал или закупуване на машини и оборудване. Дългосрочните прогнози се използват в стратегическото планиране. При разработването им трябва да се вземат предвид пазарните възможности, различни екологични фактори, вътрешните ресурси, както и неизменно съпътстващите бъдещото развитие на всяко явление несигурност и риск.

В публичния сектор прогнозирането подпомага планирането на обществени услуги и разпределението на бюджетни средства. Прогнозирането на транспортни нужди, например, се използва за предвиждане на натовареността на пътищата и оптимизиране на инфраструктурата (Vlahogianni, Karlaftis, & Golias, 2014). В здравеопазването прогнозите намират широко приложение за разходите за медикаменти (Kaushik, et al., 2020) и превенция на епидемии (Desai, et al., 2019). По време на епидемии доброто прогнозиране позволява предвиждане на разпространението на заболяването и по-

добро управление на здравните грижи. Подпомага се разпределението на болнични легла и здравен персонал според очакваната необходимост. В сферата на околната среда и климата прогнозите са важен инструмент за планиране на мерки за адаптация към климатичните промени (Fildes & Kourentzes, 2011). Дългосрочните климатични прогнози помагат на правителствата и международните институции да планират инфраструктурни проекти и мерки за намаляване на негативните ефекти върху климата. Прогностичните модели, използвани за природни бедствия, каквито са наводненията и засушаванията (Brunner, Slater, Tallaksen, & Clark, 2021), позволяват на местните власти да се подготвят за адекватни превантивни действия. В икономиката и бизнеса, например, прогнозите са полезни за финансови анализи (Granger & Poon, 2001) и управление на риска (Brooks & Persand, 2002). В сферата на логистиката и управлението на вериги за доставки прогнозите подпомагат оптимизацията на запасите и планирането на ресурсите, предотвратявайки свръхпроизводството или дефицита на стоки (Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016). В енергийния сектор прогнозите на потреблението на енергия са важен инструмент за оптимизация на производството, особено в мрежи с възобновяема енергия, където сезонността играе ключова роля (Ding, Tao, Li, & Qin, 2022).

Ключовото значение на прогнозирането се състои в разкриване на перспективите в развитието на явленията и играе важна роля както за ефективното управление на икономическите процеси на макроравнище, така и за успешното развитие на всеки бизнес на микониво. Прогнозите компенсират в определена степен неопределеността и осигуряват обективна информация, необходима за изработването на адекватни и полезни решения за бъдещето, които подпомагат навременното и информирано формиране на политики.

За успешно и навременно прогнозиране е необходимо, в рамките на една организация да се развие система за неговото осъществяване. Такава система изисква изграждане на експертност при идентифициране на проблемите, прилагане на различни подходи, избиране на подходящите сред тях за всеки конкретен случай, оценка и усъвършенстване на подходите във времето. Също така е важно да има стабилна организационна подкрепа за използването на формални методи за прогнозиране, за да се гарантира успешното им прилагане (Fildes & Hastings, 1994).

Прогнозирането обикновено включва пет основни етапа. Началото се поставя с *дефинирането на проблема*, което може да се окаже най-трудната част от процеса. Изисква се разбиране на начина, по който ще се използват прогнозите, кой има нужда от тях и какви са функциите на прогнозирането в организацията. За тази цел е необходимо прогнозиращият да комуникира с всички, които участват в събирането на данни, поддържането на базите от данни и използването на прогнозите за бъдещо планиране. Следващата стъпка е свързана със *събирането на информация*. Трябва да се има предвид,

че са необходими поне два вида информация: статистически данни и натрупаният опит на хората, които събират данните и използват прогнозите. Понякога е трудно да се осигури достатъчно количество данни за минали периоди, с които да се построи добър статистически модел. В тези случаи могат да се използват качествени методи за прогнозиране. Понякога данните за миналото развитие на явлението могат да са по-малко полезни поради структурни промени в динамиката му, но добрите статистически модели биха отчели еволюционни промени в системата. За успешното прогнозиране е необходим *предварителен (изследователски) анализ*. Той винаги започва с визуализиране на данните, при което се търси отговор на различни въпроси: дали се открояват закономерности в динамиката, забелязва ли се значима тенденция, открива ли се сезонност и важна ли е тя, присъстват ли бизнес цикли, има ли големи отклонения от тенденцията, които трябва да бъдат обяснени от експерти. За този вид анализ са разработени различни инструменти. Следва етапът за *избор и настройка на модели*. Най-подходящият модел зависи от наличието на данни за минали периоди, силата на връзките между прогностичната променлива и обяснителните променливи и начина, по който ще се използват прогнозите. Обичайно е да се сравнят няколко потенциални модела, всеки от които е създаден въз основа на набор от предположения и обикновено включва един или повече параметри, които трябва да бъдат оценени. Финалната стъпка е *използване и оценка на прогностичния модел*. След като моделът е избран и параметрите му са оценени, той се използва за създаване на прогнози. Представянето на модела може да бъде оценено само след като се получат резултатите за прогнозния период. То се основава на използването на някой от множеството методи, разработени за оценка на точността на прогнозите.

1.2. Развитие на автоматизацията в прогнозирането

Автоматизацията в прогнозирането е съвкупност от процедури, в началото на която са данните, а накрая е прогнозата за развитието на дадено явление. При това случващото се между входа и изхода е без намесата на изследователя и се извършва от специално създадени системи или софтуер. Автоматизацията може да се разглежда като противодействие на традиционните прогнози, при които реализирането на всяка стъпка от прогнозния процес – от визуализирането на данните, през избора на модел, до получаването на прогнозата, се осъществява от изследователя и зависи от неговата компетентност и индивидуалните му решения.

Съвременната динамична среда изисква точна и бърза информация за вземане на решения. Това превръща автоматизацията на прогнозите в ключов инструмент за управление на данни и ресурси. Причините биха могли да се обособят в следните аспекти:

- увеличаващият се обем от данни, които се генерират ежедневно. Обработката на такива обемни данни ръчно е практически невъзможна и

неефективна. Според (Hyndman & Athanasopoulos, 2021) автоматизацията позволява бързо и ефективно управление и анализ на големи данни;

- намаляване на времето за анализ. Това е особено важно при бързопроменящи се пазари, където е необходима незабавна информация за планиране на ресурси и стратегически решения. Според (Reilly, 1980) автоматизацията на прогнозните процеси значително подобрява оперативната ефективност, като дава възможност да се генерират прогнози за минути, вместо за дни;
- устойчивост и намаляване на субективните грешки. Традиционните прогнози са силно зависими от знанията и уменията на изследователя, което може да доведе до различия между прогнозите, генерирани от различни експерти. Автоматизацията подобрява точността на прогнозирането чрез премахване на субективни грешки и осигуряване на надеждност в прогностичния процес;
- оптимизация на ресурсите. С автоматизирането на прогнозите намалява необходимостта от експертен персонал, който да се занимава с традиционно прогнозиране и настройка на параметри. Така прогнозните методи стават достъпни за неспециалисти, с което се улеснява използването на сложни модели без нуждата от задълбочени статистически познания. Освен това се редуцират и разходите за дълги и сложни анализи;
- позволява анализ при сложни динамични редове и наличие на сезонност. Автоматизираните системи предлагат вградени алгоритми за разпознаване и обработка на сложни зависимости в данните. Автоматизацията позволява повишаване точността на прогнозите и намаляване на вероятността от неправилна интерпретация на сезонни ефекти и тенденции;
- бързо адаптиране към променящи се условия. В условия на несигурност или непрекъснати промени на пазара е важно, моделите да могат да се приспособяват към актуалната информация. Чрез автоматизираните алгоритми прогнозите могат да се обновяват почти в реално време, а взетите решенията се основават на най-новата налична информация;
- надеждност и точност в дългосрочен план. Тези характеристики са критични при дългосрочни прогнози. Автоматизираните алгоритми са надеждно решение в този случай, тъй като чрез тях може да се извърши оптимизация на параметрите съобразно структурата на данните;
- подобрена ефективност при множество динамични редове. При работа с голям брой динамични редове автоматизацията позволява едновременното им прогнозиране, което би било почти невъзможно при традиционното прогнозиране.

Развитието на автоматизираните прогностични методи преминава през няколко етапа до превръщането на автоматизацията в съвременна мощна система за прогнозиране. *Началото* е преди 70-те години на XX век, преди широкото разпространение на компютърната автоматизация, когато

се използват традиционните прогностични методи. Разчита се на обикновени статистически методи и субективната експертиза на изследователите. Това ограничава възможностите за обработка на големи масиви от данни и използване на сложни модели, поради което се прилагат достъпни и с лесни изчислителни процедури техники. Сред тях могат да се посочат:

- Средни стойности и екстраполация, които са подходящи, когато липсват значителна вариация или сезонност в данните;
- Линейна регресия, подходяща за моделиране на данни с ясна линейна тенденция, където бъдещите стойности се прогнозират въз основа на връзката между времето като независима променлива и някаква зависима променлива. Линейната регресия не е ефективна при наличие на сезонност или нелинейна зависимост;
- „Плъзгащи се средни“ (moving average) е техника, която изглажда краткосрочните колебания в динамичния ред и позволява по-ясно определяне на дългосрочните тенденции. Подходяща е при данни с висока волатилност, но има ограничения при данни с ясно изразени цикли или сезонност.

