Indeks 371262
e-ISSN 2657-9545
ISSN 0033-2372

Przegląd Statystyczny
Statistical Review

Vol. 72 | No. 2 | 2025

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

INFORMATION FOR AUTHORS

Przegląd Statystyczny. *Statistical Review* publishes original research papers on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision science and data analysis. The manuscripts considered for publication should significantly contribute to the theoretical aspects of the aforementioned fields or shed new light on the practical applications of these aspects. Manuscripts reporting important results of research projects are particularly welcome. Review papers, shorter papers reporting on major conferences in the field, and reviews of seminal monographs are eligible for submission on the Editor-in-Chief's request.

Since 1 May 2019, the journal has been publishing articles in English.

Any spelling style is acceptable as long as it is consistent within the manuscript.

All works should be submitted to the journal through the Editorial System (https://www.editorialsystem.com/pst).

For details of the submission process and editorial requirements, please visit https://ps.stat.gov.pl/ForAuthors.

Przegląd Statystyczny
Statistical Review

Vol. 72 No. 2 2025

EDITORIAL OFFICE'S ADDRESS

Statistics Poland (GUS), al. Niepodległości 208, 00-925 Warsaw, Poland

Language editing: Scientific Journals Division, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Maciej Adamowicz

Zakład Wydawnictw Statystycznych

Printed and bound: Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warsaw, Poland, zws.stat.gov.pl

Website: ps.stat.gov.pl

ISSN 0033-2372
e-ISSN 2657-9545
Index 371262

Information on the sales of the journal: Statistical Publishing Establishment
Phone no.: +48 22 608 32 10, +48 22 608 38 10

Order no. 313/2025

CONTENTS

Przegląd Statystyczny. Statistical Review, 2025, vol. 72, 2, 1–32 https://doi.org/10.59139/ps.2025.02.1
Received: 1.07.2025, revised: 28.10.2025, accepted: 10.11.2025

Skew plasticising component normal I distribution

Piotr Sulewski,[a] Damian Stoltmann[b]

Abstract. This article has two goals. The first (main) goal is to introduce a new flexible distribution defined on an infinite domain $(-\infty, \infty)$. This distribution has been named the skew plasticising component normal distribution. The second (additional) goal is to present a chronological overview of distributions belonging to the large family of normal plasticising distributions. Some properties of the proposed distribution such as the PDF, CDF, quantiles, generator, moments, skewness, kurtosis and moments of order statistics are presented. The unknown parameters of the new distribution are estimated by means of the maximum likelihood method. The Shannon entropy, the Hessian Matrix and the Fisher Information Matrix are also presented. The study provides illustrative examples of the applicability and flexibility of the introduced distribution. The most important R codes are provided in Appendix 2.

Keywords: plasticising component, bimodal model, departure from normality, Azzalini's transformation

JEL: C02, C16, C46

1. Introduction

The Gaussian distribution should be classified as a normal distribution (ND) due to the regularity and clarity of the roles of its parameters and its unique mathematical properties. However, it appears that its enormous popularity is disproportionate to its real applications. In many practical cases, empirical data exhibit skewness, heavy tails or multimodality that cannot be captured by the classical ND. The ND then needs to be plasticised.

As shown in numerous studies, various approaches have been developed to plasticise the ND, forming a broad family of normal plasticised distributions.

The relevant literature shows that there are various methods of plasticising the ND, forming a family of normal plasticising distributions.

The first group of normal plasticising distributions is a mixture of distributions, i.e. a mixture of a plasticising component and an ND. A mixture distribution, which is a combination of at least two distributions, can fit more characteristics than sample

[a] Pomeranian University in Słupsk, Institute of Exact and Technical Sciences, Department of Computer Science, ul. Arciszewskiego 22a, 76–200 Słupsk, Poland, e-mail: piotr.sulewski@apsl.edu.pl, ORCID: https://orcid.org/0000-0002-0788-6567.

[b] Pomeranian University in Słupsk, Institute of Exact and Technical Sciences, Department of Computer Science, ul. Arciszewskiego 22a, 76–200 Słupsk, Poland, e-mail: damian.stoltmann@upsl.edu.pl, ORCID: https://orcid.org/0000-0001-7053-2684.

data might contain. Owing to this property, mixture distributions have been widely used in statistical sciences (Frühwirth-Schnatter, 2006; Martínez-Flórez et al., 2022). Behboodian (1970) presents a procedure for determining whether a mixture of two NDs (also called the compound normal (CN) distribution) is unimodal or not. Stephens (2000) studied what is called the 'label switching' problem, caused by the symmetry in the likelihood of the model parameters. A common response to this problem is to remove the symmetry by using artificial identifiability constraints. Lin et al. (2007) used a mixture of skew distribution models to fit multimodal data and datasets with bimodal features. Magnus and Magnus (2019) considered a subclass of the mixture models, namely normal latent factor mixture models. Popović et al. (2017) proposed the extended mixture ND, based on a linear mixture model, whose Probability Density Function (PDF) is symmetrical. Wang and Song (2017) developed a new equivalent linearisation method for nonlinear random vibration analysis. The method employs a Gaussian mixture distribution model to approximate the probabilistic distribution of a nonlinear system response. Sulewski (2022b) defined an ND with a plasticising component (NDPC).

The second group is a family of distributions with a plasticising formula located in the exponential function of the ND. This family includes e.g. the lognormal (Gaddum, 1945; Kapteyn, 1916), SL, SB, SU (Johnson, 1949), Birnbaum–Saunders (Athayde et al., 2012; Birnbaum & Saunders, 1969; Sulewski & Stoltmann, 2023), inverse Gaussian (Chhikara & Folks, 1977), sinh-normal (Rieck & Nedelman, 1991), DS normal (Sulewski, 2021), the Sulewski Plasticizing Component (Sulewski & Volodin, 2022), SC and SD (Sulewski, 2023) distributions.

The third group is a two-piece family of distributions. The PDFs of these distributions are in Table A1 (see Appendix 1). The family of distributions includes: the two-piece skew-normal (TPSN, Kim, 2005), generalised skew-normal (GSN1, Gómez et al., 2006), extended epsilon skew-normal (EESN, Salinas et al., 2007), epsilon skew normal (ESN, Mudholkar & Hutson, 2000), flexible epsilon-skew-normal (FESN, Arellano-Valle et al., 2010), skew-two-piece skew-normal (STPSN, Jamalizadeh & Arabpour, 2011), generalised two-piece skew-normal (GTPSN, Jamalizadeh & Arabpour, 2011), generalised skew-two-piece skew-normal (GSTPSN, Jamalizadeh & Arabpour, 2011), generalised two-piece skew-normal (GTPSN, Kumar & Anusree, 2013), two-piece power normal (TPPN, Sulewski, 2021) distributions.

The fourth group is a family of distributions with PDF $f(x;\boldsymbol{\theta})\phi(x)$, where the $f(x;\boldsymbol{\theta})$ is some function with parameter vector $\boldsymbol{\theta}$ and $\phi(x)$ is a PDF of $N(0,1)$. The PDFs of these distributions are presented in Table A1 in Appendix 1. The family of distributions includes the symmetric bimodal normal (BN, Arellano-Valle & Azzalini, 2008), alpha-skew-normal (ASN, Elal-Olivero, 2010), double normal

(DN, Alavi, 2012), generalised alpha-skew-normal (GASN, Handam, 2012), Balakrishnan alpha-skew-normal (BASN, Hazarika et al., 2020), two-piece normal (TN, Salinas et al., 2023), alpha-beta skew-normal (ABSN, Shafiei et al., 2016), Balakrishnan alpha-beta-skew-normal (BABSN, Shah et al., 2021) and flexible alpha normal (FAN, Martínez-Flórez et al., 2022) distributions.

The fifth group is a family of distributions with PDF $f(x;\boldsymbol{\theta})exp(-|x|^{\gamma}/\gamma)$ $(\gamma > 0)$. The PDFs of these distributions are provided in Table A1 in Appendix 1. The family of distributions includes the generalised normal (GN, Kumar & Anusree, 2015), bimodal generalised normal (BGN, Mahmoudi et al., 2019) and alpha-skew generalised normal (ASGN, Mahmoudi et al., 2019) distributions.

The sixth group is a power normal family of distributions. The PDFs of these distributions are available in Table A1 in Appendix 1. The family of distributions includes the power normal (PN, Gupta & Gupta, 2008), generalised power-normal (GPN, Arnold et al., 2002), Durrans's power normal (Durrans, 1992) and power skew asymmetric normal (PSAN, Martínez-Flórez et al., 2014) distributions.

The seventh group is the Azzalini family of distributions. Azzalini (1985) added a skewness parameter to the Cumulative Distribution Function (CDF) of the ND and defined the skew-normal (SN) distribution with the following PDF:

$$f_{SN}(x;\lambda) = 2\phi(x)\Phi(\lambda x)\ (\lambda \in R), \tag{1}$$

where ϕ and Φ are the PDF and CDF of $N(0,1)$, respectively.

This distribution and its variations have been discussed by several authors including Azzalini (1985; 1986), Henze (1986), Azzalini & Dalla Valle (1996), Branco & Dey (2001), Loperfido (2001), Arnold et al. (2002) and Azzalini and Chiogna (2004). The PDFs of the Azzalini family of distributions are shown in Table A1 in Appendix 1. This family includes the skewed normal (SN1, Arnold et al., 2002), skew-curved normal (SCN, Arellano-Valle et al., 2004), skew-generalised normal (SGN, Arellano-Valle et al., 2004), flexible generalised skew-normal of order 3 (FGSN3, Ma & Genton, 2004), Balakrishnan skew–normal (BSN, Sharafi & Behboodian, 2008), generalised skew-normal (Gupta & Gupta, 2004), generalised skew-normal (GSN2, Jamalizadeh & Balakrishnan, 2008), two-parameter Balakrishnan skew-normal (TPBSN, Bahrami et al., 2009), generalised skew-normal (GSN, Jamalizadeh & Balakrishnan, 2008), skew bimodal normal (SBN) (Elal-Olivero et al., 2009), skew-flexible normal (SFN, Gómez et al., 2011), extended skew generalised normal I (ESGN1, Choudhury & Matin, 2011), extended skew generalised normal II (ESGN2, Choudhury & Matin, 2011), generalised mixture of standard normal and skew-normal (GMNSN, Kumar & Anusree, 2011), normal-skew-normal (NSN, Gómez et al., 2013), flexible skew-generalised normal (FSGN, Bahrami & Qasemi, 2015), flexible skew-curved normal (FSCN, Bahrami & Qasemi, 2015), extended skew generalised normal III

(ESGN3, Kumar & Anusree, 2015), shape-skew-generalised normal (SSGN, Rasekhi et al., 2017), skew-bimodal normal-normal (SBNN, Alavi & Tarhani, 2017), extended skew-normal (ESN, Seijas-Macias et al., 2017), generalised alpha-beta skew-normal (GABSN, Shah et al., 2023) and flexible alpha-skew-normal (FASN, Das et al., 2023) distributions.

Despite this extensive literature, many existing plasticising models are either computationally demanding, lacking interpretability or they fail to simultaneously model skewness and bimodality in a parsimonious way. Motivated by these limitations, we introduce a new member of the normal plasticised distributions family, namely the skew plasticising component normal (SPCN1) distribution. This model provides a simple yet flexible way to generate a wide range of unimodal and bimodal shapes while preserving a clear probabilistic interpretation of its parameters.

The SPCN1 distribution extends the idea of a compound normal model by introducing a skew plasticising component that modifies both tails and the central concentration of the ND. The proposed formulation enables continuous control over skewness and kurtosis and allows the model to adapt to empirical data exhibiting asymmetric or bimodal behaviour. Furthermore, its analytical tractability makes it suitable for estimation via maximum likelihood and for use in simulation and goodness-of-fit studies.

This article has two goals. The first (main) goal is to introduce the SPCN1 distribution defined on an infinite domain $(-\infty, \infty)$. The second (additional) goal is to provide a chronological overview of distributions belonging to the large family of normal plasticising distributions.

This paper is organised as follows. Section 2 presents the properties of the SPCN1 distribution such as the PDF, CDF, quantiles, generator, moments, skewness, kurtosis and moments of order statistics. The Shannon entropy is presented in Section 3, while the Hessian Matrix and the Fisher Information Matrix are presented in Section 4. The maximum likelihood estimation is discussed in Section 5, while illustrative examples of the applicability and flexibility of the proposed distribution are presented in Section 6. The conclusions are presented in Section 7. The most important R codes are provided in Appendix 2. The PDFs of the large family of normal plasticising distributions are given in Table 1.

2. Properties of the proposed distribution

2.1. The probability density function

The PDF and CDF of the plasticising component (PC) are given (Sulewski, 2022b) by:

$$f_{PC}(x;c) = \frac{c}{\sqrt{2\pi}}|x|^{c-1}exp\left[-\frac{1}{2}|x|^{2c}\right] = c|x|^{c-1}\phi(|x|^{c}), \tag{2}$$

$$F_{PC}(x;c) = \Phi[sgn(x)|x|^c], \tag{3}$$

where $c \geq 1$ is the shape parameter, and ϕ and Φ are the PDF and CDF of $N(0,1)$, respectively.

Definition 1. (the Azzalini transformation) The distribution of random variable X with the PDF given by:

$$f(x;c,d) = 2f_{PC}(x;c)F_{PC}(xd;c) = 2c|x|^{c-1}\phi(|x|^c)\Phi[sgn(xd)|xd|^c] \tag{4}$$

or

$$f(x;c,d) = \frac{c|x|^{c-1}exp[-0.5|x|^{2c}]}{\sqrt{2\pi}}\left\{1 + erf\left[\frac{|xd|^c sgn(xd)}{\sqrt{2}}\right]\right\},$$

or

$$f(x;c,d) = \frac{c|x|^{c-1}exp[-0.5|x|^{2c}]}{\sqrt{2\pi}}erfc\left[\frac{-|xd|^c sgn(xd)}{\sqrt{2}}\right],$$

or

$$f(x;c,d) = \frac{c}{\sqrt{2\pi}}\begin{cases} x^{c-1}exp[-0.5x^{2c}]erfc\left[\frac{-(xd)^c}{\sqrt{2}}\right], x \geq 0 \\ (-x)^{c-1}exp[-0.5(-x)^{2c}]erfc\left[\frac{(-xd)^c}{\sqrt{2}}\right], x < 0 \end{cases}$$

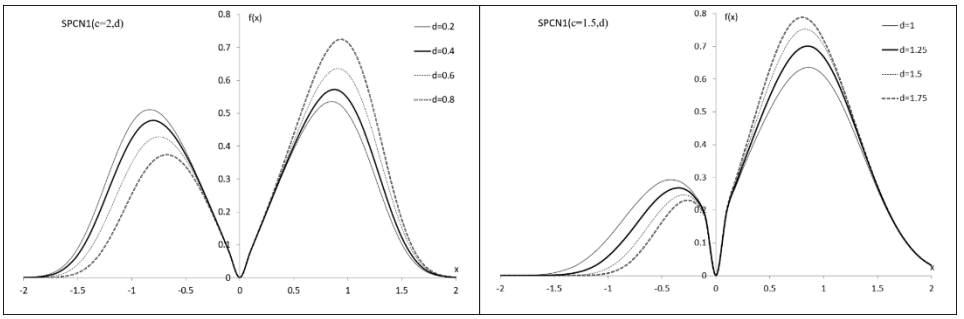
is called the skew plasticising component normal I (SPCN1) distribution, where $c \geq 1$ is the shape parameter, $d \geq 0$ is the skewness parameter, $erf(.)$ is the error function and $erfc(.)$ is the complementary error function. For $c = 1, d = 0$, we obtain the $N(0,1)$ and for $d = 0$, we obtain the PC (2). The symbol 'I' denotes the authors' first proposal for the skew plasticising component normal (SPCN) distribution. The R codes of the dSPCN1 function are presented in Appendix 2.

Figure 1 shows the PDF of the SPCN1(c, d) for some values of the parameters. If $c > 1$, the PDF has two modes of various heights.

The SPCN1(c, d) can be used to deviate from the $N(0,1)$. The similarity measure between our proposal and the $N(0,1)$ was provided by Sulewski (2022a):

$$(c,d) = \int_{-\infty}^{\infty} min[f(x;c,d), \phi(x)]dx. \tag{5}$$

Figure 1. The PDF of the SPCN1(c, d) for selected values of the parameters



Source: authors' work.

As mentioned before, the $SP\text{CN}1(1,0)$ is the $N(0,1)$, so $M_{max}(1,0) = 1$.

In addition to similarity measure M, the difference between the distributions can also be quantified using the Kullback–Leibler (KL) divergence. For two PDFs, p and q, the KL divergence is defined as (Kullback & Leibler, 1951):

$$KL(p,q) = \int_{-\infty}^{\infty} p\left(x;\boldsymbol{\theta_p}\right) log_2 \frac{p\left(x;\boldsymbol{\theta_p}\right)}{q\left(x;\boldsymbol{\theta_q}\right)} dx, \tag{6}$$

where $\boldsymbol{\theta_p}$ is the parameter vector of function $p(x)$, $\boldsymbol{\theta_q}$ is the parameter vector of function $q(x)$. KL takes the values of $(0,\infty)$.

In our context, the KL is given by:

$$KL(f,\phi) = \int_{-\infty}^{\infty} f(x;c,d) log_2 \frac{f(x;c,d)}{\phi(x)} dx.$$

Tables 1 and 2 summarise similarity measure M and the complement of the Kullback–Leibler divergence $1 - KL$ between the $SPCN1$ distribution and standard ND $N(0,1)$. The KL measure has been written as $1 - KL$ to make it easier to compare with the M similarity measure.

In both cases, as parameters c and d depart from their reference values ($c = 1$ and $d = 0$), similarity measure M decreases, indicating a gradual divergence from the ND.

The values of $1 - KL$ show a consistent trend with M: as d or c increases, divergence KL between $SPCN1$ and $N(0,1)$ grows. For small deviations of c and d, both measures suggest a strong resemblance. For larger parameter values, the $SPCN1$ distribution becomes increasingly non-normal, as evidenced by the steep decline of both measures.

In particular, Table 1 illustrates that skewness parameter d has a strong influence on similarity: even moderate departures from $d = 0$ lead to a noticeable drop in both M and $1 - KL$. Table 2 shows a similar, but slightly smoother effect for shape parameter c.

Overall, both measures (M and KL) provide consistent quantitative evidence that $SPCN1(c, d)$ continuously and controllably deviates from $N(0,1)$, confirming its flexibility and interpretability as a 'plasticised' version of the ND.

Table 1. Similarity measure $M(1, d)$ and $KL(1, d)$ between the SPCN1$(1, d)$ and $N(0,1)$

d	0	0.158	0.325	0.51	0.727	1	1.376	1.963	3.078	6.314	29.82
$M(1, d)$	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5
$1 - KL(1, d)$	1	0.989	0.955	0.898	0.82	0.721	0.603	0.468	0.32	0.162	0.035

Source: authors' work.

Table 2. Similarity measure $M(c, 0)$ and $KL(c, 0)$ between the SPCN1(c, 0) and $N(0,1)$

c	1	1.118	1.253	1.404	1.576	1.775	2.005	2.278	2.602	2.991	3.469
$M(c, 0)$	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5
$1 - KL(c, 0)$	1	0.985	0.945	0.885	0.807	0.712	0.602	0.474	0.331	0.172	−0.006

Source: authors' work.

2.2. Cumulative density function

Let X~SPCN1(c, d). The CDF of the SPCN1 distribution, based on definition 1, is given by the following formula:

$$(x; c, d) = 2c \int_{-\infty}^{x} |t|^{c-1} \phi(|t|^c) \Phi[sgn(td)|td|^c] dt. \tag{7}$$

For $x < 0$, formula (7) can be written as:

$$F(x; c, d) = \frac{c}{\sqrt{2\pi}} \int_{-\infty}^{x} \frac{(-t)^{c-1}}{e^{0.5(-t)^{2c}}} erfc\left[\frac{(-td)^c}{\sqrt{2}}\right] dt$$

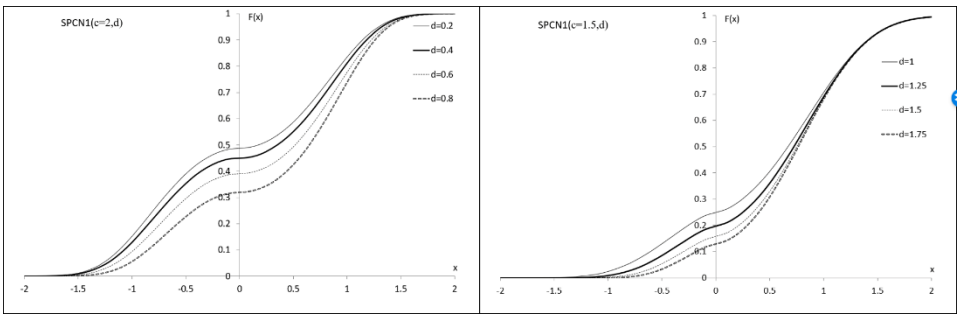
and for $x \geq 0$, we have

$$F(x; c, d) = \frac{c}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{0} \frac{(-t)^{c-1}}{e^{0.5(-t)^{2c}}} erfc\left[\frac{(-td)^c}{\sqrt{2}}\right] dt + \int_{0}^{x} \frac{t^{c-1}}{e^{0.5t^{2c}}} erfc\left[\frac{-(td)^c}{\sqrt{2}}\right] dt \right\}.$$

The R codes of the pSPCN1 function are presented in Appendix 2.

Figure 2 shows the CDF of the SPCN1(c, d) for some parameter values. For $c > 1$, we obtain two sub-CDFs placed at certain levels, which means the distribution is bimodal.

Figure 2. The CDF of the SPCN1(c, d) for some values of the parameters



Source: authors' work.

It is quite understandable that the CDF does not have a closed form, since the distribution in question has its origin in the Gaussian distribution. A similar situation, as can be seen below, concerns quantiles, the pseudo-random number generator, non-central moments, moments of order statistics, and the Shannon entropy. However, this is not a problem from the perspective of practical applications, because thanks to numerical methods, we obtain user functions written, for example in the R environment (see Appendix 2).

2.3. Quantile and pseudo-random number generator

Let $X \sim \text{SPCN1}(c, d)$. The p-th $(0 < p < 1)$ quantile x_p is a solution to equation

$$\frac{c}{\sqrt{2\pi}} \int_{-\infty}^{x_p} \frac{|x|^{c-1}}{e^{0.5|x|^{2c}}} erfc\left[\frac{-|xd|^c sgn(xd)}{\sqrt{2}}\right] dx - p = 0. \tag{8}$$

The R codes of the qSPCN1 function are presented in Appendix 2.
Let $X \sim \text{SPCN1}(c, d)$ and $R \sim Unif(0,1)$. The pseudo-random number generator of X is a solution to equation

$$\frac{c}{\sqrt{2\pi}} \int_{-\infty}^{X} \frac{|x|^{c-1}}{e^{0.5|x|^{2c}}} erfc\left[\frac{-|xd|^c sgn(xd)}{\sqrt{2}}\right] - R = 0. \tag{9}$$

The R codes of the rSPCN1 function are presented in Appendix 2.

2.4. Moments

Let $X \sim \text{SPCN1}(c, d)$. Non-central moments of X are given by:

$$\alpha_k(c,d) = E(X^k) = \frac{c}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{|x|^{c-1}}{e^{0.5|x|^{2c}}} erfc\left[\frac{-|xd|^c sgn(xd)}{\sqrt{2}}\right] dx \ \ (k = 1,2,\dots). \quad (10)$$

The R codes of the mSPCN1 function are presented in Appendix 2.

2.5. Skewness and kurtosis

Based on the order (non-central) moments and using their relationships with central moments $\mu_k = \sum_{i=0}^{k} (-1)^i \binom{k}{i} \alpha_{k-i} \alpha_1^i$, we can easily calculate skewness γ_1 and kurtosis γ_2 of the $SPCN1(c, d)$.

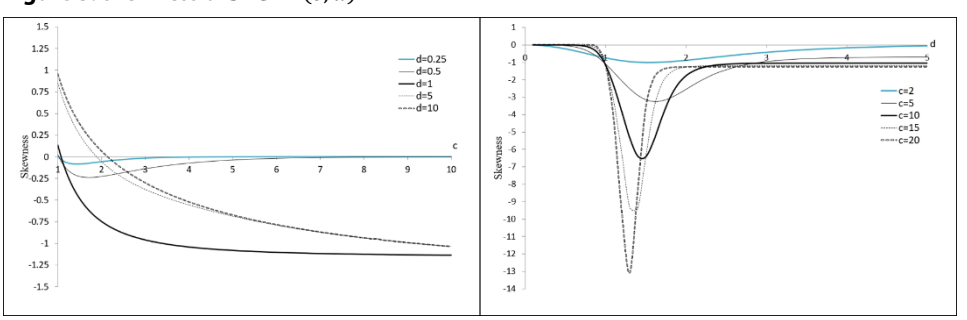
The skewness of $\text{SPCN1}(c, d)$ is defined as:

$$\gamma_1(c,d) = \frac{\mu_3}{\mu_2^{1.5}} = \frac{\alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3}{(\alpha_2 - \alpha_1^2)^{1.5}},$$

where α_i $(i = 1,2,3)$ are given by (10). The R codes of the g1SPCN1 function are presented in Appendix 2.

Figure 3 shows γ_1 as a function of c for selected d values (left) and γ_1 as a function of d for selected c values (right). $\gamma_1(c)$ is a decreasing function for $d \geq 1$, especially for the initial values of the arguments and inversely unimodal for $0 < d < 1$, e.g. $\gamma_1^{min}(1.713,0.5) = -0.239$. As d increases, $\gamma_1(c)$ decreases. $\gamma_1(d)$ is inversely unimodal for $c \geq 1$ e.g. $\gamma_1^{min}(10,1.455) = -6.54$. The $\gamma_1(d)$ function is strictly monotonical for the initial values of the arguments.

Figure 3. Skewness of $SPCN1(c, d)$



Source: authors' work.

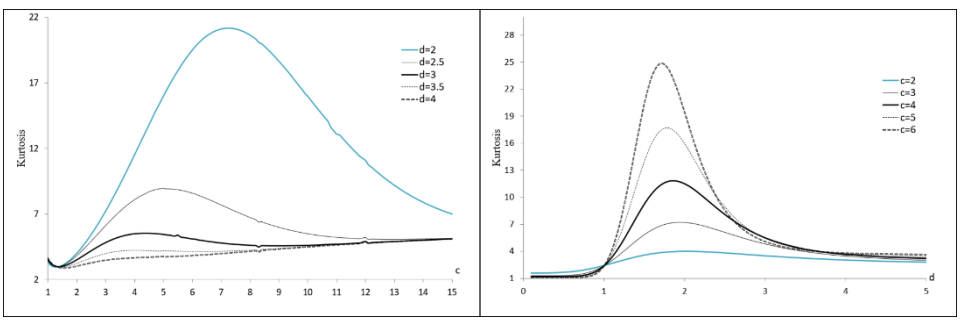
The kurtosis of SPCN1(c,d) is given by:

$$\gamma_2(c,d) = \frac{\mu_4}{\mu_2^2} = \frac{\alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4}{(\alpha_2 - \alpha_1^2)^2},$$

where α_i $(i = 1,2,\ldots,4)$ are given by (10). The R codes of the g2SPCN1 function are presented in Appendix 2.

Figure 4 shows γ_2 as a function of c for selected d values (left) and γ_2 as a function of d for selected c values (right). $\gamma_2(c)$ is the unimodal function, e.g. $\gamma_2^{max}(7.253,2) = = 21.162$. As d increases, $\gamma_2(c)$ decreases. $\gamma_2(d)$ is the unimodal function, e.g. $\gamma_2^{max}(6,1.716) = 24.87$. As c increases, $\gamma_2(c)$ also increases. As Malakhov's inequality $\gamma_2 \geq \gamma_1^2 + 1$ (Malakhov, 1978) indicates, we obtain γ_2 equal to no less than 1.

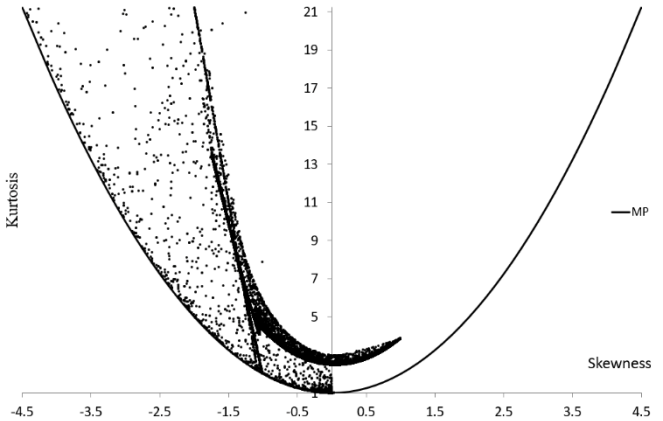
Figure 4. Kurtosis of $SPCN1(c,d)$



Source: authors' work.

We calculate γ_1 and γ_2 for 10^5 random values of $c = Unif(1,100)$ and $d = Unif(0,100)$. Figure 5 presents a set of points (γ_1, γ_2) located in a rectangle $(-4.5,4.5) \times (1,21.25)$. The symbol MP denotes the $\gamma_2 = \gamma_1^2 + 1$ Malakhov parabola. We obtain $\gamma_1 \in (-4.824,0.994)$, $\gamma_2 \in (1,21.244)$ and a very interesting shape.

Figure 5. Variability range of γ_1 and γ_2 of $SPCN1(c,d)$



Source: authors' work.

2.6. Moments of order statistics

Let $X_{i,n}$ be the i-th order statistic $\left(X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}\right)$ in a sample of size n from the SPCN1(c,d). The k-th moment of the i-th order statistic, $X_{i,n}$ is defined as:

$$\alpha_{k,i,n} = E\left(\mathrm{X}_{i,n}^{k}\right) = \int_{-\infty}^{\infty} x^k f_{i,n}(x;a,b)dx, \tag{11}$$

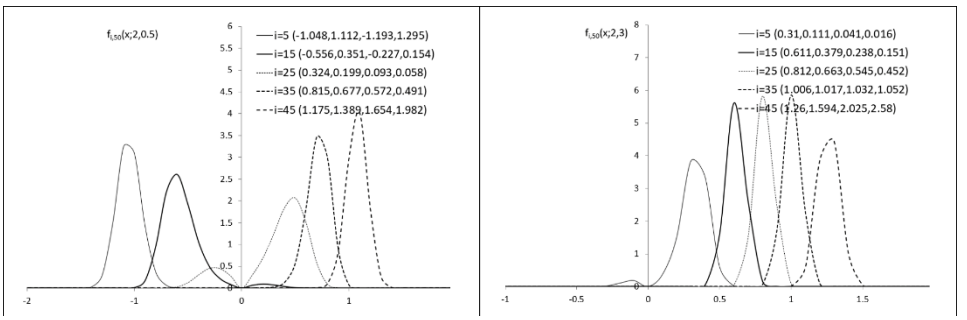
where

$$f_{i,n}(x;c,d) = i!\binom{n}{i}\frac{f(x;c,d)}{F(x;a,b)^{1-i}}[1-F(x;a,b)]^{n-i} \tag{12}$$

and $f(x;a,b), F(x;a,b)$ are given by (4) and (7). The R codes of the mOSSPCN1 function are presented in Appendix 2. Note that from (12), we have $f_{2,2}(x;c,d) = = 2f(x;c,d)F(x;a,b)$, so we obtain the Azzalini transformation without the skewness parameter.

Figure 6 shows the PDF of the $X_{5i,30}$ $(i = 1,2,3,4,5)$ of the $SPCN1(2,0.5)$ (left) and $SPCN1(2,3)$ (right), as well as $\alpha_{k,i,n}(k = 1,2,3,4)$ in brackets, respectively. The $f_{i,50}(x_m;a,b)$ value is the highest for $i = 45$ (Figure 6, left) and for $i = 35$ (Figure 6, right). The values of $\alpha_{k=1,i,n}$ and $\alpha_{k=3,i,n}$ increase along with the i value.