Освен посочените вече особености още специфични характеристики на традиционните прогностични методи ограничават тяхното прилагане в дългосрочна перспектива. Например те са зависими от експертизата на изследователя, понеже разчитат на интуицията и опита му. Това води до субективност и възможност за грешки. Освен това те не успяват да отчетат цикли, сезонност или внезапни промени в динамичните редове, което води до прогнози с ниска точност, особено в ситуации с големи колебания и непредвидени събития. Не е за пренебрегване и обстоятелството, че при наличието на множество динамични редове или обширни масиви от данни са необходими значителни време и ресурси за обработка на данните и бързо генериране на прогнози в реално време.

Първи опити за систематизация на прогнозните методи за моделиране на динамични редове правят в края на 60-те и началото на 70-те години на XX век Box и Jenkins. Целта на техния подход, по-късно формализиран в ARIMA моделите, е да преодолее ограниченията на традиционните прогностични методи чрез структуриран и последователен метод за анализ. Методологията включва идентификация на модела, оценка на параметрите и диагностика, което улеснява прогнозите и прави възможно прилагането на по-сложни модели дори при липса на експертна намеса за всяка стъпка. Освен това се намалява субективността и изследователите се насочват към по-точни прогнози. Предложената методология показва, че традиционният прогностичен процес може да бъде структуриран и стандартизиран, което по-късно се адаптира в автоматизирани алгоритми (Box, Jenkins, Reinsel, & Ljung, 2015).

Преходът към автоматизация е свързан с появата на първите софтуерни инструменти. През 80-те и 90-те години на XX век компютърните технологии започват да навлизат в анализа и прогнозирането на динамични

редове. Стига се до създаване на автоматизирани софтуерни решения за моделиране и прогнозиране:

- SAS (Statistical Analysis System) е един от първите софтуери. SAS включва автоматичен избор на модели чрез информационни критерии като Akaike Information Criterion (AIC) и Bayesian Information Criterion (BIC), което значително улеснява избора на най-подходящ модел за прогнозиране;
- SPSS (Statistical Package for the Social Sciences) е разработен за статистически анализи в социалните науки и маркетинговите изследвания, но с времето се разширява и включва възможности за автоматизация на прогнозните процеси за динамични редове;
- AUTOBOX е една от първите специализирани системи за автоматизация на динамични редове, създадена на базата на методологията на Box и Jenkins. Използва се във финансовия и търговския сектор за прогнозиране на продажби и управление на вериги за доставки (Reilly, 1980);
- BMDP (Biomedical Data Package), първоначално разработен за биомедицински изследвания, включва и функции за анализ на динамични редове. Осигурява основни инструменти за моделиране на линейни и нелинейни зависимости;
- SCA (Scientific Computing Associates) е статистически софтуер, разработен с фокус върху моделирането на динамични редове и иконометрията. Предлага автоматизирано идентифициране на параметри и модели, както и разпознаване на аномалии като структурни промени в нивото и дисперсията на данните. Използва се за анализ във финансови и икономически изследвания;
- Time Series Processor (TSP) е софтуер за иконометричен анализ, който предлага автоматизация на избор на ARIMA модели и диагностика на данни. Включени са инструменти за тестове за стационарност и корелация, което го прави предпочитан инструмент за анализ на макроикономически показатели и финансови данни;
- RATS (Regression Analysis of Time Series) е софтуерен пакет за регресионен анализ и динамични редове, който също включва автоматизация на параметрите и моделите. Полезен е за иконометрични изследвания, включително анализ на международни пазари и макроикономически прогнози;
- Forecast PRO е софтуер, специално разработен за бизнес прогнози и управление на активи. Той предоставя автоматичен избор на модели, като използва алгоритми за най-добро съвпадение на данните и се използва широко в управлението на запаси и планирането на продажби;
- EViews (Econometric Views) е иконометричен софтуер, който предоставя автоматизирани функции за моделиране на ARIMA модели и разши-

рени възможности за анализ на динамични редове. Той е предпочитан инструмент в икономически и финансови прогнози, особено в макроикономическите изследвания;

- Stata е статистически и иконометричен софтуер, който включва автоматизация на модели и параметри особено в анализите на панелни данни и динамични редове. Позволява на потребителите да създават скриптове за автоматизация, което улеснява повторяемостта на анализа и прави Stata популярен избор за академични изследвания и иконометричен анализ.

Могат да се открият някои недостатъци и предизвикателства при използването на ранната автоматизация. Софтуерните системи често имат ограничени възможности – основно предлагат автоматизация за базови модели (например, линейна регресия и ARIMA) и не са способни да обработват сложни нелинейни структури или множество сезонни компоненти, което ги прави по-малко подходящи за сложни задачи. Те са с ограничена гъвкавост, зависими са от предварително зададени параметри. Макар и автоматизирани, софтуерните системи изискват известна ръчна настройка и контрол от потребителя. Техните софтуерни и хардуерни ограничения, далече от днешните стандарти, ограничават скоростта и сложността на автоматизираните изчисления, по-бавни са и не позволяват обработка на големи масиви от данни в реално време.

Софтуерните решения от края на XX век полагат основите на автоматизацията в прогнозирането. Всяко от тях има специфични функции, насочени към определени приложения – от иконометрия и финанси до социални науки и бизнес планиране. Те предоставят първите възможности за автоматично задаване на параметри, избор на модели и разпознаване на аномалии в данните, улеснявайки и ускорявайки анализа на динамични редове. Затова имат значима роля за развитието на автоматизацията на прогнозите. С увеличаването на достъпа до компютърни ресурси и по-ниските разходи за софтуер автоматизираните прогнозни решения стават по-достъпни за бизнеса, което ускорява навлизането на тези технологии и изгражда основа за тяхното масово използване в следващите години.

Началото на XX век бележи *развитие на автоматизацията* с прехода от мощни алгоритми към платформи с машинно обучение и прогнози в реално време. Този значителен прогрес продължава и до днес. Благодарение на развитието на изчислителните технологии и въвеждането на нови алгоритми се създават по-мощни и достъпни инструменти за прогнозиране. Те предлагат на потребителите лесен достъп до автоматизирано моделиране на динамични редове и предоставят по-точни прогнози, съобразени с конкретните особености на данните.

През 2008 г. се разработва forecast пакета за езика за програмиране R (Hyndman, R., et al., 2024), който значително подобрява автоматизацията на ARIMA моделирането. Вградената функция auto.arima използва критерии AIC и BIC за автоматичен избор на параметри и настройка на модела, като

освобождава потребителите от необходимостта да извършват ръчна настройка на параметрите (Hyndman & Khandakar, 2008). Този пакет позволява включване на сезонни компоненти и SARIMA модели, което го прави подходящ за анализ на данни с изразена сезонност.

Разработването на алгоритми като STL (Seasonal-Trend decomposition using Loess) и TBATS (Trigonometric, Box-Cox transform, ARMA errors, Trend and Seasonal components) дава възможност за обработка на сложни динамични редове с множество сезонни и нелинейни компоненти (De Livera, Hyndman, & Snyder, 2011). TBATS моделът е особено полезен за данни със сложна сезонност, като месечни или тримесечни цикли, и предлага автоматично настройване на параметрите, което значително увеличава точността на прогнозите.

През последното десетилетие автоматизацията се обогатява с иновации в областта на машинното обучение и изкуствения интелект, които позволяват използването на модели с по-висока адаптивност и прогнози в реално време. Важна платформа за автоматизирано прогнозиране става Python. Пакетът statsmodels (Seabold & Perktold, 2010) предлага функции за ARIMA модели и други иконометрични методи, които улесняват анализа на динамични редове. Друг иновативен инструмент е Prophet – софтуерен пакет, разработен от Facebook. Той е проектиран да бъде лесен за употреба и да предоставя точни прогнози дори когато има липсващи стойности и непостоянна сезонност в данните (Taylor & Letham, 2018). Prophet автоматично разпознава сезонни модели и позволява на потребителите да включват специфични събития, което го прави популярен сред компании, които се нуждаят от бързи и адаптивни прогнози.

С навлизането на машинното обучение и невронните мрежи автоматизацията на прогнозите достига ново ниво. Модели като RNN (Recurrent Neural Network) и LSTM (Long Short-Term Memory) предлагат възможност за откриване на сложни зависимости и нелинейност в динамичните редове. Тези модели могат да се адаптират към динамични промени в данните и са подходящи за прогнозиране в реално време, особено при висока честота на данните (например минутни или часови измервания). Много от съвременните платформи за машинно обучение, като TensorFlow и PyTorch, включват възможности за автоматично обучение и оптимизация, което прави машинното обучение по-достъпно за прогнозиране на динамични редове (Chollet, 2017).

Развитието на облачните технологии е свързано с предлагането на услуги за прогнозиране в реално време, базирани на машинно обучение, от компании като Amazon и Google. AWS Forecast и Google Cloud AI позволяват на потребителите да качват своите данни и да използват вградени алгоритми за автоматично прогнозиране. Чрез тях могат да се обработват големи обеми от данни, което ги прави подходящи за прогнозиране при множество динамични редове в реално време.

Очаква се в бъдеще, автоматизацията на прогнозите да продължи да се усъвършенства чрез включване на нови подходи за обработка на големи данни и реализация на прогнози в реално време чрез облачни платформи и изкуствен интелект. Тези иновации ще продължат да намаляват времето и усилията, необходими за анализ на данни, и ще направят прогнозите по-достъпни за широк кръг от потребители, независимо от тяхната статистическа подготовка.