Figure 6. PDF of the $X_{5i,30}$ $(i = 1,2, \ldots ,5)$ of the $SPCN1(c,d)$ distribution



Source: authors' work.

3. Shannon entropy

Let $f(x;c,d)$ be the PDF (4). Shannon entropy S is given by (Shannon, 1948):

$$S(c,d) = -\int_{-\infty}^{\infty} f(x;c,d) ln[f(x;c,d)]dx, \tag{13}$$

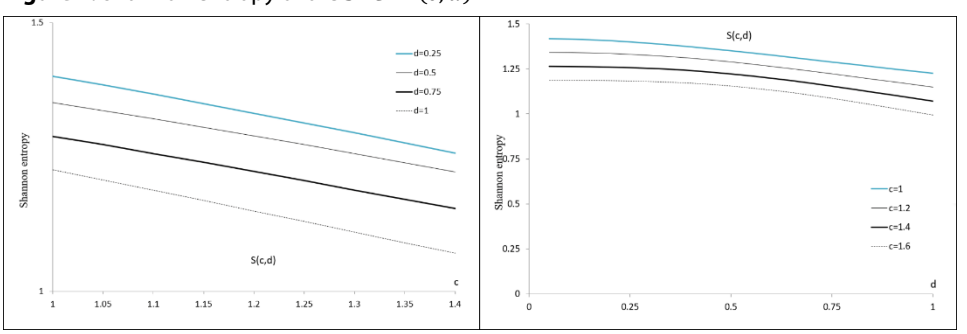
where

$$ln[f(x;c,d)] = nln\frac{c}{\sqrt{2\pi}} + (c-1)ln|x| - 0.5|x|^{2c} + ln\left\{erfc\left[\frac{-|xd|^c sgn(xd)}{\sqrt{2}}\right]\right\}. \tag{14}$$

The R codes of the sSPCN1 function are presented in Appendix 2.

Figure 7 shows the Shannon entropy as a function of c for selected d values (left) and as a function of d for selected c values (right). We obtain decreasing functions with a very small numerical range of variability.

Figure 7. Shannon entropy of the $SPCN1(c,d)$



Source: authors' work.

4. Hessian Matrix and Fisher Information Matrix

The Hessian Matrix (HM) and the Fisher Information Matrix (FIM) are critical for optimisation and statistical inference. Since the HM is related to the FIM, its singularity (non-invertibility) indicates that the log-likelihood function may have flat regions or an insufficient curvature at the given parameter values. This lack of curvature can lead to the FIM, which quantifies the precision of parameter estimates, being singular or near-singular.

The FIM quantifies the amount of information that a random variable stores about an unknown parameter. Let $f(x;c,d)$ and $ln[f(x;c,d)]$ be given by (4) and (14), respectively. If there are suitable partial derivatives of $f(x;c,d)$, then the FIM $I_{i,j}^{c,d}(i,j=1,2)$ is a square 2×2 matrix defined as:

$$I_{i,j}^{c,d} = -\begin{bmatrix} E\left\{\frac{\partial^2 ln[f(x;c,d)]}{dc^2}\right\} & E\left\{\frac{\partial^2 ln[f(x;c,d)]}{\partial c\partial d}\right\} \\ E\left\{\frac{\partial^2 ln[f(x;c,d)]}{\partial d\partial c}\right\} & E\left\{\frac{\partial^2 ln[f(x;c,d)]}{\partial d^2}\right\} \end{bmatrix}, \tag{15}$$

where $I_{1,2}^{c,d} = I_{2,1}^{cd}$, obviously. The R codes of the fimSPCN1 function are presented in Appendix 2.

A non-invertible (singular) HM can lead to the information matrix becoming singular, impacting the optimisation process and parameter estimation. If all second-order partial derivatives of $f(x;c,d)$ exist, then HM $H_{i,j}^{c,d}(i,j=1,2)$ is a square 2×2 matrix arranged as:

$$H_{i,j}^{c,d} = \begin{bmatrix} \frac{\partial^2 f(x;c,d)}{\partial c^2} & \frac{\partial^2 f(x;c,d)}{\partial c\partial d} \\ \frac{\partial^2 f(x;c,d)}{\partial d\partial c} & \frac{\partial^2 f(x;c,d)}{\partial d^2} \end{bmatrix}, \tag{16}$$

where $H_{1,2}^{c,d} = H_{2,1}^{cd}$, obviously. The R codes of the hmSPCN1 function are presented in Appendix 2.

In distribution theory, there are papers with more or less complicated FIM and HM formulas, but it is difficult to find a numerical analysis.

The values of $I_{i,j}^{c,d}\ (i,j=1,2)$ for certain parameter values, including those from Figure 1, are:

$$I_{i,j}^{1,0.2} = \begin{bmatrix} 1.78 & -0.15 \\ -0.15 & 0.61 \end{bmatrix}, I_{i,j}^{1,0.4} = \begin{bmatrix} 1.79 & -0.13 \\ -0.13 & 0.54 \end{bmatrix}, I_{i,j}^{1,0.6} = \begin{bmatrix} 1.78 & -0.07 \\ -0.07 & 0.44 \end{bmatrix}, I_{i,j}^{1,0.8} = \begin{bmatrix} 1.79 & -0.02 \\ -0.02 & 0.35 \end{bmatrix},$$

$$I_{i,j}^{1.5,0.2} = \begin{bmatrix} 0.78 & -0.05 \\ -0.05 & 0.15 \end{bmatrix}, I_{i,j}^{1.5,0.4} = \begin{bmatrix} 0.79 & -0.10 \\ -0.10 & 0.53 \end{bmatrix}, I_{i,j}^{1.5,0.6} = \begin{bmatrix} 0.79 & -0.09 \\ -0.09 & 0.69 \end{bmatrix}, I_{i,j}^{1.5,0.8} = \begin{bmatrix} 0.79 & -0.03 \\ -0.03 & 0.70 \end{bmatrix},$$

$$I_{i,j}^{2,0.2} = \begin{bmatrix} 0.43 & -0.01 \\ -0.01 & 0.10 \end{bmatrix}, I_{i,j}^{2,0.4} = \begin{bmatrix} 0.44 & -0.06 \\ -0.06 & 0.40 \end{bmatrix}, I_{i,j}^{2,0.6} = \begin{bmatrix} 0.45 & -0.08 \\ -0.08 & 0.8 \end{bmatrix}, I_{i,j}^{2,0.8} = \begin{bmatrix} 0.45 & -0.05 \\ -0.05 & 1.09 \end{bmatrix},$$

$$I_{i,j}^{1,1} = \begin{bmatrix} 1.80 & 0.02 \\ 0.02 & 0.27 \end{bmatrix}, I_{i,j}^{1,1.25} = \begin{bmatrix} 1.81 & 0.04 \\ 0.04 & 0.19 \end{bmatrix}, I_{i,j}^{1,1.5} = \begin{bmatrix} 1.82 & 0.05 \\ 0.05 & 0.14 \end{bmatrix}, I_{i,j}^{1,1.75} = \begin{bmatrix} 1.83 & 0.05 \\ 0.05 & 0.10 \end{bmatrix},$$

$$I_{i,j}^{1.25,1} = \begin{bmatrix} 1.15 & 0.02 \\ 0.02 & 0.42 \end{bmatrix}, I_{i,j}^{1.25,1.25} = \begin{bmatrix} 1.16 & 0.05 \\ 0.05 & 0.31 \end{bmatrix}, I_{i,j}^{1.25,1.5} = \begin{bmatrix} 1.17 & 0.06 \\ 0.06 & 0.21 \end{bmatrix}, I_{i,j}^{1.25,1.75} = \begin{bmatrix} 1.17 & 0.06 \\ 0.06 & 0.15 \end{bmatrix},$$

$$I_{i,j}^{1.5,1} = \begin{bmatrix} 0.80 & 0.02 \\ 0.02 & 0.61 \end{bmatrix}, I_{i,j}^{1.5,1.25} = \begin{bmatrix} 0.81 & 0.06 \\ 0.06 & 0.44 \end{bmatrix}, I_{i,j}^{1.5,1.5} = \begin{bmatrix} 0.81 & 0.06 \\ 0.06 & 0.31 \end{bmatrix}, I_{i,j}^{1.5,1.75} = \begin{bmatrix} 0.81 & 0.06 \\ 0.06 & 0.21 \end{bmatrix}.$$

We obtain positive values of $I_{i,j}^{c,d}$ $(i,j=1,2)$ except for values $I_{1,2}^{c,d}=I_{2,1}^{c,d}(d<1)$. If $c=const$ and values of d increase, then values of $I_{2,2}^{c,d}(d\geq 1)$ decrease. If $d=$ $=const$ and values of c increase, then values of $I_{1,1}^{c,d}$ decrease, values of $I_{1,2}^{c,d}=I_{2,1}^{c,d}$ are similar and values of $I_{2,2}^{c,d}(d\geq 1)$ increase.

The values of $H_{i,j}^{c,d}$ $(i,j=1,2)$ for certain parameter values, including those from Figure 1, are:

$$H_{i,j}^{1,0.2}=\begin{bmatrix}-0.19 & -0.17\\ -0.17 & -0\end{bmatrix}, H_{i,j}^{1,0.4}=\begin{bmatrix}-0.15 & -0.13\\ -0.13 & -0\end{bmatrix}, H_{i,j}^{1,0.6}=\begin{bmatrix}-0.14 & -0.10\\ -0.10 & -0\end{bmatrix}, H_{i,j}^{1,0.8}=\begin{bmatrix}-0.14 & -0.08\\ -0.08 & -0\end{bmatrix},$$

$$H_{i,j}^{1.5,0.2}=\begin{bmatrix}0.13 & -0.04\\ -0.04 & 0.03\end{bmatrix}, H_{i,j}^{1.5,0.4}=\begin{bmatrix}0.16 & -0.05\\ -0.05 & 0.02\end{bmatrix}, H_{i,j}^{1.5,0.6}=\begin{bmatrix}0.18 & -0.05\\ -0.05 & 0.02\end{bmatrix}, H_{i,j}^{1.5,0.8}=\begin{bmatrix}0.20 & -0.05\\ -0.05 & 0.02\end{bmatrix},$$

$$H_{i,j}^{2,0.2}=\begin{bmatrix}0.16 & -0.01\\ -0.01 & 0.01\end{bmatrix}, H_{i,j}^{2,0.4}=\begin{bmatrix}0.16 & -0.01\\ -0.01 & 0.01\end{bmatrix}, H_{i,j}^{2,0.6}=\begin{bmatrix}0.17 & -0.02\\ -0.02 & 0.01\end{bmatrix}, H_{i,j}^{2,0.8}=\begin{bmatrix}0.18 & -0.02\\ -0.02 & 0.01\end{bmatrix},$$

$$H_{i,j}^{1,1}=\begin{bmatrix}-0.15 & -0.07\\ -0.07 & -0\end{bmatrix}, H_{i,j}^{1,1.25}=\begin{bmatrix}-0.18 & -0.05\\ -0.05 & -0\end{bmatrix}, H_{i,j}^{1,1.5}=\begin{bmatrix}-0.2 & -0.04\\ -0.04 & -0\end{bmatrix}, H_{i,j}^{1,1.75}=\begin{bmatrix}-0.24 & -0.03\\ -0.03 & -0\end{bmatrix},$$

$$H_{i,j}^{1.25,1}=\begin{bmatrix}0.13 & -0.07\\ -0.07 & 0\end{bmatrix}, H_{i,j}^{1.25,1.25}=\begin{bmatrix}0.14 & -0.06\\ -0.06 & 0\end{bmatrix}, H_{i,j}^{1.25,1.5}=\begin{bmatrix}0.14 & -0.05\\ -0.05 & 0\end{bmatrix}, H_{i,j}^{1.25,1.75}=\begin{bmatrix}0.13 & -0.05\\ -0.05 & 0\end{bmatrix},$$

$$H_{i,j}^{1.5,1}=\begin{bmatrix}0.21 & -0.05\\ -0.05 & 0.01\end{bmatrix}, H_{i,j}^{1.5,1.25}=\begin{bmatrix}0.23 & -0.05\\ -0.05 & 0.01\end{bmatrix}, H_{i,j}^{1.5,1.5}=\begin{bmatrix}0.24 & -0.05\\ -0.05 & 0.01\end{bmatrix}, H_{i,j}^{1.5,1.75}=\begin{bmatrix}0.25 & -0.05\\ -0.05 & 0.01\end{bmatrix}.$$

We get $H_{1,2}^{c,d}=H_{2,1}^{c,d}<0$. Values of $H_{2,2}^{c,d}$ are very close to zero, e.g. $H_{2,2}^{1,0.2}=$ $=-0.0004988119$, $H_{2,2}^{1,0.4}=-0.0009952323$, $H_{2,2}^{1,0.6}=-0.001486889$. If $d=$ $=const$ and values of c increase then values of $H_{1,2}^{c,d}(d<1)$ increase, values of $H_{1,2}^{c,d}=H_{2,1}^{c,d}$ are similar and values of $H_{1,2}^{c,d}=H_{2,1}^{c,d}(d<1)$ and $H_{1,1}^{c,d}(d\geq 1)$ increase.

5. Maximum likelihood estimation

In this section, we present a location-scale SPCN1 distribution characterised by location parameter $\mu\in R$ and scale parameter $\sigma>0$. This distribution is formulated through the $Y=\mu+\sigma X$ transformation:

$$f(y;\mu,\sigma,c,d)=\frac{2c}{\sigma}\left|\frac{y-\mu}{\sigma}\right|^{c-1}\phi\left(\left|\frac{y-\mu}{\sigma}\right|^{c}\right)\Phi\left[sgn\left(d\,\frac{y-\mu}{\sigma}\right)\left|d\,\frac{y-\mu}{\sigma}\right|^{c}\right]. \tag{17}$$

Let $y_1^*,y_2^*,\ldots,y_n^*$ be a random sample of size n from the $SPCN1(\mu,\sigma,c,d)$. Our target is to estimate the unknown μ,σ,c,d parameters. The likelihood function based on (2) is given by:

$$L=\frac{2c}{\sigma}\prod_{i=1}^{n}\left|\frac{y_i^*-\mu}{\sigma}\right|^{c-1}\phi\left(\left|\frac{y_i^*-\mu}{\sigma}\right|^{c}\right)\Phi\left[sgn\left(d\,\frac{y_i^*-\mu}{\sigma}\right)\left|d\,\frac{y_i^*-\mu}{\sigma}\right|^{c}\right], \tag{18}$$

then the log-likelihood function $l = lnL$ is defined as:

$$l = nln\frac{2c}{\sigma} + (c-1)\sum_{i=1}^{n} ln\left|\frac{y_i^*-\mu}{\sigma}\right| + \sum_{i=1}^{n} ln\left[\phi\left(\left|\frac{y_i^*-\mu}{\sigma}\right|^c\right)\right] + \\ + \sum_{i=1}^{n} ln\left\{\Phi\left[sgn\left(d\frac{y_i^*-\mu}{\sigma}\right)\left|d\frac{y_i^*-\mu}{\sigma}\right|^c\right]\right\}. \tag{19}$$

There is no need to present formulas $\frac{dl}{d\mu}, \frac{dl}{d\sigma}, \frac{dl}{dc}, \frac{dl}{dd}$, because they have a complicated form. To simplify the process, we can use one of the advanced computational environments with embedded optimisation procedures. These include Mathcad, Mathematica, Excel or R. For the purpose of this paper, the maximum likelihood estimates (MLEs) of the μ, σ, c, d parameters were calculated in the R environment.

For estimated parameter Θ, the bias (BIAS) of estimator $\hat{\Theta}$ is defined as:

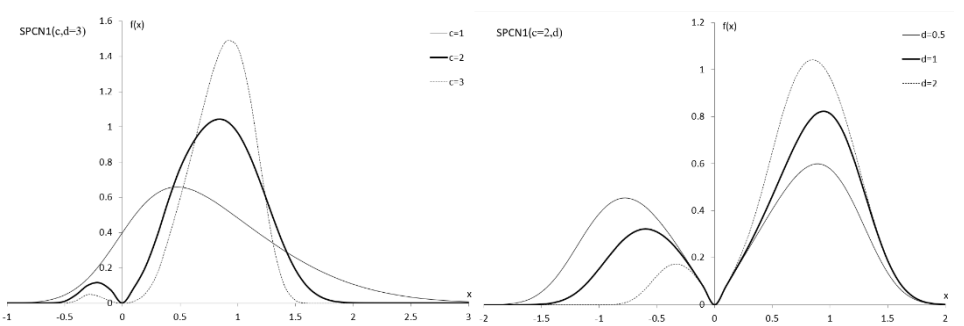
$$BIAS(\hat{\Theta}) = \frac{1}{n}\sum_{i=1}^{n} \hat{\Theta}(y_1^*, y_2^*, \ldots, y_n^*) - \Theta$$

and the root mean squared error (RMSE) is given by:

$$RMSE(\hat{\Theta}) = \sqrt{E\left[\left(\hat{\Theta} - \Theta\right)^2\right]}.$$

These characteristics of the MLEs are shown in Tables 3 and 4. The simulation study was performed with 10^3 samples using sample sizes of $25, 50, 100, 200$. The samples, as shown in Figure 8, were drawn from the $SPCN1(c, 3)$, $c = (1, 2, 3)$ and $SPCN1(2, d)$, $d = (0.5, 1, 2)$. Our MLE analysis is for a unimodal and slightly bimodal distribution (left) as well as a clearly bimodal distribution (right).

Figure 8. PDF curves of the SPCN1 distribution for parameter values used in the MLE



Source: authors' work.

Table 3. Biases and RMSEs of the MLEs from $SPCN1(0, 1, c, 3)$

c	n	$\hat{\mu}$		$\hat{\sigma}$		$\hat{c}$		$\hat{d}$	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
1	25	−0.087	0.028	0.117	0.028	0.202	0.069	0.327	0.125
	50	−0.076	0.019	0.104	0.021	0.154	0.043	0.316	0.119
	100	−0.062	0.011	0.086	0.014	0.105	0.022	0.313	0.116
	200	−0.053	0.007	0.073	0.009	0.079	0.012	0.298	0.108
2	25	−0.065	0.019	0.087	0.018	0.265	0.099	0.279	0.100
	50	−0.052	0.011	0.071	0.011	0.218	0.075	0.270	0.096
	100	−0.038	0.006	0.058	0.007	0.173	0.049	0.266	0.092
	200	−0.024	0.002	0.043	0.003	0.124	0.025	0.238	0.078
3	25	−0.055	0.011	0.066	0.011	0.280	0.105	0.275	0.097
	50	−0.045	0.007	0.058	0.007	0.262	0.095	0.272	0.095
	100	−0.036	0.004	0.048	0.005	0.222	0.071	0.262	0.090
	200	−0.030	0.002	0.041	0.003	0.194	0.055	0.243	0.080

Source: authors' work.

Table 4. Biases and RMSEs of the MLEs from $SPCN1(0, 1, 2, d)$

d	n	$\hat{\mu}$		$\hat{\sigma}$		$\hat{c}$		$\hat{d}$	
		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
0.5	25	−0.023	0.008	0.048	0.005	0.226	0.076	0.172	0.051
	50	−0.017	0.005	0.043	0.003	0.188	0.056	0.143	0.034
	100	−0.013	0.002	0.035	0.002	0.158	0.039	0.121	0.023
	200	−0.007	0.001	0.033	0.002	0.139	0.030	0.101	0.016
1	25	−0.043	0.011	0.068	0.010	0.211	0.070	0.182	0.055
	50	−0.032	0.006	0.058	0.006	0.182	0.053	0.150	0.039
	100	−0.020	0.003	0.047	0.004	0.140	0.033	0.115	0.023
	200	−0.016	0.002	0.041	0.002	0.121	0.024	0.098	0.015
2	25	−0.050	0.012	0.076	0.013	0.242	0.086	0.263	0.096
	50	−0.040	0.007	0.064	0.008	0.206	0.066	0.243	0.084
	100	−0.031	0.004	0.053	0.005	0.162	0.044	0.212	0.068
	200	−0.023	0.002	0.045	0.003	0.127	0.027	0.187	0.052

Source: authors' work.

Tables 3 and 4 summarise the simulation results for the bias and RMSE of the MLEs under different sample sizes and parameter settings. As observed, the estimates converge to the true parameter values as sample size n increases, which confirms the consistency of the proposed estimators.

In both tables, the bias of location parameter $\hat{\mu}$ is slightly negative for all cases, indicating a small systematic underestimation. The lowest bias is obtained for the location parameter, suggesting that it is estimated most accurately among all parameters. For all parameters, both bias and RMSE decrease as the sample size increases.

The estimates of $\hat{d}$ are generally more variable but exhibit the same pattern of convergence as n increases.

A comparison between Tables 3 and 4 reveals that a higher d slightly increases the bias of $\hat{d}$. Overall, the simulation confirms that the maximum likelihood estimators of the SPCN1 parameters are consistent and perform well even for small sample sizes.

6. Application

6.1. Goodness-of-fit tests

Sulewski and Stoltmann (2023) divided alternatives into nine groups according to their skewness (γ_1) and excess kurtosis ($\bar{\gamma}_2$) signs. Groups O-H are defined in Table 5. Our proposal belongs to all analysed groups except group C.

Table 6 presents parameter vectors $\theta = (0, \sigma, c, d)$ together with the corresponding values of the mean (μ_a), standard deviation (σ_a), skewness (γ_1), excess kurtosis ($\bar{\gamma}_2$) and similarity measure $M(\theta;\ \mu, \sigma)$ for selected configurations. As similarity measure M increases from 0.5 to 0.9, the parameters (σ, c, d) change smoothly, leading to different distributional shapes. The SPCN distribution is capable of producing light- and heavy-tailed, symmetric and asymmetric, as well as both unimodal and bimodal forms. Figure 9 illustrates these shapes graphically. The transition between the groups demonstrates that the SPCN family provides a coherent parametric framework for controlling skewness and kurtosis independently, while maintaining analytical tractability. These results confirm that SPCN is a highly flexible model encompassing empirical data patterns encountered in practice. We observe both unimodal and bimodal shapes.

Table 5. Groups of alternatives with signs of γ_1 and $\bar{\gamma}_2$

Group	γ_1	$\bar{\gamma}_2$
O	zero	zero
A	positive	positive
B	negative	positive
C	zero	positive
D	zero	negative
E	positive	negative
F	negative	negative
G	positive	zero
H	negative	zero

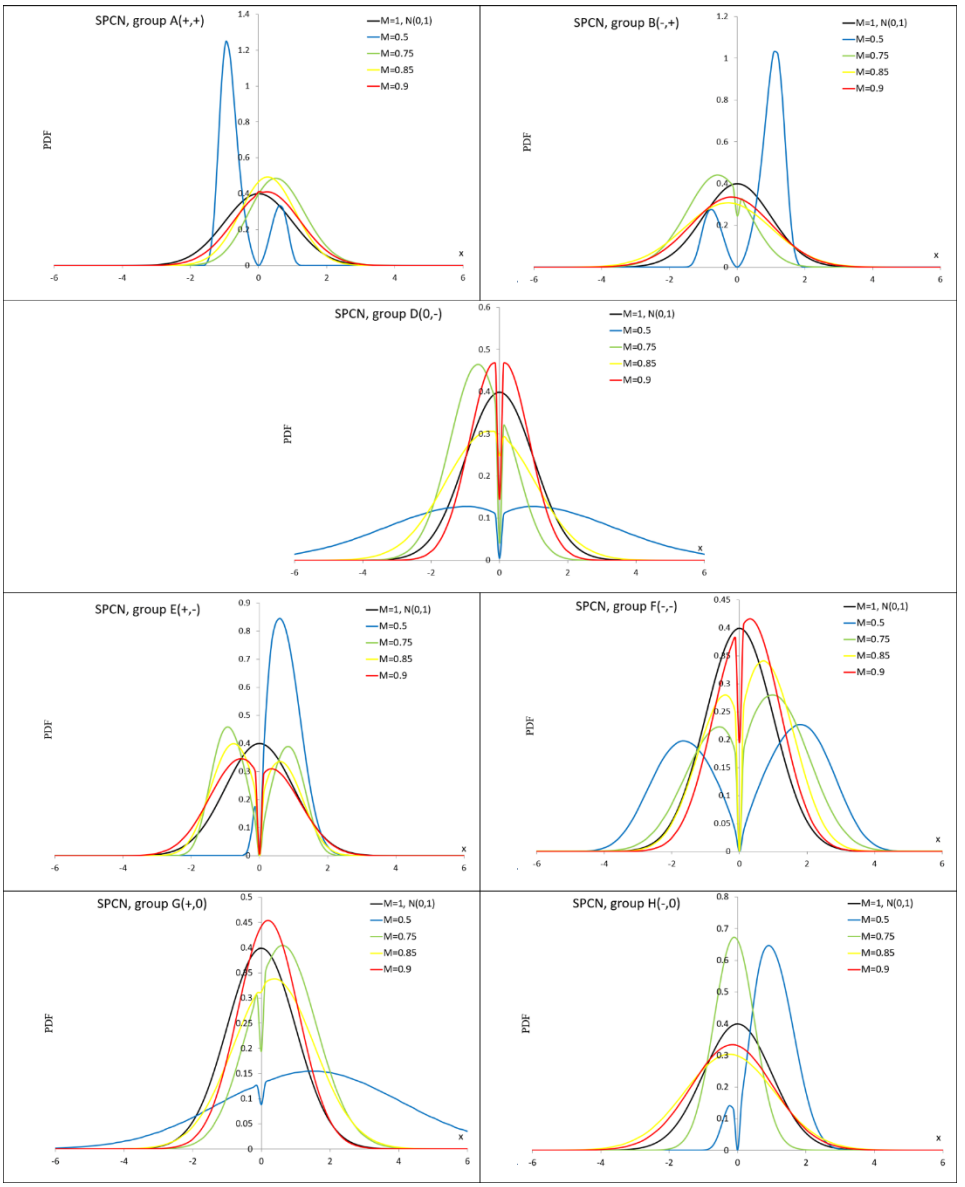
Source: authors' work.

Table 6. Vectors of SPCN parameter $\boldsymbol{\theta}$, mean μ, standard deviation σ, skewness γ_1, excess kurtosis $\bar{\gamma}_2$ and similarity measure M. Groups O-B, D-H

Group	$\boldsymbol{\theta} = (0, \sigma, c, d)$	μ	σ	$\gamma_{1,}$	$\bar{\gamma}_2$	$M(\boldsymbol{\theta}; \mu, \sigma)$
O	0,1,1,0	0	1	0	0	$M(\boldsymbol{\theta}; 0,1) = 1$
A	0,0.96,2.682, −1.141	−0.578	0.634	1.158	0.071	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.001,1,1	0.565	0.827	0.137	0.062	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,0.855,1,0.445	0.277	0.809	0.017	0.004	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,1.005,0.999,0.325	0.248	0.975	0.008	0.005	$M(\boldsymbol{\theta}; 0,1) = 0.9$
B	0,1.164,2.71,1.144	0.704	0.766	−1.172	0.1	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.102,1.012, −0.975	−0.612	0.911	−0.107	0.023	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.322,1, −0.277	−0.282	1.292	−0.005	0.001	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,1.202,1, −0.181	−0.171	1.189	−0.001	0.001	$M(\boldsymbol{\theta}; 0,1) = 0.9$
D	0,2.94,1.101,0	0	2.849	0	−0.343	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.061,1.069, −0.952	−0.573	0.865	0	−0.126	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.325,1.006, −0.267	−0.27	1.294	0	−0.023	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,0.817,1.038,0	0	0.806	0	−0.141	$M(\boldsymbol{\theta}; 0,1) = 0.9$
E	0,0.948,1.372,4.472	0.746	0.458	0.328	−0.1	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.123,1.77, −0.351	−0.122	1.001	0.13	−1.272	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.099,1.373, −0.32	−0.164	1.001	0.117	−0.874	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,1.156,1.126, −0.246	−0.179	1.099	0.05	−0.402	$M(\boldsymbol{\theta}; 0,1) = 0.9$
F	0,2.21,1.741,0.314	0.204	1.976	−0.108	−1.26	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.546,1.237,0.413	0.368	1.408	−0.13	−0.621	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.236,1.181,0.378	0.284	1.141	−0.1	−0.514	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,0.981,1.023,0.33	0.238	0.944	−0.01	−0.082	$M(\boldsymbol{\theta}; 0,1) = 0.9$
G	0,3.028,1.012,0.833	1.539	2.593	0.07	0	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,1.189,1.017,0.93	0.643	0.992	0.085	0	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.238,1.001,0.409	0.374	1.18	0.013	0	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,0.902,1,0.291	0.201	0.879	0.005	0	$M(\boldsymbol{\theta}; 0,1) = 0.9$
H	0,1.293,1.487,2.851	0.997	0.631	−0.055	0	$M(\boldsymbol{\theta}; 0,1) = 0.5$
	0,0.602,1, −0.213	−0.1	0.594	−0.002	0	$M(\boldsymbol{\theta}; 0,1) = 0.75$
	0,1.338,1, −0.229	−0.238	1.317	−0.002	0	$M(\boldsymbol{\theta}; 0,1) = 0.85$
	0,1.206,1, −0.168	−0.159	1.196	−0.001	0	$M(\boldsymbol{\theta}; 0,1) = 0.9$

Source: authors' work.

Figure 9. PDF curves of the SPCN1 distribution for parameter values presented in Table 6



Source: authors' work.

6.2. Real data example

In this Section, we present two real data examples to demonstrate the flexibility and applicability of the SPCN1 distribution. A total of ten distributions were involved in Monte Carlo simulations.

The models selected for comparison with the SPCN1 are: ESGN3, SGN, GMNSN, FGSN3, SN1, FSCN, BABSN, FASN, SBNN, and SSGN. The PDFs of the used models are shown in Appendix 1.

The estimation of the model parameters is carried out using the maximum likelihood method. To avoid local maxima of the logarithmic likelihood function, the optimisation process is run 100 times with several different initial values widely scattered in the parameter space. AIC, BIC and HQIC were used for model comparisons. Let us recall that

$$AIC = -2l + 2p, BIC = -2l + pln(n), HQIC = -2l + 2pln(ln(n)), \qquad (20)$$

where l is the log-likelihood function, n is the sample size and p is the number of model parameters. Tables 7–8 display the values of the MLEs, the information criteria (AIC, BIC and HQIC) for the analysed models. The lowest values of the information criteria are marked in bold, indicating the best-fitting model according to that criterion. It can be observed that different models perform better depending on the dataset, highlighting the flexibility and suitability of certain distributions for capturing the characteristics of the data, but the SPCN1 model achieves the best results. Plots of the estimated PDF of the analysed models are given in Figures 10 and 11. Overall, these results allow for a comprehensive assessment of model adequacy and can guide the selection of the most appropriate distribution for further statistical analysis.