2. ARIMA модели и тяхната автоматизация

2.1. Същност на ARIMA моделите

Един от основните подходи за прогнозиране на динамични редове се базира на разработените от Box и Jenkins (Box, Jenkins, Reinsel, & Ljung, 2015) клас линейни стохастични модели – $ARIMA$. Те осигуряват систематизиран подход за анализ на динамични редове въз основа на взаимната обвързаност между членовете на реда. Основната идея на $ARIMA$ моделите е да се разкрият вътрешните закономерности в миналите значения на реда и на базата на тях да се прогнозират бъдещите стойности. Този клас модели се основава на предположението, че динамичният ред представлява реализация на стохастичен процес – последователност от случайни величини, всяка от които притежава свое вероятностно разпределение.

Основно изискване за прилагането на $ARIMA$ моделите е, динамичните редове да бъдат стационарни[2]. Дефиницията за стационарност в широк смисъл гласи, че динамичният ред се характеризира чрез постоянни, независещи от времето, математическо очакване, дисперсия и автоковариация (Иванов, Касабова, & Шопова, 2017). На практика поради инерционния характер на социално-икономическите явления (Димитров, 1980) много от динамичните редове са нестационарни. Това налага да се извърши трансформация, с която нестационарният ред да се сведе до стационарен. В контекста на моделирането с $ARIMA$ тази трансформация е филтриране с последователни разлики. Стационарните редове са интегрирани от нулев порядък – $I(0)$. Динамичен ред, интегриран от порядък d и означаван с $I(d)$, е такъв ред, който се свежда до стационарен чрез филтриране с последователни разлики d пъти.

Общата формулировка на $ARIMA(p, d, q)$ модел е следната:

$$\boldsymbol{\phi(L)(1-L)^d y_t = c + \theta(L)\varepsilon_t}, \qquad \textbf{(1)}$$

[2] Следва да се отбележи, че ARIMA моделите могат да се прилагат и за нестационарни редове, но след трансформация в стационарни чрез прилагане на последователни разлики. Това е изяснено по-нататък в изложението.

където:

y_t е наблюдаваният динамичен ред в момент t;

ε_t представлява „бял шум“ със средна стойност 0 и дисперсия σ_a^2;

$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ е полином на авторегресията (AR) от порядък p, който описва зависимостта между текущите и предишните стойности;

$\theta(L) = 1 - \theta_1 L - \cdots - \theta_p L^q$ е полином на плъзгащи се средни (MA) от порядък q, който описва зависимостта между текущата стойност и предишните случайни отклонения;

L е лагов оператор, при който $Ly_t = y_{t-1}$;

d е порядъкът на последователните разлики, прилагани за постигане на стационарност;

$(1 - L)^d$ означава, че са приложени d пъти последователни разлики, за да се сведе редът до стационарен.

$ARIMA(p, d, q)$ моделите са изградени от три основни компонента: авторегресионните модели (AR), порядък на интегрираност (I) и модели на плъзгащи се средни (MA). При авторегресионните модели текущите значения на динамичния ред се представят като линейна комбинация от краен брой минали значения на реда. Параметърът p показва колко предишни стойности са включени в моделирането. При моделите с плъзгащи се средни текущите значения се изразяват като линейна комбинация от краен брой предшестващи значения на случайните отклонения. Параметърът q показва колко предходни случайни отклонения се включват в модела. Порядъкът на интегрираност (I) представлява броя на последователните разлики, необходими за постигане на стационарност на реда. Този брой се изразява с параметъра d.

Често срещана особеност на динамичните редове, характеризиращи социално-икономически явления, е сезонността. Сезонният компонент се свързва с периодично повтарящи се колебания през определени интервали от време, като месеци, тримесечия или години, с относително постоянна амплитуда. Те често отразяват регулярни промени в условията, при които протичат събитията – климатични условия, празници или специфики на производствената дейност. Със сезонност се характеризират потреблението на топлинна енергия от домакинствата, пътуванията с цел туризъм и търсенето на туристически услуги, заетостта в земеделието, продажбите на дребно на сезонни плодове и зеленчуци и др.

Отчитането на сезонността може значително да подобри точността на прогнозите. Но тя създава предизвикателства за стандартните $ARIMA$ модели, които не са в състояние да уловят периодичните колебания във времето. Когато сезонните закономерности са ясно изразени, прогнозирането без тяхното отчитане може да доведе до значителни грешки и неточности.

За справяне с този проблем е необходимо разширение на $ARIMA$ модела, което позволява включването на сезонни компоненти и така осигурява по-реалистично моделиране на данните. Това разширение, известно като сезонен $ARIMA$ ($SARIMA$) модел, добавя допълнителни параметри, които са специално предназначени за улавяне на сезонността в динамичните редове.

Общата формулировка на $SARIMA(p, d, q, P, D, Q)$ модел е следната:

$$\boldsymbol{\Phi(L^s)\phi(L)(1-L)^d(1-L^s)^D y_t = \Theta(L^s)\theta(L)\varepsilon_t}, \qquad \textbf{(2)}$$

където:

y_t е наблюдаваният динамичен ред в момент t;

L е лаговият оператор, като $L^k y_t = y_{t-k}$;

s е сезонната честота (например, 12 за месечни данни или 4 за тримесечни данни);

$\phi(L)$ и $\theta(L)$ са полиномите на авторегресия и плъзгащи се средни за несезонните компоненти;

$\Phi(L^s)$ и $\Theta(L^s)$ са сезонните авторегресионен полином и полином на плъзгащи се средни;

d е порядъкът на последователните разлики, необходими за постигане на стационарност;

D е порядъкът на последователните разлики на сезонните честоти за премахване на сезонната интегрираност.

$SARIMA(p, d, q, P, D, Q)$ моделът разширява класическия $ARIMA(p, d, q)$ модел, за да отчете сезонните зависимости в динамичните редове, като добавя сезонна авторегресия, сезонна плъзгаща се средна и сезонна интегрираност:

Сезонната авторегресия SAR улавя връзките между текущата стойност на динамичния ред и стойностите от същия сезон на предходни периоди. Този компонент се описва чрез сезонен авторегресионен полином $\Phi(L^s)$ с порядък P. Например при $P = 1$ и сезонност от 12 месеца текущата стойност y_t ще зависи от стойността преди 12 месеца y_{t-12}, което е типично за годишни сезонни цикли.

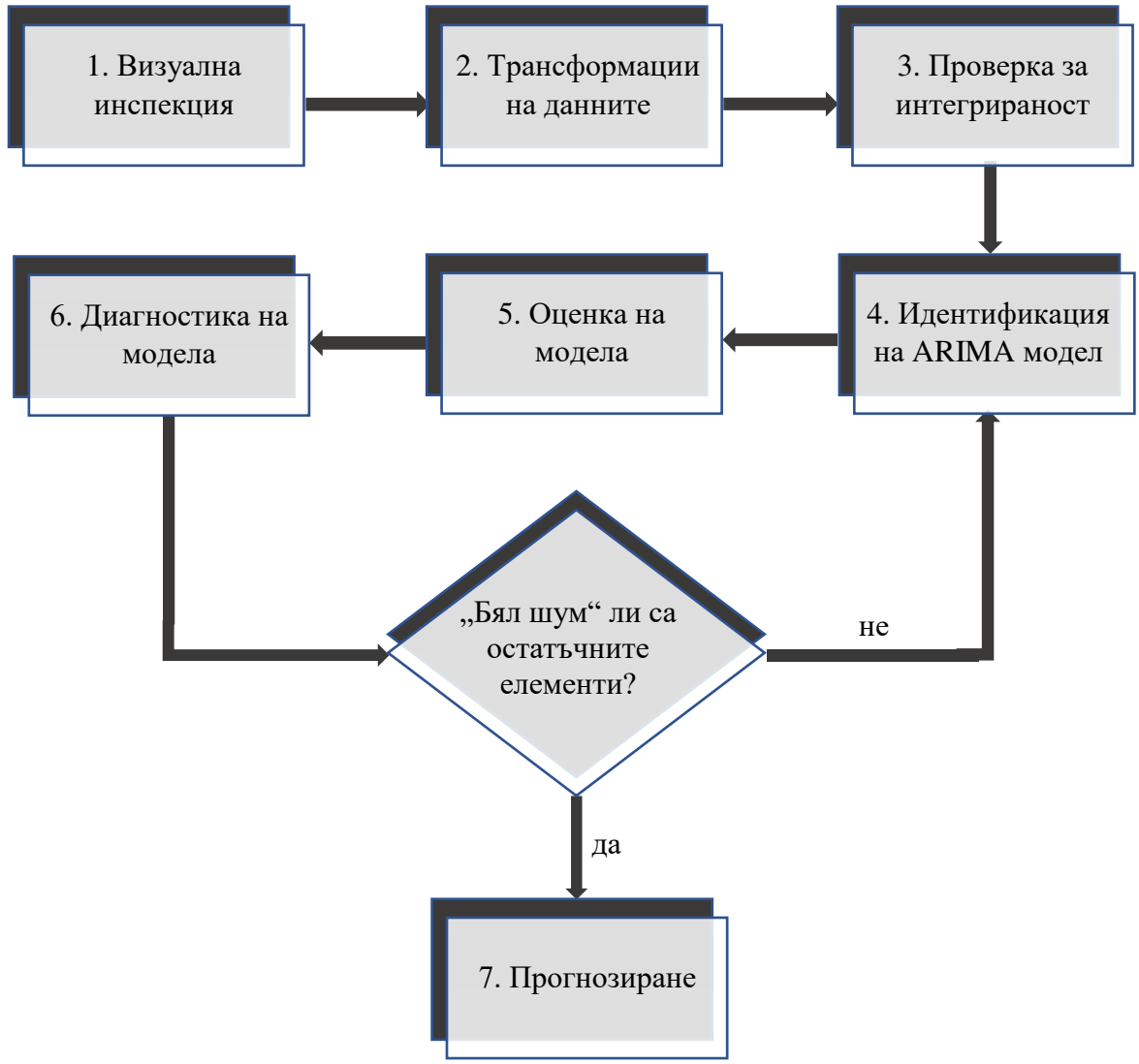
Сезонната плъзгаща се средна SMA отчита влиянието на сезонните случайни отклонения от предходни сезони върху текущата стойност. Този компонент се представя чрез полином на сезонната подвижна средна $\Theta(L^s)$ с порядък Q. Например при $Q = 1$ и сезонност от 12 месеца текущата стойност y_t се влияе от случайното отклонение, възникнало преди 12 месеца ε_{t-12}, което е характерно за модели с годишна сезонна структура.

Сезонната интегрираност SI се включва, за да се постигне стационарност при наличието на сезонност. Тя се означава с параметъра D. Сезонното интегриране е процес, при който се премахват повторяеми сезонни ефекти чрез изваждане на стойности, които са отдалечени на разстояние, кратно на

сезонната честота s. Например при сезонност $s = 12$ и $D = 1$, от всяка стойност y_t ще бъде извадена стойността y_{t-12}, което премахва годишните цикли и прави реда стационарен по отношение на сезонните колебания.

2.2. Процес на ARIMA моделиране

Процесът на прогнозиране с ARIMA модели включва последователност от етапи, всеки от който е от значение за осигуряването на точни и надеждни прогнози (фиг.1). Първите три етапа могат да се разглеждат като предварителни, които да улеснят същинското моделиране през следващите три етапа.



Фигура 1. Етапи на прогнозиране с ARIMA модели
Източник: авторите, по идея от (Hyndman & Athanasopoulos, 2021)

Визуалната инспекция е основен етап в анализа на динамичните редове, който позволява на анализатора да идентифицира ключови характеристики като трендове, сезонност и екстремални стойности. Тя предоставя

интуитивно разбиране на динамиката на реда, което е от съществено значение за правилното моделиране и за избора на подходящи трансформации.

По отношение на порядъка на интегрираност могат да се използват линейни диаграми на изходния ред и на реда, съставен от първите последователни разлики. Индикативни за порядъка на интегрираност могат да бъдат и корелограмите, построени на основата на автокорелационните коефициенти. Теоретично, при напълно случаен ред (бял шум) автокорелационните коефициенти би трябвало да имат близки до нула стойности. При стационарен ред автокорелационните коефициенти следва да затихват до нула с нарастването на лаговете. Ако автокорелационните коефициенти остават с високи стойности за много на брой лагове, т.е. затихват плавно и бавно, то тогава редът е интегриран.

Сезонността в динамичните редове може да се установи визуално чрез линейните диаграми на изходния ред и на реда, получен след филтриране с последователни разлики за сезонните честоти, а също и чрез сезонни диаграми. Последните дават възможност да се проследи изменението на динамичния ред през различни периоди, определящи сезонността – например месеци, тримесечия и други.

Полезен инструмент за идентифициране на сезонност в динамичните редове е визуалният анализ на автокорелационните функции на изходния ред, реда на първите последователни разлики, реда на първите разлики на сезонните честоти и реда, получен от едновременното прилагане на първи разлики и разлики на сезонните честоти. При наличието на сезонност в корелограмата на изходния ред автокорелационните коефициенти, при лагове, съответстващи на сезонния период, биха имали високи стойности. Например при тримесечни данни, ако редът има годишна сезонност, ще се наблюдават пикове в автокорелационната функция при лагове, кратни на 4 (4, 8, 12 и т.н.). Разбира се, ако редът е интегриран, те биха се забелязали поотчетливо на корелограмата на реда на първите разлики.

Наличието на екстремални значения в наблюдавания ред е съществено за ARIMA моделирането. На практика често тяхното визуализиране е затруднено, заради другите характеристики на реда. Отдалечените наблюдения визуално могат да се открояят при инспекцията на остатъчните елементи от модела или след прилагане на STL декомпозиция.

Въпреки значението си, визуалната инспекция може да бъде времеемка и трудоемка, особено когато обемът от данни е голям или когато има множество динамични редове за анализ. Освен това автоматизацията на визуалния анализ е трудна задача, тъй като процесът разчита в голяма степен на човешката интуиция и способността за визуално разпознаване на закономерности, които трудно могат да бъдат формализирани в алгоритми.

Автоматичните методи често изпитват затруднения при правилното идентифициране на сезонни модели или при разпознаването на едва доловими трендове и цикли, които могат да бъдат очевидни за човешкото око.

Освен това екстремалните стойности или структурните промени понякога се проявяват в нетипични форми, които алгоритмите трудно могат да уловят без предварително зададени специфични правила. Тези предизвикателства правят автоматизирането на визуалната инспекция частично възможно, но с ограничена ефективност, подчертавайки важността на комбинирания подход между автоматични и човешки методи за осигуряване на висока точност и надеждност в анализа на динамични редове.

Целта на ***трансформацията на данните*** е да се стабилизира вариацията в изследваните редове. Ако данните показват вариации, които се увеличават или намаляват в зависимост от нивото на реда, тогава трансформацията може да бъде полезна. Например логаритмичната трансформация често е подходяща. Ако означим оригиналните наблюдения с $y_1, y_2,\ldots, y_T$ и трансформираните наблюдения с $w_1, w_2,\ldots, w_T$, тогава $w_t = log_a(y_t)$. Логаритмите са полезни, защото са лесни за интерпретация: промените в логаритмичните стойности са относителни (или процентни) промени в оригиналната скала. Например, ако се използва десетичен логаритъм (т.е. с основа $a = 10$), увеличение с 1 в логаритмичната скала съответства на умножение по 10 в оригиналната скала. Логаритмична трансформация не е възможна, ако някоя стойност в оригиналния ред е нула или отрицателна.

Понякога се използват и други трансформации (макар че те не са толкова лесни за интерпретация). Например могат да се приложат квадратни или кубични корени. Тези трансформации се наричат степенни трансформации, тъй като могат да бъдат записани във формата $w_t = y_t^p$.

Полезна фамилия трансформации, която включва както логаритмични, така и степенни трансформации, е тази на Box-Cox (1964), които зависят от параметъра λ и са дефинирани по следния начин:

$$\boldsymbol{w_t} = \begin{cases} \boldsymbol{log_a(y_t)} & \text{, ако } \boldsymbol{\lambda = 0} \\ \boldsymbol{\frac{sign(y_t)|y_t|^{\lambda}-1}{\lambda}} & \text{, ако } \boldsymbol{\lambda \neq 0} \end{cases} \quad \textbf{(3)}$$

Това всъщност е модифицирана трансформация на Box-Cox, описана от Bickel & Doksum (1981), която позволява отрицателни стойности на y_t, ако $\lambda > 0$.

Логаритъмът в трансформацията на Box-Cox винаги е естествен логаритъм (т.е. с основа $a = e$). Затова, ако $\lambda = 0$, се използва естествен логаритъм $ln(y_t)$. При $\lambda \neq 0$, се прилага степенна трансформация.

Ако $\lambda = 1$, то $w_t = y_t - 1$. Така трансформираните данни се изместват надолу, но няма промяна в динамиката на реда. За всички други стойности на λ такава промяна ще настъпи.

Проверка за интегрираност. Проверката за порядъка на интегрираност се осъществява чрез тестове за единичен корен. Те заемат ключова роля в процеса на $ARIMA(p, d, q)$ моделирането, по-конкретно – при установяване стойността на параметъра d. Чрез тях се определя дали даден ред е

стационарен, или съдържа единичен корен, което предполага нестационарност. В практиката се среща разнообразие от тестове за единичен корен, които могат да се класифицират по различни критерии – например, спрямо нулевата хипотеза, сезонността на данните и подхода към проверката. Тестовете могат да бъдат разделени основно на такива, които приемат за нулева хипотеза наличието на единичен корен, например тестът на Dickey-Fuller (1979), (1981) и неговите разширени версии като ADF (Said & Dickey, 1984) и такива, при които нулевата хипотеза е стационарност – като теста на Kwiatkowski, Phillips, Schmidt, & Shin (1992) – KPSS.

Наличието на сезонност в динамичните редове е предпоставка да е затруднено прилагането на горните тестове. За отчитане на сезонните компоненти са подходящи например тестовете на Hylleberg, Engle, Granger, & Yoo (1990) и на Canova & Hansen (1995). Друг проблем при прилагането на тестовете за единичен корен е чувствителността им към структурни разриви. В този случай може да се използва тестът на Zivot & Andrews (2002).

Автоматизацията на процеса на тестове за единичен корен е възможна до известна степен чрез алгоритми, които автоматично прилагат различни тестове и избират най-подходящия според характеристиките на данните. Въпреки това автоматизирането на процеса не премахва напълно нуждата от експертна оценка, тъй като някои аспекти изискват допълнителен анализ и могат да бъдат трудни за улавяне от автоматичните алгоритми. Това налага, изследователите да използват комбиниран подход между автоматизация и допълнителна преценка.