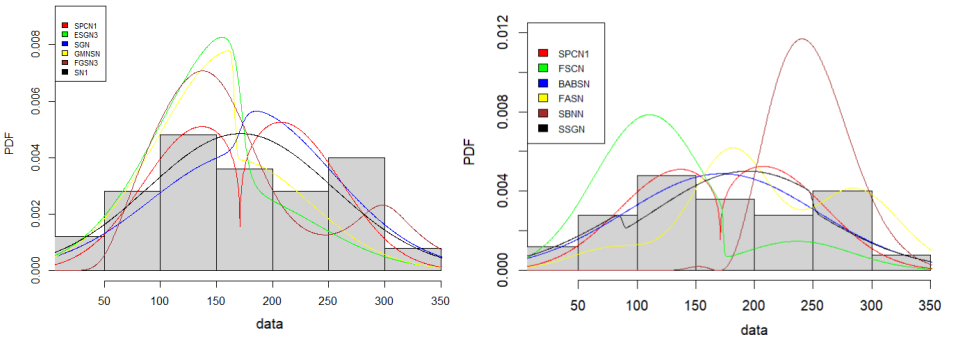
Example 1. The first dataset contains data on arrests per 100,000 residents for assault in each of the 50 US states in 1973 $(n = 50)$ (see R codes USArrests[2]). The descriptive statistics are: $\mu_a \approx 170.76, \sigma_a \approx 83.338, \gamma_1 \approx 0.227, \bar{\gamma}_2 \approx 1.931.$

Table 7. MLEs, AIC, BIC and HQIC (first dataset)

Model	Estimated parameters of the given model					AIC	BIC	HQIC
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{d}$	$\hat{e}$			
SPCN1	170.885	73.377	1.204	0.077		**590.689**	**598.337**	**593.601**
ESGN3	172.436	70.182	38.819	–6.098	56.707	606.892	616.452	610.532
SGN	171.026	80.465	2.235	96.356		594.334	601.982	597.246
GMNSN	165.992	67.932	–0.498	39.027		598.795	606.443	601.707
FGSN3	173.533	77.009	–1.708	0.876		603.324	610.972	606.236
SN1	171.515	82.248	41.258	–11.863		591.205	598.853	594.118
FSCN	173.546	46.995	–1.338	–47.931		608.326	615.974	611.239
BABSN	171.948	81.713	69.507	5.661		591.194	598.842	594.106
FASN	174.431	51.598	6.471	0.384		608.138	615.786	611.051
SBNN	169.436	50.301	–37.110			621.637	627.373	623.821
SSGN	168.524	83.882	46.199	0.429	62.820	596.266	605.826	599.906

Source: authors' work.

Figure 10. Estimated PDF of the analysed distributions, first dataset



Source: authors' work.

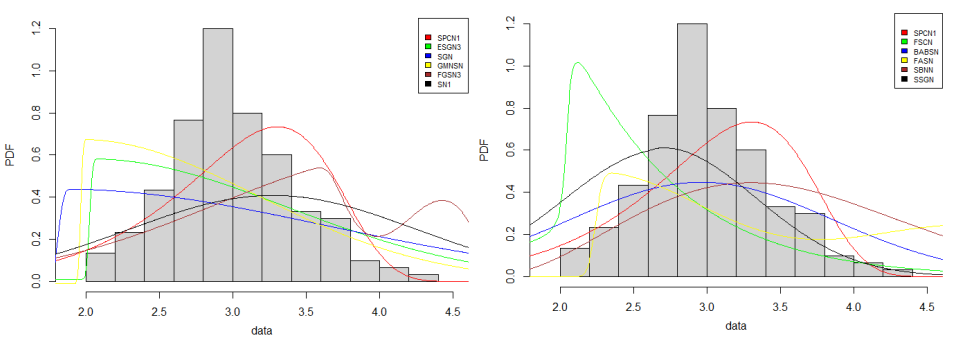
Example 2. The second set of data are measurements in centimetres of the variable sepal width for 50 flowers from each of the 3 species of iris ($n = 150$). The species are Iris setosa, versicolor, and virginica (see R codes iris[2]). The descriptive statistics are: $\mu_a \approx 3.057, \sigma_a \approx 0.436, \gamma_1 \approx 0.316, \bar{\gamma}_2 \approx 3.181$.

Table 8. MLEs, AIC, BIC and HQIC (second dataset)

Model	Estimated parameters					AIC	BIC	HQIC
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{d}$	$\hat{e}$			
SPCN1	−0.606	3.971	5.979	16.328		**223.498**	**235.541**	**228.391**
ESGN3	2.026	1.346	52.699	86.264	88.325	281.479	296.533	287.595
SGN	1.814	1.817	66.218	92.863		333.708	345.750	338.600
GMNSN	1.965	1.201	−71.349	98.800		269.996	282.039	274.889
FGSN3	4.212	1.314	3.213	−37.129		308.097	320.139	312.989
SN1	3.273	0.977	98.154	−26.668		313.642	325.684	318.534
FSCN	2.048	2.904	3.854	58.423		383.613	395.655	388.505
BABSN	2.962	0.891	48.565	27.686	139.243	286.486	301.540	292.602
FASN	2.241	1.207	93.347	30.114		419.210	431.253	424.103
SBNN	1.436	1.317	−23.335			298.112	307.144	301.781
SSGN	2.638	0.660	0.339	0.747	62.115	273.228	288.281	279.343

Source: authors' work.

Figure 11. Estimated PDF of the analysed distributions, second dataset



Source: authors' work.

7. Conclusions

In this paper, we propose bimodal distributions that can be used as an alternative to the other bimodal distributions in modelling bimodal-distributed data, including compound normal and Laplace distributions. The characterisation of the skew plasticising component normal I distribution is investigated. Simulation examples showed that such estimation procedures performed well. Our proposal is a very interesting alternative distribution for goodness-of-fit tests. Real data examples demonstrate that the skew plasticising component normal I distribution is a flexible, parsimonious and competitive model that deserves to be added to the existing distributions in modelling unimodal- (see example II) and bimodal-distributed data (see example I).

References

Alavi, M. R. (2012). On a New Bimodal Normal Family. *Journal of Statistical Research of Iran*, *8*(2), 163–176. https://doi.org/10.18869/acadpub.jsri.8.2.163.

Alavi, S. M. R., & Tarhani, M. (2017). On a Skew Bimodal Normal-Normal distribution fitted to the Old-Faithful geyser data. *Communications in Statistics – Theory and Methods*, *46*(15), 7301–7312. https://doi.org/10.1080/03610926.2016.1148731.

Arellano-Valle, R. B., & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *99*(7), 1362–1382. https://doi.org/10.1016/j.jmva.2008.01.020.

Arellano-Valle, R. B., Cortés, M. A., & Gómez, H. W. (2010). An extension of the epsilon-skew-normal distribution. *Communications in Statistics – Theory and Methods*, *39*(5), 912–922. https://doi.org/10.1080/03610920902807903.

Arellano-Valle, R. B., Gómez, H. W., & Quintana, F. A. (2004). A new class of skew-normal distributions. *Communications in Statistics – Theory and Methods*, *33*(7), 1465–1480. https://doi.org/10.1081/STA-120037254.

Arnold, B., Beaver, R. J., Azzalini, A., Balakrishnan, N., Bhaumik, A., Dey, D. K., Cuadras, C. M., Sarabia, J. M., Arnold, B. C., & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, *11*, 7–54. https://doi.org/10.1007/BF02595728.

Athayde, E., Azevedo, C., Leiva, V., & Sanhueza, A. (2012). About Birnbaum-Saunders distributions based on the Johnson system. *Communications in Statistics – Theory and Methods*, *41*(11), 2061–2079. http://dx.doi.org/10.1080/03610926.2010.551454.

Azzalini, A. (1985). A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics*, *12*(2), 171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, *46*(2), 199–208. https://doi.org/10.6092/issn.1973-2201/711.

Azzalini, A., & Chiogna, M. (2004). Some results on the stress–strength model for skew-normal variates. *METRON*, *62*(3), 315–326.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*(4), 715–726. https://doi.org/10.1093/biomet/83.4.715.

Bahrami, W., Agahi, H., & Rangin, H. (2009). A Two-parameter Balakrishnan Skew-normal Distribution. *Journal of Statistical Research of Iran*, *6*(2), 231–242. http://dx.doi.org/10.18869/acadpub.jsri.6.2.231.

Bahrami, W., & Qasemi, E. (2015). A Flexible Skew-Generalized Normal Distribution. *Journal of Statistical Research of Iran JSRI*, *11*(2), 131–145. https://doi.org/10.18869/acadpub.jsri.11.2.131.

Behboodian, J. (1970). On the modes of a mixture of two normal distributions. *Technometrics*, *12*(1), 131–139. https://doi.org/10.2307/1267357.

Birnbaum, Z. W., & Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, *6*(2), 319–327. https://doi.org/10.2307/3212003.

Branco, M. D., & Dey, D. K. (2001). A General Class of Multivariate Skew – Elliptical Distributions. *Journal of Multivariate Analysis*, *79*(1), 99–113. https://doi.org/10.1006/jmva.2000.1960.

Chhikara, R. S., & Folks, J. L. (1977). The inverse Gaussian distribution as a lifetime model. *Technometrics*, *19*(4), 461–468. https://doi.org/10.2307/1267886.

Choudhury, K., & Matin, M. A. (2011). Extended skew generalized normal distribution. *METRON*, *69*(3), 265–278. https://doi.org/10.1007/BF03263561.

Das, J., Pathak, D., Hazarika, P. J., Chakraborty, S., & Hamedani, G. G. (2023). A New Flexible Alpha Skew-normal Distribution. *Journal of the Indian Society for Probability and Statistics*, *24*(2), 485–507. https://doi.org/10.1007/s41096-023-00163-8.

Durrans, S. R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research*, *28*(6), 1649–1655. https://doi.org/10.1029/92WR00554.

Elal-Olivero, D. (2010). Alpha-skew-normal distribution. *Proyecciones. Journal of Mathematics*, *29*(3), 224–240. http://dx.doi.org/10.4067/S0716-09172010000300006.

Elal-Olivero, D., Gómez, H. W., & Quintana, F. A. (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, *139*(4), 1484–1492. https://doi.org/10.1016/j.jspi.2008.07.016.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer. https://doi.org/10.1007/978-0-387-35768-3.

Gaddum, J. H. (1945). Lognormal distributions. *Nature*, *156*, 463–466. http://dx.doi.org/10.1038/156463a0.

Gómez, H. W., Elal-Olivero, D., Salinas, H. S., & Bolfarine, H. (2011). Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics*, *22*(1), 50–62. https://doi.org/10.1002/env.1026.

Gómez, H. W., Salinas, H. S., & Bolfarine, H. (2006). Generalized skew-normal models: properties and inference. *Statistics. A Journal of Theoretical and Applied Statistics*, *40*(6), 495–505. https://doi.org/10.1080/02331880600723168.

Gómez, H. W., Varela, H., & Vidal, I. (2013). A new class of skew-symmetric distributions and related families. *Statistics. A Journal of Theoretical and Applied Statistics*, *47*(2), 411–421. https://doi.org/10.1080/02331888.2011.589904.

Gupta, R. C., & Gupta, R. D. (2004). Generalized skew-normal model. *Test*, *13*(2), 501–524. https://doi.org/10.1007/BF02595784.

Gupta, R. D., & Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, *17*(1), 197–210. https://doi.org/10.1007/s11749-006-0030-x.

Handam, A. H. (2012). A note on generalized alpha-skew-normal distribution. *International Journal of Pure and Applied Mathematics*, *74*(4), 491–496. https://www.ijpam.eu/contents/2012-74-4/6/6.pdf.

Hazarika, P. J., Shah, S., & Chakraborty, S. (2020). Balakrishnan Alpha Skew Normal Distribution: Properties and Applications. *Malaysian Journal of Science*, *39*(2), 71–91. https://doi.org/10.48550/arXiv.1906.07424.

Henze, N. (1986). A Probabilistic Representation of the 'Skew-normal' Distribution. *Scandinavian Journal of Statistics*, *13*(4), 271–275.

Jamalizadeh, A., & Arabpour, A. R. N. (2011). A generalized skew two-piece skew-normal distribution. *Statistical Papers*, *52*(2), 431–446. https://doi.org/10.1007/s00362-009-0240-x.

Jamalizadeh, A., & Balakrishnan, N. (2008). On order statistics from bivariate skew-normal and skew-t_ν distributions. *Journal of Statistical Planning and Inference*, *138*(12), 4187–4197. https://doi.org/10.1016/j.jspi.2008.03.035.

Johnson, N. L. (1949). System of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176. https://doi.org/10.2307/2332539.

Kapteyn, J. C. (1916). Skew frequency curves in Biology and Statistics. *Recueil des travaux botaniques néerlandais*, *13*(2), 105–157. https://natuurtijdschriften.nl/pub/552499/RTBN1916013002002.pdf.

Kim, H. J. (2005). On a class of two-piece skew-normal distributions. *Statistics*, *39*(6), 537–553. https://doi.org/10.1080/02331880500366027.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694.

Kumar, C. S., & Anusree, M. R. (2011). On a generalized mixture of standard normal and skew normal distributions. *Statistics & Probability Letters*, *81*(12), 1813–1821. https://doi.org/10.1016/j.spl.2011.07.009.

Kumar, C. S., & Anusree, M. R. (2013). A generalized two-piece skew normal distribution and some of its properties. *Statistics*, *47*(6), 1370–1380. https://doi.org/10.1080/02331888.2012.697269.

Kumar, C. S., & Anusree, M. R. (2015). On modified generalized skew normal distribution and some of its properties. *Journal of Statistical Theory and Practice*, *9*(3), 489–505. https://doi.org/10.1080/15598608.2014.935617.

Lin, T. I., Lee, J. C., & Yen., S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, *17*(3), 909–927.

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statistics & Probability Letters*, *54*(4), 381–387. https://doi.org/10.1016/S0167-7152(01)00103-1.

Ma, Y., & Genton, M. G. (2004). Flexible Class of Skew-Symmetric Distributions. *Scandinavian Journal of Statistics*, *31*(3), 459–468. https://doi.org/10.1111/j.1467-9469.2004.03_007.x.

Magnus, G., & Magnus, J. R. (2019). The estimation of normal mixtures with latent variables. *Communications in Statistics – Theory and Methods*, *48*(5), 1255–1269. https://doi.org/10.1080/03610926.2018.1429625.

Mahmoudi, E., Jafari, H., & Meshkat, R. S. (2019). Alpha-Skew Generalized Normal Distribution and its Applications. *Applications and Applied Mathematics: An International Journal (AAM)*, *14*(2), 784–804. https://digitalcommons.pvamu.edu/aam/vol14/iss2/10.

Malakhov, A. N. (1978). *A cumulant analysis of random non-Gaussian processes and their transformations*. Soviet Radio.

Martínez-Flórez, G., Bolfarine, H., & Gómez, H. W. (2014). Skew-normal alpha-power model. *Statistics*, *48*(6), 1414–1428. https://doi.org/10.1080/02331888.2013.826659.

Martínez-Flórez, G., Tovar-Falón, R., & Elal-Olivero, D. (2022). Some new flexible classes of normal distribution for fitting multimodal data. *Statistics*, *56*(1), 182–205. http://dx.doi.org/10.1080/02331888.2022.2041642.

Mudholkar, G. S., & Hutson, A. D. (2000). The epsilon skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, *83*(1), 291–309. https://doi.org/10.1016/S0378-3758(99)00096-8.

Popović, B. V., Cordeiro, G., Ortega, E. M., & Pascoa, M. A. R. (2017). A new extended mixture normal distribution. *Mathematical Communications*, *22*(1), 53–73. https://www.mathos.unios.hr/mc/index.php/mc/article/view/1409.

Rasekhi, M., Hamedani, G. G., & Chinipardaz, R. (2017). A flexible extension of skew generalized normal distribution. *Metron*, *75*(1), 87–107. https://doi.org/10.1007/s40300-017-0106-2.

Rieck, J. R., & Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, *33*(1), 51–60. https://doi.org/10.1080/00401706.1991.10484769.

Salinas, H. S., Arellano-Valle, R. B., & Gómez, H. W. (2007). The extended skew-exponential power distribution and its derivation. *Communications in Statistics – Theory and Methods*, *36*(9), 1673–1689. https://doi.org/10.1080/03610920601126118.

Salinas, H., Bakouch, H., Qarmalah, N., & Martínez-Flórez, G. (2023). A Flexible Class of Two-Piece Normal Distribution with a Regression Illustration to Biaxial Fatigue Data. *Mathematics*, *11*(5), 1–14. https://doi.org/10.3390/math11051271.

Seijas-Macias, A., Oliveira, A., & Oliveira, T. (2017). The Presence of Distortions in the Extended Skew-normal Distribution. In *Proceedings of The International Statistical Institute Regional Statistics Conference 2017 "Enhancing Statistics, Prospering Human Life". Bali, 20–24 March 2017* (pp. 840–846). Bank Indonesia, International Statistical Institute Regional Statistics Conference. https://isi-web.org/sites/default/files/2025-04/Proceeding-International-Statistic-Institute.pdf.