Същинското ARIMA моделиране на динамични редове е итеративен процес, включващ етапите идентификация, оценка на параметрите и диагностика на модела. Този процес е подробно разгледан в класическата статистическа литература (Box, Jenkins, Reinsel, & Ljung, 2015), (Tiao, 2000).

Основната задача през етапа на ***идентификация на модела*** е да се изберат подходящите стойности за параметрите на авторегресията $(AR) – p$ и на сезонната авторегресия $(SAR) – P$, на плъзгащата се средна $(MA) – q$ и на сезонната плъзгаща се средна $(SMA) – Q$. Изследователският инструментариум на този етап се състои от статистически тестове за определяне на порядъка на интегрираност и автокорелационен анализ (Wilson, 2000).

Идентификацията на $ARIMA$ моделите, според класическата методология на Box и Jenkins, се основава на теоретичните свойства на автокорелационната и частната автокорелационна функция. За авторегресионен модел от порядък $AR(p)$ частната автокорелационна функция прекъсва след p-тия лаг, докато автокорелационната функция постепенно намалява. Обратно, при модел $MA(q)$ на плъзгаща се средна от ред q автокорелационната функция прекъсва след q-тия лаг, докато частната автокорелационна функция постепенно намалява. Тези теоретични свойства се използват като основа за идентифициране на реда на AR и MA компонентите в даден динамичен ред.

Методът на Box и Jenkins за идентификация на моделите разчита в голяма степен на анализа на линейните диаграми на динамичните редове и на корелограмите на автокорелационната и частната автокорелационна функции. Тези инструменти, и съответно самата методология, могат да бъдат ефективни при идентификация на чисти авторегресионни или модели с плъзгаща се средна, но не толкова ефективни при модели, които съдържат едновременно двата компонента.

След идентификацията на модела се преминава към следващия етап – ***оценка*** на неговите параметри. Използват се методите на максималното правдоподобие и на най-малките квадрати. През този етап се намират стойностите на параметрите на модела, с които оптимално се описват закономерностите в динамиката на реда. Въз основа на оценените модели се изчисляват и стойностите на информационните критерии за избор на най-подходящ модел.

Етапът на ***диагностика на модела*** е от съществено значение за проверка на адекватността на модела и се основава на анализ на остатъчните му елементи. Те трябва да бъдат „бял шум“, т.е. да са случайни, некорелирани и хомоскедастични. По този начин се гарантира, че моделът е обхванал вътрешните закономерности в изследвания ред и е подходящ да се използва за прогнозиране. На този етап се използват тестовете на Ljung & Box (1978), Breusch & Pagan (1979) и др. Ако остатъците не отговарят на изискванията за бял шум, е необходимо да се преразгледат параметрите или да се избере друг модел.

3. Автоматизиране на идентификацията на ARIMA моделите

Процедурата за моделиране с ARIMA, разгледана в параграф 2.2, е про-цес, който е предизвикателство за пълно автоматизиране. В специализираната литература идентификацията на модела често е възприемана като „изкуство“ и е силно обвързана с експертизата на конкретния изследовател.

Съществуват много причини, поради които е необходимо колкото е възможно по-голямо автоматизиране на етапа на идентификация на ARIMA модела. Те могат да се сведат до две. Първо, трябва да се елиминират колко-то е възможно повече рутинните и механичните задачи, които би могло да се изпълнят от компютър. По този начин се повишава производи-телността на анализатора. Втората причина е свързана с обективността на етапа на идентификация. Желателно е, той да не се подлага на евристични методи и ad hoc процедури, които варират при различните експерти по динамични редове.

Проблематиката, свързана със субективността на класическата методология за идентификация, е задълбочено изследвана в научната литература от края на XX век и началото на XXI век. В резултат са разработени редица

подходи, които обективизират и автоматизират този процес. Подробна класификация може да бъде открита например в (Gómez & Maravall, 2000), (Choi, 2012), (de Gooijer, Abraham, Gould, & Robinson, 1985).

В рамките на настоящото изследване обособяваме две групи подходи за автоматизация на идентификацията на $ARIMA$ модели, които се различават както хронологично, така и според техния аналитичен подход и степента на сложност. Чрез подобна класификация обхващаме както по-ранните подходи към автоматизацията, така и по-новите интегрирани алгоритми. Исторически погледнато, първите автоматизирани подходи възникват с въвеждането на информационните критерии и еталонните (pattern) методи. Информационните критерии предлагат сравнително лесен начин за автоматичен избор на параметри на модела чрез минимализиране логаритъма на максималната стойност на функцията на правдоподобието. Еталонните методи се базират на съответствието между теоретичните автокорелационни структури за спецификациите на $ARIMA$ моделите (които служат за еталон) и емпирично изчислените такива от конкретни динамични редове. При тези методи се използват таблични средства за онагледяване, което ги прави популярни в първоначалните етапи на автоматизация.

С напредването на технологиите и нуждата от по-точни и гъвкави модели за прогнозиране се развиват комплексни алгоритми, които комбинират информационни критерии с евристични правила и автоматично откриване на сезонност. Този клас алгоритми надграждат класическите подходи, като подобряват автоматизацията и позволяват надеждна идентификация дори при динамични редове с по-сложни вътрешни закономерности.

3.1. Информационни критерии

Идентификацията при този клас методи, в основата на които са информационни критерии, се постига чрез най-ниската стойност на оценката на дисперсията ($\hat{\sigma}^2_{p,q}$) на остатъчните елементи от съответната спецификация на $ARIMA$ модела. Тъй като с включването на допълнителни компоненти на авторегресия (p) и плъзгащи се средни (q) оценката на дисперсията намалява, не е удачно да се избере модел, за която тя е минимална. По тази причина се прибавя компонент за корекция, отчитащ броя на наблюденията в реда и броя на включените компоненти на авторегресия (p) и плъзгащи се средни (q). Общата форма на функцията, която трябва да се минимализира е:

$$\boldsymbol{P(p,q) = ln\hat{\sigma}^2_{p,q} + (p+q)\frac{C(n)}{n},} \qquad \textbf{(4)}$$

$C(n)$ в (4) може да се замени с друга величина, например:

• ако $C(n) = 2$, се получава информационният критерий AIC на Akaike (1974);

• бейсовият информационен критерий BIC е резултатът при $C(n) = ln(n)$ (Schwarz, 1978);

• замяната $C(n) = 2ln\big(ln(n)\big)$ дава критерия HQ на Hannan и Quinn (1979).

По отношение на стойността на компонента за корекция критериите могат да се ранжират по следния начин. Най-ниска стойност се получава за AIC, а най-висока за BIC. Поради това може да се очаква, че спецификацията на модел, избрана с AIC, ще бъде с повече параметри p и q от тази с BIC. Това от своя страна е предпоставка, избраният модел да обяснява прекалено добре закономерностите в динамиката на реда (overfitting), но да не е подходящ за последващо прогнозиране.

Бейсовият информационен критерий (BIC) налага по-строга санкция за броя на параметрите, отколкото AIC, като по този начин често се избират по-пестеливи модели. Това прави критерия особено полезен при определяне на коректния порядък на модела при дълги динамични редове.

Критерият на Hannan-Quinn (HQ) предоставя компонент за корекция между този на AIC и BIC, балансирайки между склонността на AIC към „свръх напасване“ и консервативния подход на BIC.

Трябва да се има предвид, че при използването на информационни критерии за сравняване на спецификации на $ARIMA$ модели е необходимо да се съблюдава следното правило: сравними са тези модели, при които върху динамичните редове са приложени едни и същи трансформации. Например $ARIMA(1,1,1)$ не може да се сравнява с $ARIMA(1,0,1)$, тъй като в първия случай динамичният ред се трансформира с първи последователни разлики, а във втория случай се използва изходният ред.

Друга особеност на информационните критерии е, че чрез BIC се получават устойчиви оценки за порядъка на модела за разлика от AIC. Това не е причина да се предпочита BIC пред AIC, защото устойчивостта се основава на предположението, че съществува „истински“ $ARIMA$ модел за реда, а това е съмнително твърдение.

В обобщение, моделите трябва да се разглеждат като изкуствени конструкции, които са полезни за определени цели, но представляват само грубо приближение на реалността. В този смисъл критериите, използвани за избор на модели, особено при конкурентни спецификации, зависят от целта на изследването. Например някои критерии могат да бъдат подходящи за прогнозиране, но неприложими за установяване на теоретични зависимости. Трябва да се има предвид, че обикновено липсва еднозначно решение на проблема с идентификацията чрез информационни критерии.

3.2. Еталонни (pattern) методи

В основата на този клас подходи за идентификация лежи връзката между частните автокорелационни коефициенти и автокорелационните коефициенти, изразени чрез системата от уравнения на Yule-Walker (Chan,

1999). Те се основават на определени теоретични функции (най-често изведени от автокорелационните коефициенти). Така се създава двумерен масив за всеки $ARMA(p,q)$ модел, който да служи за еталон. Използвайки емпиричния аналог за динамичния ред на този двумерен масив, $ARMA$ моделът се идентифицира чрез съпоставяне с теоретичния еталон, който най-много наподобява емпиричния. Сред многото методи за идентификация по еталон, предложени в литературата, могат да се посочат методът на R и S масивите на Gray et al. (1978), „ъгловият" метод (Corner method) на Beguin et al. (1980), методът на разширена автокорелация (EACF) на Tsay и Tiao (1984), методът на най-малка канонична корелация (SCAN) на Tsay и Tiao (1985). Последните два метода са ефективни при несезонни динамични редове и могат да се използват с нестационарни редове. Същевременно методът на R и S масивите и ъгловият метод, които могат да се прилагат само при стационарни редове, не дават задоволителни резултати дори за редове без сезонност.

Woodward & Gray (1981) въвеждат нова концепция, наречена генерализирана частна автокорелационна функция (GPAC), която разширява традиционната частна автокорелационна функция. Тази генерализирана функция предоставя повече информация за порядъка на $ARMA$ моделите, дори когато са налице смесени компоненти (както авторегресионни, така и плъзгащи се средни). GPAC служи като връзка между метода на S масива и метода на Box и Jenkins, като подчертава ситуации, в които методът на S масива може да подобри подхода на Box и Jenkins, особено при работа със сложни $ARMA$ модели с ненулеви стойности на p и q.

Еталонните методи за идентификация получават смесени отзиви от анализаторите на динамични редове. Някои, като Liu и Hanssens (1982), Lii (1985) и Rezayat & Anandalingam (1988), подкрепят ъгловия метод. Други, като de Gooijer & Heuts (1981) и Davies & Petruccelli (1984), са по-малко оптимистични. Последните представят аргумент срещу използването на метода GPAC. Gooijer и Heuts (1981) отбелязват, че сложността и теоретичните трудности на метода на R и S масивите са ги възпрели от неговото използване.

Tsay и Tiao (1984) предлагат унифициран подход за определяне на реда на смесени стационарни и нестационарни $ARMA$ модели. Авторите се фокусират върху подобряване на устойчивостта в оценката на параметрите при моделиране с авторегресия и плъзгащи се средни. Те представят итеративен процес на регресия за получаване на устойчиви оценки на авторегресионните параметри. Така се преодолява проблемът, при който оценките на параметрите, получени по метода на най-малките квадрати, не винаги са устойчиви особено при смесени $ARMA$ модели.

Основната иновация е разширената автокорелационна функция (EACF). Тя служи като инструмент за идентификация на модела, като разглежда поведението на автокорелацията при различни лагове. EACF има

свойството на „отсичане“, което позволява на изследователите да идентифицират редовете на модела за $ARMA$ процеси без нужда от предварително определяне на порядъка на интегрираност. Този подход улеснява процеса на идентификация на смесени модели на авторегресия и плъзгащи се средни.

Идентифицирането на $ARMA$ моделите става чрез изследване на двумерна матрица от автокорелационни стойности, което помага за определянето на редовете на авторегресионния (AR) и на плъзгащата се средна (MA) компоненти. EACF е разработена, за да преодолее някои от ограниченията на традиционния анализ с автокорелационна функция (ACF) и частична автокорелационна функция (PACF) особено при смесени $ARMA$ модели.

Методът EACF създава таблица, в която всеки елемент съответства на конкретна комбинация от редове на AR и MA. Таблицата съдържа двупосочен масив от автокорелационни стойности, представляващ различни възможни комбинации от параметри на AR и MA. За всяка комбинация се изчисляват автокорелационните стойности за различни лагове и се търси специфичен модел на нулеви стойности, който подсказва подходящия $ARMA(p, q)$ модел.

Този метод може да се прилага както за сезонни, така и за несезонни модели, представяйки универсален подход за идентификация на модели на динамични редове. Като цяло Tsay и Tiao представят систематична процедура за идентификация на $ARMA$ модели, която намалява субективните решения и увеличава ефективността при специфицирането на модела.

3.3. Комплексни алгоритми

Hannan & Rissanen (1982) предлагат метод за определяне на порядъка на моделите на авторегресия и плъзгащи се средни за стационарен ред. В техния метод остатъчните елементи могат да бъдат получени чрез конструиране на авторегресионен модел от висок порядък за изходния ред. След това функцията на максималното правдоподобие на потенциалните модели се изчислява чрез последователност от линейни регресии.

Алгоритъмът включва три етапа:

- Най-напред се построява модел $ARMA(3,0)$ за несезонната част. След това се изчислява информационният критерий BIC за всички модели, за които сезонната част удовлетворява условията $0 \leq P, Q \leq m_s$. Числото m_s се избира от потребителя, като стойността по подразбиране е 1. Избира се моделът, за който стойността на BIC е най-малка;
- Фиксира се моделът за сезонните компоненти, избран на предходния етап. Въз основа на минималната стойност на BIC критерия за несезонните модели, които удовлетворяват условията $0 \leq p, q \leq m_r$, се избират порядъка на авторегресия p и на плъзгащи се средни q. Числото m_r може да варира в зависимост от решението на потребителя, като стойността по подразбиране е 3;

• Фиксира се несезонният модел, избран на предходния етап. Въз основа на BIC се избира модел за сезонната част при същите условия, както при първия етап – $0 \leq P, Q \leq m_s$.

Gómez (1998) разширява метода за идентификация на Hannan & Rissanen, като включва идентификация на мултипликативен сезонен ARIMA модел. Gómez & Maravall (1998) прилагат тази автоматична процедура за идентификация в софтуера TRAMO и SEATS. За даден ред алгоритъмът се опитва да намери модела с минимален BIC.

Liu (1989) предлага метод за идентификация на сезонни ARIMA модели, използвайки метод на филтриране и определени евристични правила; този алгоритъм се използва в софтуера SCA-Expert. Друг подход е описан от Mélard & Pasteels (2000), чийто алгоритъм за $ARIMA$ модели също позволява анализ на интервенции. Той е внедрен в софтуерния пакет „Time Series Expert“ (TSE-AX).

Други алгоритми се използват в комерсиален софтуер, въпреки че не са документирани в литературата. По-специално, Forecast Pro (Goodrich, 2000) е добре известен със своя добър автоматизиран $ARIMA$ алгоритъм, който е използван в състезанието за прогнозиране M3 (Makridakis & Hibon, 2000). Друг патентован алгоритъм е внедрен в Autobox (Reilly, D., 2000). Ord и Lowe (1996) предоставят преглед на някои от търговските софтуерни продукти, които внедряват автоматизирано прогнозиране с $ARIMA$.

Алгоритъмът на Hyndman-Khandakar (2008) е един от най-широко използваните алгоритми за идентификация на $ARIMA$ модели. Неговата популярност се дължи не само на неговата висока ефективност. Благодарение на реализацията в R чрез функцията auto.arima, която е част от пакета forecast, той е и публично достъпен. Алгоритъмът не е ограничен само до идентификацията. Чрез него се оценяват и параметрите на $ARIMA$ модел. След това моделът може лесно да се използва за прогнозиране. Поради това функцията auto.arima се утвърждава като стандарт в прогнозирането на динамични редове. Все пак трябва да се отбележи, че алгоритъмът не включва автоматизиран процес на диагностична проверка. Тази задача е оставена на експерта, който го прилага.

Алгоритъмът на Hyndman-Khandakar за автоматично $ARIMA$ моделиране може да се опише със следните етапи:

1. Броят на разликите d $(0 \leq d \leq 2)$ се определя с помощта на многократни тестове KPSS.

2. Стойностите на p и q след това се избират чрез минимализиране на коригирания критерий на Akaike AICc след изчисляване на последователни разлики на данните d пъти. Вместо да се разглежда всяка възможна комбинация от p и q, алгоритъмът използва поетапно търсене, за да премине през пространството на потенциално възможните модели.

2.1. Прилагат се четири първоначални модела:

- $ARIMA(0, d, 0)$,

- $ARIMA(2, d, 2)$,
- $ARIMA(1, d, 0)$,
- $ARIMA(0, d, 1)$.

Към всеки от тях се включва константа, с изключение на случаите, при които $d = 2$. Ако $d \leq 1$, се създава и допълнителен модел:

- $ARIMA(0, d, 0)$ без константа.

2.2. Най-добрият модел (с най-малката стойност на AICc) от използваните в стъпка 2.1 се определя като „текущ модел“.

2.3. Разглеждат се варианти на текущия модел:

- вариране на p и/или q от текущия модел с ± 1;
- включване/изключване на константата c от текущия модел.

Най-добрият модел, разгледан досега (или текущият модел, или един от тези варианти), става новият текущ модел. Стъпка 2.3 се повтаря, докато не е възможно да бъде намерен модел с по-нисък AICc.

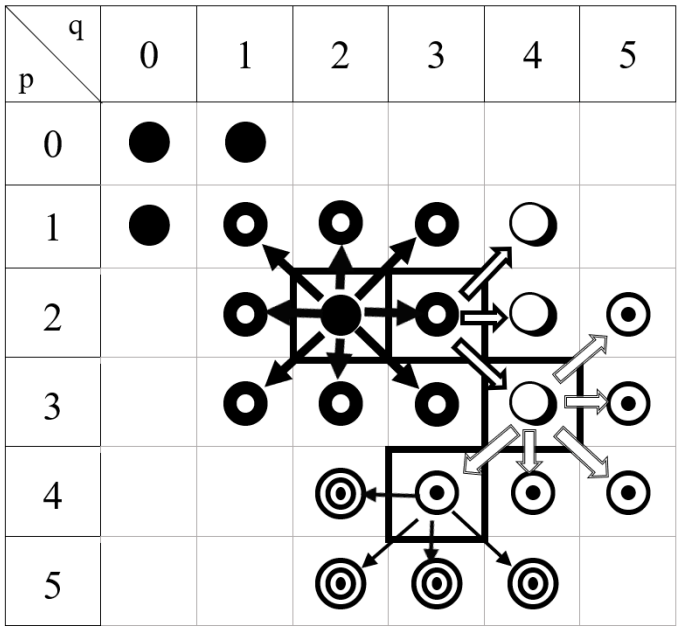
По подразбиране, без да преминава през всички съседни на „текущия модел“, определен на стъпка 2.2, алгоритъмът преминава към нов „текущ модел“, веднага щом бъде идентифициран по-добър модел.