Shafiei, S., Doostparast, M., & Jamalizadeh, A. (2016). The alpha–beta skew-normal distribution: Properties and applications. *Statistics*, *50*(2), 338–349. http://dx.doi.org/10.1080/02331888.2015.1096938.

Shah, S., Hazarika, P. J., Chakraborty, S., & Ali, M. M. (2021). The Balakrishnan-Alpha-Beta-Skew-Normal Distribution: Properties and Applications. *Pakistan Journal of Statistics and Operation Research*, *17*(2), 367–380. http://dx.doi.org/10.18187/pjsor.v17i2.3731.

Shah, S., Hazarika, P. J., Chakraborty, S., & Ali, M. M. (2023). A generalized-alpha–beta-skew-normal distribution with applications. *Annals of Data Science*, *10*(4), 1127–1155. https://doi.org/10.1007/s40745-021-00325-0.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

Sharafi, M., & Behboodian, J. (2008). The Balakrishnan skew–normal density. *Statistical Papers*, *49*(4), 769–778. https://doi.org/10.1007/s00362-006-0038-z.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *62*(4), 795–809.

Sulewski, P. (2021). DS normal distribution: properties and applications. *Lobachevskii Journal of Mathematics*, *42*(12), 2980–2999. http://dx.doi.org/10.1134/S1995080221120337.

Sulewski, P. (2022a). Modified Lilliefors goodness-of-fit test for normality. *Communications in Statistics – Simulation and Computation*, *51*(3), 1199–1219. https://doi.org/10.1080/03610918.2019.1664580.

Sulewski, P. (2022b). Normal distribution with plasticizing component. *Communications in Statistics – Theory and Methods*, *51*(11), 3806–3835. https://doi.org/10.1080/03610926.2020.1837881.

Sulewski, P. (2023). New Members of The Johnson Family of Probability Distributions: Properties and Application. *REVSTAT-Statistical Journal*, *21*(4), 535–556. https://doi.org/10.57805/revstat.v21i4.429.

Sulewski, P., & Stoltmann, D. (2023). Modified Cramer-von Mises goodness-of-fit test for normality. *Przegląd Statystyczny. Statistical Review*, *70*(4), 1–36. https://doi.org/10.59139/ps.2023.04.1.

Sulewski, P., & Volodin, A (2022). Sulewski Plasticizing Component Distribution: Properties and Applications. *Lobachevskii Journal of Mathematics*, *43*(8), 2286–2300. http://dx.doi.org/10.1134/S1995080222110270.

Wang, Z., & Song, J. (2017). Equivalent linearization method using Gaussian mixture (GM-ELM) for nonlinear random vibration analysis. *Structural Safety*, *64*, 9–19. https://doi.org/10.1016/j.strusafe.2016.08.005.

Appendix 1

Table A1. PDFs of distributions from Groups 3–7

Group no.	PDFs
3	$f_{TPSN}(x;\alpha)=\frac{2\pi}{\pi+2tan^{-1}(\alpha)}\phi(x)\Phi(\alpha\lvert x\rvert)\ (\alpha\in R).$
3	$f_{GSN1}(x;\alpha,\delta)=\begin{cases}2\phi\left(\frac{x}{1+\delta}\right)\left[\frac{\delta}{1+\delta}+\frac{1-\delta}{1+\delta}\Phi\left(\frac{\alpha x}{1+\delta}\right)\right] & x<0\\ 2\phi\left(\frac{x}{1-\delta}\right)\Phi\left(\frac{\alpha x}{1-\delta}\right) & x\geq 0\end{cases}\ (\alpha\in R,\delta\in[0,1)),$
3	$f_{EESN}(x;\alpha,\delta)=\begin{cases}2\left[\frac{\delta}{1+\delta}+\frac{1-\delta}{1+\delta}\Phi\left(\frac{\alpha x}{1+\delta}\right)\right]\phi\left(\frac{x}{1+\delta}\right), & x<0\\ 2\Phi\left(\frac{\alpha x}{1-\delta}\right)\phi\left(\frac{x}{1-\delta}\right), & x\geq 0\end{cases}\ (\alpha\in \mathrm{R},\delta\in[0,1)),$
3	$f_{ESN}(x;\varepsilon)=\phi\left(\frac{x}{1+\varepsilon}\right)I(x<0)+\phi\left(\frac{x}{1-\varepsilon}\right)I(x\geq 0)\ (\lvert\varepsilon\rvert<1),$
3	$f_{FESN}(x;\alpha,\varepsilon)=\frac{1}{2-2\Phi(\delta)}\begin{cases}\phi\left(\frac{x}{1+\varepsilon}-\alpha\right) & x<0\\ \phi\left(\frac{x}{1-\varepsilon}+\alpha\right) & x\geq 0\end{cases}(\alpha\in \mathrm{R},\lvert\varepsilon\rvert<1),$
3	$f_{STPSN}(x;\alpha,\beta)=\frac{4\pi}{\pi+2tan^{-1}(\beta)}\phi(x)\Phi(\alpha x)\Phi(\beta\lvert x\rvert)\ (\alpha,\beta\in R),$
3	$f_{GTPSN}(x;\alpha,\beta,\varepsilon)=\frac{2\pi\phi(x)\Phi_2(\alpha\lvert x\rvert,\beta\lvert x\rvert;\varepsilon)}{cos^{-1}\left(\frac{-\varepsilon-\alpha\beta}{\sqrt{1+\alpha^2}\sqrt{1+\beta^2}}\right)+tan^{-1}(\alpha)+tan^{-1}(\beta)}\ (\alpha,\beta\in R,\lvert\varepsilon\rvert<1),$ where $\Phi_2(\alpha\lvert x\rvert,\beta\lvert x\rvert;\varepsilon)$ denotes the CDF of $N_2(0,0,1,1,\varepsilon)$,
3	$f_{GSTPSN}(x;\alpha,\beta,\varepsilon)=c(\alpha,\beta,\rho)\phi(x)\Phi_2(\alpha x,\beta\lvert x\rvert;\varepsilon)\ (\alpha,\beta\in R,\lvert\varepsilon\rvert<1),$ where $\Phi_2(\lambda_1 x,\lambda_2\lvert x\rvert;\varepsilon)$ denotes the CDF of $N_2(0,0,1,1,\rho)$ and $c(\lambda_1,\lambda_2,\varepsilon)=\frac{4\pi}{cos^{-1}\left(\frac{-\varepsilon-\alpha\beta}{\sqrt{1+\alpha^2}\sqrt{1+\beta^2}}\right)+cos^{-1}\left(\frac{-\varepsilon+\alpha\beta}{\sqrt{1+\alpha^2}\sqrt{1+\beta^2}}\right)+2tan^{-1}(\beta)},$
3	$f_{GTPSN}(x;\alpha,\varepsilon)=\frac{2\pi\phi(x)}{\pi+tan^{-1}(\alpha)+tan^{-1}(\alpha\varepsilon)}\begin{cases}\Phi(\alpha x) & x<0\\ \Phi(\alpha\varepsilon x) & x\geq 0\end{cases}\ (\alpha\in R,\lvert\varepsilon\rvert<1),$
3	$f_{TPPN}(x,\sigma_1,\sigma_2,c)=\begin{cases}\frac{c}{\sigma_1\sqrt{2\pi}}\cdot\left(\frac{-x}{\sigma_1}\right)^{c-1}exp\left[-\frac{1}{2}\left(\frac{-x}{\sigma_1}\right)^{2c}\right] & x<0\\ 0\ x=0 & \\ \frac{c}{\sigma_2\sqrt{2\pi}}\cdot\left(\frac{x}{\sigma_2}\right)^{c-1}exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_2}\right)^{2c}\right] & x>0\end{cases}(\sigma_1,\sigma_2>0,c\geq 1).$
4	$f_{BN}(x;\lambda)=\left(\frac{1+\alpha\lambda}{1+\lambda}\right)\phi(x)\ (\lambda\geq 0),$
4	$f_{ASN}(x;\alpha)=\frac{(1-\alpha x)^2+1}{2+\alpha^2}\phi(x)\ (\alpha\in \mathrm{R}),$
4	$f_{DN}(x;\gamma)=\frac{\sqrt{\pi}\lvert x\rvert^{\gamma}}{\Gamma(\gamma+0.5)2^{\gamma}}\phi(x)\ (\gamma\geq 0),$
4	$f_{GASN}(x;\alpha,n)=\frac{(1-\alpha x)^{2n}+1}{2+\sum_{i=1}^{n}\binom{2n}{2i}\alpha^{2i}\prod_{j=1}^{i}(2j-1)}\phi(x)\ (\alpha\in \mathrm{R},n\in \mathrm{N}-\{0\}),$
4	$f_{BASN}(x;\alpha)=\frac{\left[(1-\alpha x)^2+1\right]^2}{3\alpha^4+8\alpha^2+4}\phi(x)\ (\alpha\in \mathrm{R}),$
4	$f_{TN}(x;\gamma)=\exp(-0.5\gamma^2)cosh(\gamma x)\phi(x)\ (\gamma>0),$
4	$f_{ABSN}(x;\alpha,\beta)=\frac{\left(1-\alpha x-\beta x^3\right)^2+1}{\alpha^2+15\beta^2+6\alpha\beta+2}\phi(x)\ (\alpha,\beta\in \mathrm{R}),$

Table A1. PDFs of distributions from Groups 3–7 (cont.)

Group no.	PDFs
4	$f_{BABSN}(x;\alpha,\beta)=\frac{\left[\left(1-\alpha x-\beta x^3\right)^2+1\right]^2}{c(\alpha,\beta)}\phi(x)\ (\alpha,\beta\in \mathrm{R}),$ where $c(\alpha,\beta)=3\alpha^4+8\alpha^2+4+60\alpha^3\beta+12\alpha\beta(4+315\beta^2)+630\alpha^2\beta^2+15\beta^2(8+693\beta^2).$
4	$f_{FAN}(x;\gamma)=\frac{2+0.5\gamma\left[\left(x^2-1\right)^2+2\right]}{1+\gamma}\phi(x)\ (\gamma\geq 0).$
5	$f_{GN}(x;\gamma)=\frac{1}{2\gamma^{1/\gamma}\Gamma(1/\gamma)}exp\left(-\frac{\lvert x\rvert^\gamma}{\gamma}\right)(\gamma>0),$
5	$f_{BGN}(x;\gamma)=\frac{\gamma^{(\gamma-3)/\gamma}}{2\Gamma(3/\gamma)}x^2exp\left(-\frac{\lvert x\rvert^\gamma}{\gamma}\right)(\gamma>0),$
5	$f_{ASGN}(x;\gamma,\omega)=\frac{\gamma^{1-1/\gamma}\left[(1-\omega x)^2+1\right]}{2\left[\omega^2\gamma^{2/\gamma}\Gamma(3/\gamma)+2\Gamma(1/\gamma)\right]}exp\left(-\frac{\lvert x\rvert^\gamma}{\gamma}\right)(\gamma,\omega>0).$
6	$f_{PN}(x;\alpha)=\alpha\phi(x)[\Phi(x)]^{\alpha-1}(\alpha>0),$
6	$f_{GPN}(x;\alpha,\lambda)=\mathrm{k}(\alpha,\lambda)\phi(x)[\Phi(\lambda x)]^{\alpha-1}\ (\alpha>0,\lambda\in R),$
7	$f_{PSAN}(x;\alpha,\lambda)=\alpha\phi_\lambda(x)[\Phi_\lambda(x)]^{\alpha-1}(\alpha>0,\lambda\in R),$ where $\phi_\lambda(x)=2\phi(x)\Phi(x)$ and $\Phi_\lambda(x)=\int_{-\infty}^{x}\phi_\lambda(t)dt,$
7	$f_{SN1}(x;\lambda_0,\lambda_1)=\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)^{-1}\phi(x)\Phi(\lambda_0+\lambda_1x)(\lambda_0,\lambda_1\in R),$
7	$f_{SCN}(x;\lambda)=2\phi(x)\Phi\left(\frac{\lambda x}{\sqrt{1+\lambda^2x^2}}\right)(\lambda\in R),$
7	$f_{SGN}(x;\lambda_1,\lambda_2)=2\phi(x)\Phi\left(\frac{\lambda_1x}{\sqrt{1+\lambda_2x^2}}\right)(\lambda_1\in R,\lambda_2\geq 0),$
7	$f_{FGSN3}(x;\lambda_1,\lambda_2)=2\phi(x)\Phi(\lambda_1x+\lambda_2x^3)\ (\lambda_1,\lambda_2\in R),$
7	$f_{BSN}(x;\lambda,\mathrm{n})=\frac{\phi(x)\Phi(\lambda x)^n}{b_n(\lambda)}\ (\lambda\in R,n\geq 1),$ where $b_n(\lambda)=E[\Phi(\lambda U)^n], U\sim N(0,1)$. For $n=1,2,3$ we have closed form for $b_n(\lambda)$, i.e. $b_1(\lambda)=\frac{1}{2},b_2(\lambda)=\frac{1}{4}+\frac{1}{2\pi}sin^{-1}\left(\frac{\lambda^2}{1+\lambda^2}\right),b_3(\lambda)=\frac{1}{8}+\frac{3}{4\pi}sin^{-1}\left(\frac{\lambda^2}{1+\lambda^2}\right).$
7	$f_{GSN2}(x;\lambda_1,\lambda_2,\rho)=\frac{2\pi\phi(x)\Phi_2(\lambda_1x,\lambda_2x;\rho)}{cos^{-1}\left(\frac{-\rho-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\right)},\lambda_1,\lambda_2\in R,\lvert\rho\rvert<1,$ where $\Phi_2(\lambda_1x,\lambda_2x;\rho)$ denotes the CDF of $N_2(0,0,1,1,\rho),$
7	$f_{TPBSN}(x;\lambda_1,\lambda_2)=\frac{1}{c_{n,m}(\lambda_1,\lambda_2)}\phi(x)[\Phi(\lambda_1x)]^n[\Phi(\lambda_2x)]^m\ (\lambda_1,\lambda_2\in R),$ where $c_{n,m}(\lambda_1,\lambda_2)=E\{[\Phi(\lambda_1U)]^n[\Phi(\lambda_2U)]^m\}, U\sim N(0,1),$
7	$f_{GSN}(x;\lambda_1,\lambda_2,\rho)=\frac{2\pi\phi(x)\Phi_2(\lambda_1x,\lambda_2x;\rho)}{cos^{-1}\left(\frac{-\rho-\lambda_1\lambda_2}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\right)}\ (\lambda_1,\lambda_2\in R,\lvert\rho\rvert<1),$
7	$f_{SBN}(x;\alpha,\lambda)=2\left(\frac{1+\alpha x^2}{1+\alpha}\right)\phi(x)\Phi(\lambda x)\ (\alpha\geq 0,\lambda\in \mathrm{R}),$
7	$f_{SFN}(x;\theta,\lambda)=\frac{\phi(\lvert x\rvert+\theta)\Phi(\lambda x)}{1-\Phi(\theta)}\ (\theta,\lambda\in R),$
7	$f_{ESGN1}(x;\lambda_1,\lambda_2,\lambda_3)=2\phi(x)\Phi\left(\frac{\lambda_1x}{\sqrt{\lambda_2x^2+\lambda_3x^4}}\right)(\lambda_1\in R,\lambda_2,\lambda_3>0),$
7	$f_{ESGN2}(x;\lambda_1,\lambda_2,\lambda_3)=2\phi(x)\Phi\left(\frac{\lambda_1x}{\sqrt{1+\lambda_2x^2+\lambda_3x^4}}\right)(\lambda_1\in R,\lambda_2,\lambda_3>0),$

Table A1. PDFs of distributions from Groups 3–7 (cont.)