На Фигура 2 е представено преминаването през пространството на $ARMA(p, q)$ моделите чрез алгоритъма на Hyndman-Khandakar. В таблицата са обхванати различни комбинации $ARMA(0,0)$, като порядъкът на AR се увеличава надолу по колоните, а порядъкът MA се увеличава по редовете. Плътните черни точки показват първоначалния набор от модели, разглеждани от алгоритъма. На схемата моделът $ARMA(2,2)$ има най-ниската стойност на AICc сред приложените модели. Той се нарича „текущ модел“ и е отбелязан в черната клетка. След това алгоритъмът претърсва съседните модели (клетки на таблицата), както е показано с плътните черни стрелки. Ако бъде намерен по-добър модел, той става новият „текущ модел“. В примера новият „текущ модел“ е $ARMA(2,3)$. Алгоритъмът продължава по този начин, докато не бъде намерен по-добрия модел. В примера се преминава през модела $ARMA(3,4)$ и се стига до модела $ARMA(4,3)$.

От сравнението между представените комплексни алгоритми могат да се направят изводи за техните качества и ограничения по различни критерии:

- *Изчислителна сложност*. Алгоритъмът на Hyndman-Khandakar е по-сложен в изчислително отношение от метода на Hannan & Rissanen, но предлага по-голяма гъвкавост при справянето с различни видове сезонност и трансформации. Методът на Liu също е сравнително лек в изчислително отношение, но изисква някои настройки от потребителя, особено при по-сложни модели;
- *Гъвкавост и автоматизация*. Алгоритъмът на Hyndman-Khandakar е най-автоматизиран и гъвкав от разгледаните. Той може автоматично да

избира правилните параметри и трансформации и предоставя добри резултати дори при големи набори от данни и сложни сезонни модели;



Фигура 2. Процес на поетапно търсене на нов „текущ модел“ чрез алгоритъма на Hyndman-Khandakar
Източник: авторите, по идея от (Hyndman & Athanasopoulos, 2021).

• *Надеждност и точност*. Алгоритъмът на Hyndman-Khandakar се отличава с висока точност благодарение на критерия AICc и способността за автоматично обработване на сложни динамични редове, включително чрез трансформации. Методът на Hannan & Rissanen също е надежден за *ARMA* модели, но има ограничения при сезонни динамични редове;

• *Приложимост.* Всички методи намират широко приложение в анализа на динамични редове, но алгоритъмът на Hyndman-Khandakar е предпочитан за анализи в реално време и за големи, комплексни времеви серии благодарение на своята висока автоматизация и адаптивност. Алгоритмите на Hannan & Rissanen и на Liu са подходящи за по-прости модели и по-малки набори от данни.

Заключение

В настоящото изследване са разгледани предимствата и ограниченията на различни автоматизирани алгоритми за идентификация на ARIMA модели в контекста на тенденцията за автоматизиране на прогнозирането на динамични редове. В условията на бързо променяща се икономическа и социална среда прогнозирането заема ключова роля за ефективното планиране и управление както в частния, така и в публичния сектор. Въпреки своето

стратегическо значение, създаването на точни прогнози остава трудоемък и времеемък процес, който изисква високо ниво на експертиза и е подложен на субективни грешки. Това е особено валидно при обработката на големи обеми от данни, където традиционният анализ става практически невъзможен и рисков. Подобни предизвикателства правят автоматизацията на процесите на прогнозиране все по-необходима в съвременната практика, осигурявайки обективност, ускоряване на анализа и повишаване на точността на прогнозите. Чрез проследяването на еволюцията на процеса по автоматизиране на прогнозите се откроиха както различните етапи, през които той преминава, така и основните му достижения, които го утвърждават като стандарт в съвременното прогнозиране.

Един от основните методи за прогнозиране са ARIMA моделите, разработени от Box и Jenkins. Те заемат важно място в традицията на прогнозирането на динамика. Въпреки значението и широкото си приложение, ARIMA моделирането е трудоемък процес, който изисква преминаване през множество етапи. Всеки от тях поставя различни предизвикателства и крие риск от субективни грешки, които могат да намалят точността на прогнозите. Това прави процеса на моделиране не само бавен и комплексен, но и изискващ висока степен на статистическа експертиза. Съществено място, с оглед оптимизирането на изследователския процес, заема етапът на идентификация на модела. От една страна, за успешното прогнозиране е необходима коректна идентификация на модела. От друга страна, класическата методология на Box и Jenkins е ефективна при идентификация на чисти авторегресионни или модели с плъзгаща се средна, но не е толкова ефективни при модели, които съдържат едновременно двата компонента. Автоматизирането на този етап чрез обективни методи и алгоритми за избор на параметри минимализира риска от субективни грешки и спомага за значително ускоряване на процеса. По този начин автоматизацията не само улеснява и стандартизира етапа на идентификация, но и повишава надеждността и точността на получените впоследствие прогнози.

Разработването на подходи за автоматизация, насочени към преодоляване на предизвикателствата в процеса на идентификация на ARIMA модели, е обект на изследване в множество научни публикации. Ние обособихме три основни групи – информационни критерии, еталонни методи и комплексни алгоритми. Те се разграничават хронологично, като първите две са от края на XX век, а третата – от началото XXI век.

В обобщение, настоящото изследване потвърждава изследователската теза, че автоматизираните подходи за идентификация на ARIMA модели значително намаляват субективния фактор и времето, необходимо за анализ, но не могат напълно да елиминират намесата на изследователя.

Научните приноси на настоящата разработка могат да се обособят в две направления. От една страна, обоснована е важната роля на идентификацията в процеса на ARIMA моделиране, като е обърнато внимание на

проблемите, породени от субективизма и времеемкостта, характерни за класическото прилагане на процедурата. Направен е извод, че в основата на решаването на тези проблеми са опитите за автоматизация и обективизация на този процес. От друга страна, въз основа на прегледа на теоретични и емпирични източници, в които е представена и методологичната база на прилаганите подходи, е извършена тяхната систематизация и е направен критичен анализ с извеждане на силните и слабите им страни.

Резултатите от научното изследване са полезни за практиката при решаване на конкретни прогностични задачи. Те позволяват да се избере конкретен автоматизиран алгоритъм, който да минимализира както времето за извършване на идентификацията на ARIMA моделите, така и субективизма на изследователя.

Използвани източници

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Beguin, J., Gourieroux, C., & Monfort, A. (1980). Identification of a mixed autoregressive-moving average process: The corner method,. От O. D. Anderson (Ред.), *Time Series* (стр. 423-436). Amsterdam: North-Holland.

Bickel, P., & Doksum, K. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association, 76*(374), 296–311. doi:10.1080/01621459.1981.10477649

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th изд.). New Jersey: John Wiley & Sons.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Statistical Methodology, 26*(2), 211–252. doi:10.1111/j.2517-6161.1964.tb00553.x

Breusch , T. S., & Pagan, A. R. (September 1979 r.). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), 1287-1294. doi:10.2307/1911963

Brooks, C., & Persand, G. (2002). Volatility forecasting for risk management. *Journal of Forecasting, 22*(1), 1-22. doi:10.1002/for.841

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water, 8*(3), 1-32. doi:10.1002/wat2.1520

Canova, F., & Hansen, B. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics, 13*(3), 237-252.

Chan, W.-S. (March 1999 r.). A comparison of some of pattern identification methods for order determination of mixed ARMA models. *Statistics & Probability Letters, 42*(1), 69-79. doi:10.1016/S0167-7152(98)00195-3

Choi, B. (2012). *ARMA Model Identification.* Springer Science & Business Media.

Chollet, F. (2017). *Deep Learning with Python.* Shelter Island, NY 11964: Manning Publications Co.

de Gooijer, J., Abraham, B., Gould, A., & Robinson, L. (December 1985 r.). Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey. *International Statistical Review, 51*(3), 301-329. doi:10.2307/1402894

de Gooijer, L., & Heuts, R. (1981). The Corner Method: An Investigation of an Order Discrimination Procedure for General ARMA Processes. *Journal of the Operational Research Society, 32*(11), 1039–1042. doi:10.1057/jors.1981.212

De Livera, A., Hyndman, R., & Snyder, R. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association, 106*(496), 1513 - 1527. doi:10.1198/jasa.2011.tm09771

Desai, A. N., Kraemer, M. U., Cori, A., Nouvellet, P., Herringer, M., Cohn, E., . . . Lassmann, B. (2019). Real-time Epidemic Forecasting: Challenges and Opportunities. *Health Security, 17*(4), 268–275. doi:10.1089/hs.2019.0022

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics forautoregressive time series with a unit root. *Econometrica: Journal ofthe Econometric Society*, 1057-1072.

Dickey, D., & Fuller, W. (1979). Distribution of the estimators forautoregressive time series with a unit root. *Journal of the Americanstatistical association, 74*(366a), 427-431.

Ding, S., Tao, Z., Li, R., & Qin, X. (December 2022 r.). A novel seasonal adaptive grey model with the data-restacking technique for monthly renewable energy consumption forecasting. *Expert Systems with Applications, 208*. doi:10.1016/j.eswa.2022.118115

Fildes, R., & Hastings, R. (1994). The Organization and Improvement of Market Forecasting. *Journal of the Operational Research Society, 45*(1), 1-16. doi:10.1057/jors.1994.1

Fildes, R., & Kourentzes, N. (October–December 2011 r.). Validation and forecasting accuracy in models of climate change. *International Journal of Forecasting, 27*(4), 968-995. doi:10.1016/j.ijforecast.2011.03.008

Gómez, V. (1998). *Automatic Model Identification in the Presence of Missing Observations and Outliners.* Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.

Gómez, V., & Maravall, A. (1998). *Guide for Using the Programs TRAMO and SEATS (Beta Version: December 1997), Working Papers 9805.* Banco de España.

Gómez, V., & Maravall, A. (2000). Automatic Modeling Methods for Univariate Series (chapter 7). От G. C. D. Peña (Ред.), *A Course in Time Series Analysis.* John Wiley & Sons, Inc. doi:10.1002/9781118032978.ch7

Goodrich, R. (2000). The Forecast Pro methodology. *International Journal of Forecasting, 4*(16), 533-535.

Granger, C. W., & Poon, S.-H. (2001). Forecasting Financial Market Volatility: A Review. 1-43. doi:http://dx.doi.org/10.2139/ssrn.268866

Gray, H., Kelley, G., & Mc Intire, D. (1978). A new approach to arma modeling. *Communications in Statistics - Simulation and Computation, 7*(1), 1-77. doi:10.1080/03610917808812057

Hannan, E., & Quinn, B. (January 1979 r.). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society: Series B (Methodological), 41*(2), 190-195. doi:10.1111/j.2517-6161.1979.tb01072.x

Hannan, E., & Rissanen, J. (April 1982 r.). Recursive estimation of mixed autoregressive-moving average order. *Biometrika, 69*(1), 81–94. doi:10.1093/biomet/69.1.81

Hylleberg, S., Engle, R., Granger, C., & Yoo, B. (1990). Seasonal in-tegration and cointegration. *Journal of econometrics, 44*(1-2), 215-238.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice, 3rd edition.* Melbourne, Australia: OTexts. Извлечено от OTexts.com/fpp3

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software, 27*(3), 1-22. doi:10.18637/jss.v027.i03

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Yasmeen, F. (2024). *Forecasting functions for time series and linear models. R package version 8.23.0*. Извлечено от forecast: https://pkg.robjhyndman.com/forecast/

Kaushik, S., Choudhury, A., Sheron, P. K., Dasgupta, N., Natarajan, S., Pickett, L. A., & Dutt, V. (2020). AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Frontiers in Big Data, 3*(4), 1-17. doi:10.3389/fdata.2020.00004

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics, 54*, 159-178.

Lii, K.-S. (May 1985 r.). Transfer function model order and parameter estimation. *Journal of Time Series Analysis, 6*(3), 153-169. doi:10.1111/j.1467-9892.1985.tb00406.x

Liu, L. (1989). Identification of seasonal arima models using a filtering method. *Communications in Statistics - Theory and Methods, 18*(6), 2279–2288. doi:10.1080/03610928908830035

Liu, L.-M., & Hanssens, D. (Jun 1982 r.). Identification of multiple-input transfer function models. *Communications in Statistics - Theory and Methods, 11*(3), 297-314. doi:10.1080/03610928208828236

Ljung, G. M., & Box, G. E. (August 1978 r.). On a measure of lack of fit in time series models. *Article Navigation, 65*(2), 297–303. doi:10.1093/biomet/65.2.297

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451-476. doi:10.1016/S0169-2070(00)00057-1

Mélard, G., & Pasteels, J. (2000). Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting, 16*(4), 497-508.

Ord, K., & Lowe, S. (1996). Automatic Forecasting. *The American Statistician, 50*(1), 88094. doi:10.1080/00031305.1996.10473549

Petruccelli, J., & Davies, N. (1984). Some restrictions on the use of corner method hypothesis tests. *Communications in Statistics - Theory and Methods, 13*(5), 543–551. doi:10.1080/03610928408828699

Reilly, D. (2000). The AUTOBOX system. *International Journal of Forecasting, 16*(4), 531-533.

Reilly, D. R. (1980). Experiences with n automatic Box-Jenkins modelling algorithm. От O. D. (Ed.), *Time series analysis.* North-Holland.

Rezayat, F., & Anandalingam, G. (1988). Using instrumental variables for selecting the order of arma models. *Communications in Statistics - Theory and Methods, 17*(9), 3029–3065. doi:10.1080/03610928808829787

Said, S., & Dickey, D. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika, 71*(3), 599-607.

Schwarz, G. (Mar. 1978 r.). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464. Извлечено от https://www.jstor.org/stable/2958889

Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. *SciPy, 7*(1).

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (July 2016 r.). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1-26. doi:10.1016/j.ejor.2015.11.010

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician, 72*(1), 37-45. doi:10.1080/00031305.2017.1380080

Tiao, G. (2000). Univariate Autoregressive Moving-Average Models (Chapter 3). От G. C. D. Peña (Ред.), *A Course in Time Series Analysis* (стр. 53-85). John Wiley & Sons, Inc. doi:10.1002/9781118032978.ch3

Tsay, R., & Tiao, G. (July 1984 r.). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *Journal of the American Statistical Association, 79*(385), 84-96. doi:10.1080/01621459.1984.10477068

Tsay, R., & Tiao, G. (August 1985 r.). Use of canonical analysis in time series model identification. *Biometrika, 72*(2), 299-315. doi:10.1093/biomet/72.2.299

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (June 2014 r.). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies, 43*(Part 1), 3-19. doi:10.1016/j.trc.2014.01.005

Wilson, G. T. (2000). Model Fitting and Checking, and the Kalman Filter (Chapter 4). От G. C. D. Peña (Ред.), *A Course in Time Series Analysis* (стр. 86-110). John Wiley & Sons, Inc. doi:10.1002/9781118032978

Woodward, W., & Gray, H. (Sep. 1981 r.). On the Relationship Between the S Array and the Box-Jenkins Method of ARMA Model Identification. *Journal of the American Statistical Association, 76*(375), 579-587. doi:10.2307/2287515

Zivot, E., & Andrews, D. (2002). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics, 20*(1), 25-44.

Димитров, А. (1980). За инерционността в икономиката. *Икономическа мисъл, 8*, 41-52.

Иванов, Л., Касабова, С., & Шопова, М. (2017). *Статитическо изследване и прогнозиране на развитието.* Свищов: АИ "Ценов".

том 33, 2025 г.

ИНСТИТУТ ЗА НАУЧНИ ИЗСЛЕДВАНИЯ
ПРИ СТОПАНСКА АКАДЕМИЯ
„Д. А. ЦЕНОВ“ - СВИЩОВ

АЛМАНАХ
НАУЧНИ ИЗСЛЕДВАНИЯ

ИНОВАЦИИ, УСТОЙЧИВОСТ, ДИГИТАЛНА ТРАНСФОРМАЦИЯ – ЗА УСКОРЕНО ИКОНОМИЧЕСКО РАЗВИТИЕ

том 33, 2025 г.

Академично издателство „ЦЕНОВ“
Свищов - 2025 г.

АЛМАНАХ НАУЧНИ ИЗСЛЕДВАНИЯ

Издава се със средства от целевата субсидия за научна дейност на СА „Д. А. Ценов”, съгласно Наредбата за условията и реда за оценката и планирането, разпределението и разходването на средствата от държавния бюджет за финансиране на присъщата на държавните висши училища научна или художествено творческа дейност.

ISSN 1312-3815

This issue is funded by the state “Ordinance on the terms and procedure for the evaluation and planning, allocation and spending of the state budget funds in financing scientific or artistic activities, intrinsic to state higher schools” for the inherent to the "D. A. Tsenov" Academy of Economics scientific activity.

ISSN 1312-3815

СЪДЪРЖАНИЕ

Раздел I
Ускорено икономическо развитие и иновации

Раздел II
Устойчиво и балансирано развитие

Раздел III
Дигитална трансформация на икономиката и обществото

CONTENTS

Section I
Accelerated Economic Development and Innovation

Section II
Sustainable and Balanced Development

Section III
Digital Transformation of Economy and Society

СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ”

АЛМАНАХ НАУЧНИ ИЗСЛЕДВАНИЯ

ТОМ 33

ИНОВАЦИИ, УСТОЙЧИВОСТ, ДИГИТАЛНА ТРАНСФОРМАЦИЯ – ЗА УСКОРЕНО ИКОНОМИЧЕСКО РАЗВИТИЕ

Даден за печат на 20.02.2025 г., излязъл от печат на 28.02.2025 г.
Поръчка № 18923, тираж: 50 бр.

Издателство и печат: Академично издателство „Ценов”
Свищов, ул. Цанко Церковски № 11А

ISSN 1312-3815