Group no.	PDFs
7	$f_{GMNSN}(x;\alpha,\lambda)=\frac{2}{\alpha+2}\phi(x)[1+\alpha\Phi(\lambda x)]\ (\alpha>-2,\lambda\in R),$
7	$f_{NSN}(x;\alpha,\beta)=\left[\frac{1}{2}-\frac{1}{\pi}tan^{-1}\left(\frac{\beta}{\sqrt{1+\alpha^2(1+\beta^2)}}\right)\right]^{-1}\phi(x)\Phi_\beta(\alpha x)(\alpha,\beta\in R),$ where $\Phi_\beta(\alpha x)=2\int_{-\infty}^{\alpha x}\phi(y)(\beta y)dy,$
7	$f_{FSGN}(x;\theta,\lambda_1,\lambda_2)=\frac{\phi(\lvert x\rvert+\theta)}{1-\Phi(\theta)}\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}\right)\ (\theta,\lambda_1\in R,\lambda_2>0),$
7	$f_{FSCN}(x;\lambda,\lambda_1)=\frac{\phi(\lvert x\rvert+\lambda)}{1-\Phi(\lambda)}\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_1^2 x^2}}\right)\ (\lambda,\lambda_1\in R),$
7	$f_{ESGN3}(x;\alpha,\lambda_1,\lambda_2)=\frac{2}{\alpha+2}\phi(x)\left[1+\alpha\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}\right)\right]\ (\alpha\geq 1,\lambda_1\in R,\lambda_2\geq 0),$
7	$f_{SSGN}(x;\alpha,\lambda_1,\lambda_2)=2\phi(x)\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2\lvert x\rvert^{2\alpha}}}\right)(\alpha\neq 0,\lambda_1\in R,\lambda_2>0),$
7	$f_{SBNN}(x;\lambda)=2x^2\phi(x)\Phi(\lambda x),\lambda\in R,$
7	$f_{BABSN}(x;\alpha,\beta)=\phi(x)\frac{\Phi(\beta\sqrt{\alpha^2+1}+\alpha x)}{\Phi(\beta)}(\alpha,\beta\in \mathrm{R}),$
7	$f_{GABSN}(x;\alpha,\beta,\lambda)=\frac{\left(1-\alpha x-\beta x^3\right)^2+1}{c(\alpha,\beta,\lambda)}\phi(x)\Phi(\lambda x)\ (\alpha,\beta,\lambda\in \mathrm{R}),$ where $c(\alpha,\beta,\lambda)=1+3\alpha\beta-\alpha\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}-\beta\sqrt{\frac{2}{\pi}}\frac{\lambda(3+2\lambda^2)}{(1+\lambda^2)^{1.5}}+\frac{\alpha^2}{2}+\frac{15\beta^2}{2},$
7	$f_{FASN}(x;\alpha,\lambda)=\frac{2+0.5\alpha\left[\left(x^2-1\right)^2+2\right]}{1+\alpha}\phi(x)\Phi(\lambda x)\ (\alpha\geq 0,\lambda\in \mathrm{R}).$

Note: Functions $\phi(x)$ and $\Phi(x)$ are the PDF and CDF of the $N(0,1)$, respectively.
Source: authors' work.

Appendix 2

The following code contains R codes for the PDF, CDF, quantile, mode, k-th order moment, skewness, kurtosis, pdf of order statistics, moments of order statistics and pseudo-random number generator which is also available at github.com/PiotrSule/SPCN1.

```
library(RelDists)
library(zipfR)
library(pracma)
library(flexsurv)
library(xlsx)
library(ggamma)
library(gsl)
library(PSDistr)
library(Deriv)
library(splines)
```

```
#normalization condition
norm_cond<-function(c,d) {
 esgn1 <- function(x) dSPCN1(x,c,d)
 return (as.numeric(integrate(Vectorize(esgn1), lower = -Inf, upper = Inf)[1]))}

#CDF
library(PSDistr)
dSPCN1 <- function(x,c,d){
 return(2*dpc(x,0,1,c)*ppc(x*d,0,1,c))
}

#PDF
pSPCN1 <- function(x,c,d){
 return(integral(function(x) dSPCN1(x,c,d), -100, x, reltol = 1e-12, method =
"Simpson"))
}

#quantile
qSPCN1=function(p,c,d){
 u11 = function(x,c,d) pSPCN1(x,c,d)-p
 return(uniroot(u11, c(-5,5), tol = 0.0000000001, f.lower = -5, c=c, d=d)$root)
}

#generator
rSPCN1 =function(n,c,d) {
 x=numeric(n)
 for (i in 1:n) x[i]=qSPCN1(runif(1,0,1),c,d)
 return(sort(x))
}

#ordinary moments
mSPCN1=function(k,c,d) {
 return(integral(function(x) x^k*dSPCN1(x,c,d), -Inf, Inf, reltol = 1e-12, method =
"Simpson"))
}

#skewness
g1SPCN1=function(c,d){
 w1=mSPCN1(3,c,d)-3*mSPCN1(1,c,d)*mSPCN1(2,c,d)+2*mSPCN1(1,c,d)^3
```

```
 w2=mSPCN1(2,c,d)-mSPCN1(1,c,d)^2
 return(w1/w2^(1.5))
}

#kurtosis
g2SPCN1=function(c,d){
 w1=mSPCN1(4,c,d)-4*mSPCN1(1,c,d)*mSPCN1(3,c,d)+6*mSPCN1(1,c,d)^2*
 mSPCN1(2,c,d)-3*mSPCN1(1,c,d)^4
 w2=mSPCN1(2,c,d)-mSPCN1(1,c,d)^2
 return(w1/w2^2)
}

# PDF of order statistics
 dOSSPCN1=function(x,i,n,c,d) {
 return(fact(n)/fact(i-1)/fact(n-i)*dSPCN1(x,c,d)*pSPCN1(x,c,d)^(i-1)
 *(1-pSPCN1(x,c,d))^(n-i))
 }
# moments of order statistics
mOSSPCN1=function(k,i,n,c,d) {
 return(integral(function(x) x^k*dOSSPCN1(x,i,n,c,d), -Inf, Inf, reltol = 1e-12,
method = "Simpson"))
}

# Shannon entropy
sSPCN1=function(c,d){
 return(integral(function(x) -dSPCN1(x,c,d)*log(dSPCN1(x,c,d)), -Inf, Inf, reltol =
1e-12, method = "Simpson"))
}

I11 <- function(x, c, d){
 eval(Deriv(Deriv(expression(n*log(2*c)+(c-
1)*log(abs(x))+log(dnorm(abs(x)^c,0,1))+log(porm(sign(d*x),0,1)*abs(x*d)^c)),'c'),'
d'))
}
I12 <- function(x, c, d){
 eval(Deriv(Deriv(expression(n*log(2*c)+(c-
1)*log(abs(x))+log(dnorm(abs(x)^c,0,1))+log(porm(sign(d*x),0,1)*abs(x*d)^c)),'c'),'
d'))
}
```

```
I21 <- function(x, c, d) return(I12(x,c,d))
I22 <- function(x, c, d){
 eval(Deriv(Deriv(expression(n*log(2*c)+(c-
1)*log(abs(x))+log(dnorm(abs(x)^c,0,1))+log(porm(sign(d*x),0,1)*abs(x*d)^c)),'d'),
'd'))
}

# Fisher Information Matrix
fimSPCN1=function(c,d,xg){
 FIM=numeric(4)
 FIM[1]=-integral(function(x) I11(x,c,d)*dSPCN1(x,c,d), -xg, xg, reltol = 1e-9,
method = "Kronrod")
 FIM[2]=-integral(function(x) I12(x,c,d)*dSPCN1(x,c,d), -xg, xg, reltol = 1e-9,
method = "Kronrod")
 FIM[3]=FIM[2]
 FIM[4]=-integral(function(x) I22(x,c,d)*dSPCN1(x,c,d), -xg, xg, reltol = 1e-9,
method = "Kronrod")
 return(FIM)
}

#Hessian Matrix
hmSPCN1=function(c,d){
 HM=numeric(4)
 HM[1]=eval(Deriv(Deriv(expression(2*c*abs(x)^(c-1)*
 dnorm(abs(x)^c,0,1)*pnorm(sign(d*x)*abs(x*d)^c,0,1)),'c'),'c'))
 HM[2]=eval(Deriv(Deriv(expression(2*c*abs(x)^(c-1)*
 dnorm(abs(x)^c,0,1)*pnorm(sign(d*x)*abs(x*d)^c,0,1)),'c'),'d'))
 HM[3]=HM[2]
 HM[4]=eval(Deriv(Deriv(expression(2*c*abs(x)^(c-1)*
 dnorm(abs(x)^c,0,1)*pnorm(sign(d*x)*abs(x*d)^c,0,1)),'d'),'d'))
 return(HM)
}
```

Przegląd Statystyczny. Statistical Review, 2025, vol. 72, 2, 33–53 https://doi.org/10.59139/ps.2025.02.2
Received: 4.04.2025, revised: 20.10.2025, accepted: 21.10.2025

Does the deposit interest rate stimulate savings in West Africa? An application of dynamic-panel data analysis

Abdurrauf Babalola,[a] Abdulazeez Vatsa Attahiru[b]

Abstract. Saving is a crucial step towards investing. Several factors influence the decision to save, including religion, economic conditions, cultural attitudes and the interest rate on savings accounts. This study investigates low savings rates in West Africa and tries to find out whether higher interest rate would stimulate regional saving. The lack of empirical understanding of how interest rate affects saving behaviour challenges economic development. Using the life-cycle hypothesis, the study analysed savings as a function of a deposit interest rate, per capita income and the inflation rate through panel autoregressive distributed lag analysis. The findings show that deposit interest rate does not significantly impact savings in the region, in contrast to per capita income and inflation. The conclusion of the study is that the relationship between interest rates and savings is complex and influenced by multiple factors other than the deposit interest rate. The study suggests implementing policies that promote long-term investment strategies beyond relying on interest rates, which helps balance immediate investments with savings and encourages firms to set aside funds for future use.

Keywords: interest rate, dynamic panel, savings, West Africa

JEL: C23, E21, E43, O16

1. Introduction

Savings are vital for the economic development of West African countries, providing a basis for investment, consumption and financial stability. Several factors influence the decision to save, including religion, economic conditions, cultural attitudes and interest rates (deposit rates) on savings accounts. Interest rates significantly affect the saving behaviour. Higher rates generally encourage more savings due to attractive returns, while lower rates may lead consumers to spend instead. This relationship can impact a country's overall savings rate and, subsequently, its economic growth.

In West Africa, people's responses to changes in interest rates are influenced by factors like inflation, economic situation and financial literacy (World Bank, 2018), while Babalola and Abdul (2022) suggested that the financial security situation, the

[a] Al-Hikmah University, Faculty of Management Sciences, Department of Economics, Adeta Road, P.M.B. 1601, Ilorin, Nigeria, e-mail: abdclement@yahoo.com, ORCID: https://orcid.org/0000-0001-8389-6639.

[b] Al-Hikmah University, Faculty of Management Sciences, Department of Economics, Adeta Road, P.M.B. 1601, Ilorin, Nigeria, abdulazeezvasta1@gmail.com, ORCID: https://orcid.org/0009-0009-5253-5414.

access to the saved funds and the possibility of quick cash transfers, among other factors, are key to savings. High interest rates may not encourage savings during periods of high inflation, and cultural practices, such as traditional saving methods and informal savings clubs, can sometimes outweigh the influence of banks. Policymakers need to consider these cultural factors to effectively increase savings rates.

Several economic sectors are operated in the region, e.g. agriculture, mining, manufacturing and services (African Development Bank Group, 2021). At the same time, issues like poverty, income inequality and limited access to financial services remain significant challenges (World Bank, 2018). Understanding this economic landscape is essential for addressing attitudes towards savings and the effect of interest rates.

The rates at which savings accounts earn interest play a crucial role in shaping saving habits and are determined by the central banks' monetary policy choices and factors such as inflation, fluctuations of the currency exchange rates, availability of funds and the economic data.

The savings buildup in West Africa affects investment and economic growth (African Development Bank Group, 2021). Whether it is caused by appealing interest rates or by some other factors, it can offer a secure funding source for investing. More investments can lead to a boost in economic growth, employment opportunities and a decrease in poverty levels. Hence, it is essential to examine how interest rates affect savings to comprehend their possible effects on the region's economic growth and advancement.

In West Africa, several factors enhance savings among individuals and communities. A key motivator is the desire for financial security, as many save for emergencies, healthcare, education or small business investments. This need drives people to set aside funds whenever possible. Access to formal financial institutions is vital in increasing savings rates. When banks and financial services are available and people understand the benefits of saving, a culture of saving develops. Financial literacy initiatives can help individuals see the advantages of formal accounts over informal methods of savings. Cultural attitudes play a role as well. Informal savings groups, like rotating savings and credit associations, provide social support and encourage collective saving, often being more trusted than formal banking systems in areas where the trust in banks is low.

This study explores low savings levels in West Africa and, as stated above, tries to find out how interest rates (more specifically, deposit interest rates) influence the saving behaviour. The region's diverse economic conditions, varying financial literacy and cultural attitudes all complicate policy development. Fluctuating Treasury bill rates and inflation deter individuals from saving, while limited financial literacy and

access to savings products promote a consumption-oriented mindset. Therefore, the objective of this study is to determine if higher interest rates effectively increase savings in the West African region.

2. Literature review

2.1. Conceptual review

Interest rates are a fundamental economic tool, serving as a crucial indicator of financial health and the direction of monetary policy. The Central Bank of Nigeria (CBN) defines interest rates as the cost of borrowing funds or the return on deposited funds, expressed as a percentage of the principal amount. The CBN emphasises that the rates are influenced by monetary policy decisions, inflation expectations, and the broader economic environment, highlighting their dual role in facilitating economic growth and maintaining stability (CBN, 2023).

The Central Bank of West African States (French: Banque Centrale des États de l'Afrique de l'Ouest, BCEAO) sees interest rates as a compensation for the use of capital, reflecting the associated risks and opportunity costs of lending and borrowing. The BCEAO adjusts the rates according to the economic conditions to maintain financial stability and promote investment within the West African Economic and Monetary Union (Diop & Diaw, 2023).

The Bank of Ghana describes interest rates as the cost of money, highlighting their impact on savings, investment and consumption, all of which are vital for the economic development (Bank of Ghana, 2022).

The Central Bank of Liberia, on the other hand, defines interest rates as a compensation that borrowers pay to lenders, which can significantly affect consumer spending and business investment vital for growth (Central Bank of Liberia, 2023).

Finally, according to the Bank of Sierra Leone, interest rates indicate the cost of financing and are influenced by inflation, monetary policy and the demand for credit (African Development Bank Group, 2025).

In summary, as observed by Babalola (2021), Babalola and Abdul (2022) and Babalola et al. (2023), interest rates represent the costs of loans or the returns for providing capital, assuming various forms such as deposit (savings) interest rate, fixed/time interest rate and treasury bill rates. Using the concept of most central banks in the region, there are two categories of interest rates, namely the lending interest rate, which is the cost of borrowing or obtaining a loan, and the deposit interest rate, which is the reward for depositing funds as savings. This study employed the deposit interest rate as a measure of interest rate since it affects savings to the largest extent in most regions of the world.

Savings have been defined by scholars from various perspectives, referring both to individual and aggregate levels. Mankiw (2014) defines savings as the portion of disposable income that is not consumed but instead set aside for future use. This definition focuses on individual behaviour and emphasises the act of withholding a part of income for future purposes. Embracing a macroeconomic standpoint, Keynes (1936) introduced the concept of aggregate savings, which refers to the sum of individual savings within an economy. Keynes viewed savings as the difference between income and consumption, highlighting its role in determining the level of investment and overall economic activity.

Building on Keynes' perspective, Feldstein (1974) defined national savings as the sum of private and government savings, where government savings represent the difference between government revenue and expenditure. This broader definition acknowledges the impact of both private and public sectors on the overall savings within an economy.

Several years later, Blanchard et al. (2017) came up with a similar definition of savings. Their focus was on finding out how various sectors influence overall savings in an economy.

2.2. Theoretical review

The classical theory of savings is a fundamental economic concept that explains how saving behaviour is affected by interest rates, income and consumption preferences. At the core of this theory is the belief that interest rates serve as critical incentives for individuals to save. When interest rates are higher, individuals are more likely to increase their savings, as they can earn a greater return on their deposited funds. Conversely, lower interest rates may discourage savings, nudging individuals toward immediate consumption. This relationship between interest rates and the saving behaviour underscores the importance of monetary policy in shaping economic conditions (Gertler & Kiyotaki, 2010). The above theory refers to the deposit interest rate rather than the lending interest rate or the official interest rate, which is the policy rate in most economies.

The life-cycle theory of savings posits that individuals save throughout their lifetime to smooth consumption patterns and maintain their desired standard of living (Modigliani & Brumberg, 1954). This theory suggests that individuals anticipate changes in income and expenses over their lifetime and adjust their savings accordingly. During their younger years, individuals typically save a smaller portion of their income to support consumption during retirement. As they age and approach retirement, their savings usually increase to ensure a comfortable post-work life. The life-cycle theory

emphasises the importance of long-term planning and intertemporal consumption decisions in shaping the saving behaviour.

The permanent income theory states that people base their savings decisions on their long-term average income rather than their current income (Friedman, 1957). According to this theory, individuals view temporary fluctuations in income as transitory and adjust their savings accordingly. They save a portion of their income to maintain a stable level of consumption over time, even when faced with temporary changes in income. The permanent income theory indicates that individuals prioritise stable consumption patterns and use savings as a tool to smooth income fluctuations.

2.3. Empirical review

Muntanga (2020) researched how interest rates affected savings and investment in Zambia from 1980 to 2018. The study used simple linear regression techniques. The research results indicate that interest rates had a significant effect on net savings in Zambia, among other variables. However, the study's result was based on a simple analysis which is static, meaning that this work analysed the instant effect, but in reality, it takes some time before a variable responds to change, i.e. there is a lag period.

Obeh and Brotoboh (2021) conducted a study on how the interest rate spread affected savings in Nigeria from 1981 to 2019. Multiple regression analysis and the Johansen co-integration test were used to validate the long-term connection in the model. Their findings indicate that the interest rate spread had no significant effect on savings. The issue with this work, however, is that they ran the analysis through two eras of interest rate policies, i.e. before and after the liberalisation of the interest rates in 1986, and their analysis did not capture any structural break period. This liberalisation made the interest rate to float according to the market forces of demand and supply.

Babalola and Abdul (2022) analysed the use of interest rates to encourage saving, which is a popular topic. In this connection, earlier authors expressed their opposition to the practice of charging interest. The persistently low interest rates in Nigeria might fail to motivate developing countries, as they grapple with financial exclusion in the face of interest-based competitors. This situation highlights the pressing need for innovative financial solutions that empower all market participants. The above-mentioned research investigated if interest rates boosted savings in Nigeria from 1987 to 2021, using vector autoregressive/error correction methods to examine the data and draw statistical conclusions. The findings indicate that the deposit rate in Nigeria had no significant impact on savings in that country. The latter are, on the other hand, significantly affected by the treasury bill rates, among other variables.

Loaba (2022) analysed the impact of the use of mobile banking services on saving behaviour in West Africa. Using the Global Findex Database 2017 and jointly estimating a multinomial logit and probit models, the author found that the use of mobile banking services increased the likelihood of formal and informal saving by 2.4% and 0.83%, respectively. Women were more likely to engage in informal savings, but the likelihood of them starting formal savings increased when they used mobile banking services.

Obi (2022) examined how interest rates influenced savings and investment in Nigeria from 1981 to 2020. The research utilised modern econometric methods like cointegration and error correction mechanisms to identify the permanent link between the selected variables. Using the monetary policy rate to represent interest, he found that the interest rate greatly influenced savings and investment in Nigeria. However, the monetary policy rate he used was not ideally matched to the purpose of the research (a deposit interest rate would be more suitable here).

The study by Umoru and Tedunjaiye (2023) investigated the impact of interest rate volatility and exchange rate devaluation on aggregate savings within the Economic Community of West African States (ECOWAS) region. Using a panel-group-means (PMG) estimator and GARCH/ARCH (1,1) models, the authors showed that interest rate volatility had a significant impact on savings in some countries of the region, while in some other it did not. Their aggregate results demonstrated that, in the short run, interest rate movement does not have a significant impact on savings, unlike in the long run.

The study by Muse (2024) explored the impact of different interest rate regimes on savings in Nigeria. Using co-integration regression and a VAR-based impulse-response model, it analysed pre- and post-regime changes. The author found that the prime lending rate had a marginally negative effect on savings, and the policy of the liberalisation of interest rates did not significantly impact savings. However, the type of interest rate regime did moderate the relationship between interest rates and the savings volume in Nigeria.

Idi and Jabil (2024) studied the effects of interest rates on savings and investment in Nigeria from 1980 to 2023. Using econometric techniques such as cointegration and error correction, they found that interest rates significantly impacted both savings and investment in Nigeria. They used the monetary policy rate in their research (which does not directly affect savings) instead of the deposit interest rate. This study used a single interest rate to examine its influence on savings and investment, which might not be the most relevant choice of measurement.

Fundji (2024) studied the influence of interest rates on savings growth in the whole of Africa (apart from North Africa) between 2009 and 2021, using fully modified ordinary least squares. The results confirmed a statistically significant

impact of interest rates on savings growth across all countries. Fundji's (2024) method proved effective since the variables were stationary of the same order; however, the deposit interest rate that is most relevant to savings could have been employed for a more adequate representation. Besides, the claim that interest rates benefit savings growth across all income levels is too general and should consider the distinct economic contexts of high-income versus low-income countries.

2.4. Research gap

The relationship between interest rates and savings behaviour is complex, yet there is a notable research gap in this area in relation to West Africa. Most studies focus on individual countries or developed economies, neglecting the unique factors in the region. Furthermore, examining how financial literacy and varying employment rates affect savings behaviour in response to changes in interest rates (deposit interest rate) could help fill another important gap in the existing body of research. The work of Umoru and Tedunjaiye (2023) that investigated the impact of interest rate volatility and exchange rate devaluation on aggregate savings within the ECOWAS region could have been exhaustive, but the authors used real interest rates to work out the interest rate volatility instead of the deposit interest rate. In contrast, our study uses the latter, which is the most suitable type of interest rate for this kind of research.

3. Methodology

3.1. Theoretical framework

The life-cycle hypothesis (LCH) is a key framework developed by Franco Modigliani and Richard Brumberg in the 1950s that explains how interest rates influence savings behaviour. It asserts that individuals plan their consumption and savings over their lifetime, saving during their working years for retirement. A crucial aspect of the LCH is intertemporal choice, where higher interest rates incentivise saving by increasing the opportunity cost of current consumption, while lower rates encourage immediate spending. The LCH also highlights the impact of the expectations about future income and interest rates on savings-related decisions. In West Africa, the LCH offers insights into how interest rate fluctuations affect saving behaviour, which is vital for policymakers seeking to create effective financial tools that foster savings and enhance economic stability. Overall, it serves as a valuable framework for analysing savings behaviour and informing policies that promote regional economic resilience.

The hypothesis of the study is as follows:

H_0: Interest rate does not significantly impact savings in West Africa

3.2. Model specification

This section introduces the designated model addressing the impact of interest rates on savings in West Africa. The volume of savings is adopted as the dependent variable and the deposit interest rate assumes the role of the focus variable, while the *per capita* income and inflation rate are the control variables. In this research, the ordinary least squares (OLS) technique of multiple regression analysis will be employed to estimate the model. The model proposed by Babalola and Abdul (2022) was adapted to the needs of this study. Below is the outline of Babalola and Abdul's study:

$$GDS = f(DR, TBR, INS, INF, CBB). \tag{1}$$

The model is thus modified below as:

$$SAV = f(DR, PCI, INF). \tag{2}$$

Expressing equation (2) econometrically, we have:

$$SAV_{it} = \beta_0 + \beta_1 DR_{it} + \beta_2 LPCI_{it} + \beta_3 INF_{it} + \mu_{it}, \tag{3}$$

where

SAV_{it} is the savings rate for country i at time t (proxied by total savings as a % of GDP),

$LPCI_{it}$ is the natural log of *per capita* income for country i at time t. The natural log is taken to equalise the variable,

DR_{it} is the deposit rate for country i at time t,

INF_{it} is the inflation rate for country i at time t,

μ_{it} is the stochastic error term, which assumes a constant variance and normal distribution.

Where β_0 (is constant), β_1–β_4 are the parameters of variables for estimation. The subscript i (i = 1… N) represents the nation i in our sample, N is equal to 8, while t (t = 1… T) specifies the period (year). The study examines eight nations throughout 39 years, so there are more years (T) than nations (N). The study population is therefore T x N = 312 observations.

A priori Expectations and Measurement of Variables are as follows:

- **SAV (Savings)**

The savings rate is measured as the total volume of savings expressed as a percentage of GDP. This captures the overall saving behaviour of an economy relative to its size.

- **DR (Deposit Rate)**

Measurement: The deposit rate is measured as the interest rate offered by banks on deposits, reflecting the returns customers receive for their savings.

A priori expectation: A higher deposit rate is expected to positively influence the savings rate (SAV_{it}), as it incentivises individuals to save more due to better returns ($\beta_1 > 0$).

- **PCI (*Per Capita* Income)**

Measurement: The *per capita* income is the average income per individual in an economy. Here it is measured as the GDP divided by the population of the country. It is expressed in constant 2015 USD. This facilitates a fair comparison between countries.

A priori expectation: The PCI is expected to positively affect the savings rate (SAV_{it}), as higher income would encourage saving ($\beta_2 < 0$).

- **INF (Inflation rate)**

Measurement: The inflation rate is measured as a percentage change in the consumer price index.

A priori expectation: The inflation rate is expected to negatively affect the savings rate (SAV_{it}), as higher inflation rates may discourage saving in favour of consumption and investment/ buying capital goods ($\beta_3 < 0$).

3.3. Sources of data

We used secondary data from time series with yearly frequency sourced from the World Development Indicators (WDI, World Bank Group, 2025). Our aim was to examine the impact of interest rates on the saving behaviour in West Africa by analysing eight countries, four Anglophone and four Francophone ones, presented in Table 1. Covering the period from 1986 to 2023, the study consists of 312 observations, capturing significant economic events and trends in the region.

Table 1. The selected West African countries

S/N	Anglophones	S/N	Francophones
1	The Gambia	1	Guinea
2	Ghana	2	Cote D'Ivoire
3	Nigeria	3	Mali
4	Sierra Leone	4	Niger

Source: authors' work.

3.4. Estimation technique

We employed a dynamic panel data analysis (Panel Auto-Regressive Dynamic Lag) after we carried out the usual pre-estimation tests such as descriptive statistics, correlation matrix and stationarity, using Eviews 9 of IHS Global (2016).

Equation (4) below represents a general panel ARDL(p, q_1, q_2, q_3) specification for the functional model outlined in Equation (1). Equation (5) presents the error-correction model (ECM) reparameterisation, which is useful for analysing short-run versus long-run effects.

According to Bismans and Damette (2025), Pesaran and Shin (2003) and Arellano (2003), the econometric model specification for the technique has the form presented below.

General Panel ARDL(p, q_1, q_2, q_3) is specified as:

$$SAV_{it} = \alpha_i + \sum_{k=1}^{p} \phi_{i,k}\, SAV_{i,t-k} + \sum_{k=0}^{q1} \beta_{i,k}\, DR_{i,t-k} + \sum_{k=0}^{q2} \gamma_{i,k}\, LPCI_{t-k} + \sum_{i=0}^{q3} \delta_{i,k}\, INF_{i,t-k} + + \mu_{it}, \quad (4)$$

where αi are individual (country) fixed effects or can include time effects.
Lags p, q1, q2, and q3 can differ by variable, or be set equal for simplicity.

ECM (Error-Correction Model) form – panel ARDL reparameterisation

A convenient reparameterisation isolates short-run dynamics and the long-run equilibrium. For compactness, write an ARDL(k,k,k,k) ECM; the general case is analogous.

$$\Delta SAV_{it} = \mu i + \lambda i\big(SAV_{i,t-k} - \phi_{1,i,}, iDR_{i,t-k} - \phi_{2,i,} LPCI_{i,t-k} - \phi_{3,i,} INF_{i,t-k}\big) + \sum_{j=0}^{s1} \pi_{1,i,j}\, \Delta DR_{i,t-j} + \sum_{j=0}^{s2} \pi_{2,i,k}\, \Delta LPCI_{i,t-j} + \sum_{j=0}^{s3} \pi_{3,i,j}\, \Delta INF_{i,t-j} + + uit, \quad (5)$$

where λi is the speed-of-adjustment coefficient (expected < 0 if disequilibria correct toward long run).
The long-run coefficients are $\phi_{1,i,}$, $\phi_{2,i,}$, $\phi_{3,i,}$.
Short-run dynamics are captured by the π-coefficients on first differences.

4. Research results

4.1. Descriptive statistics

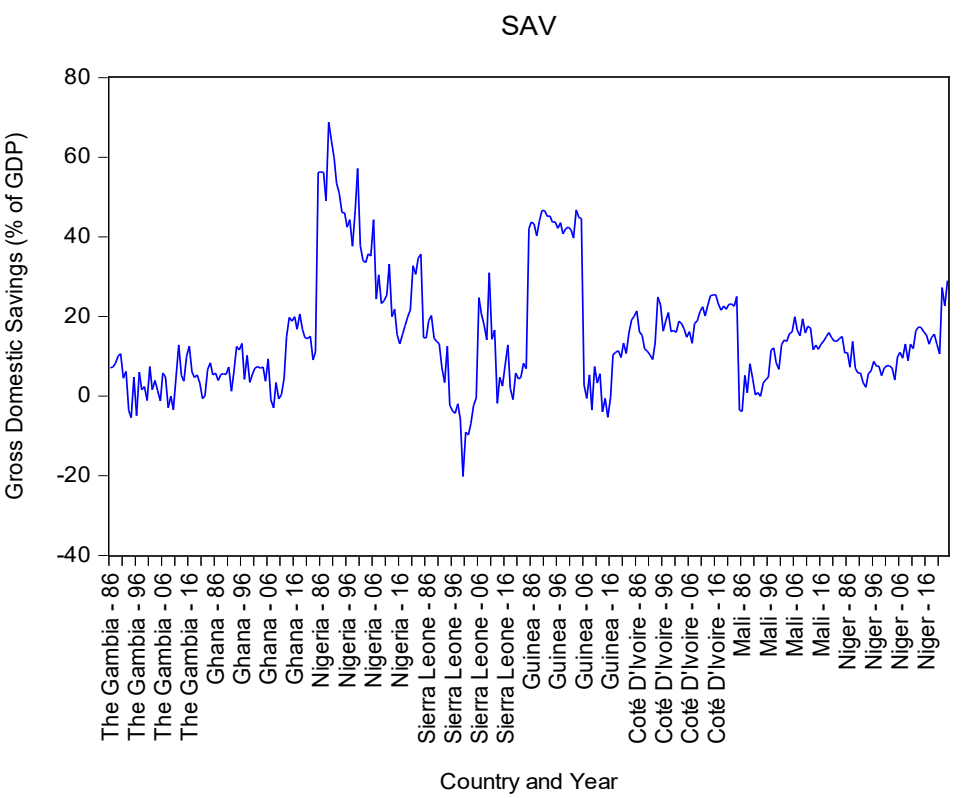
The provided summary statistics for the variables SAV (savings), DPR (deposit rate), PCI (*per capita* income) and INF (inflation rate) offer valuable insights into West Africa's financial landscape.

Result of Descriptive Statistics

Figures 1–4 present an overview of the eight studied countries in relation to various variables, including savings and deposit interest rates, income *per capita* and inflation rates, using line graphs.

In Figure 1, the behaviour of gross domestic savings measured as a percentage of GDP shows a marginal percentage of less than 20% throughout the studied period in the countries under consideration. Even though Nigeria, Sierra Leone and Guinea could boast savings at more than 20% of GDP (Nigeria having the highest savings up to 64%, followed by Guinea with up to 43% and Sierra Leone with up to 30% of GDP), the whole region, represented by eight countries, had a mean percentage of savings smaller than 20% of GDP, which indicates a low savings culture. Figure 1 also shows periods of dissaving in Gambia, Ghana, Sierra Leone and Guinea, when the percentage of savings was smaller than zero.

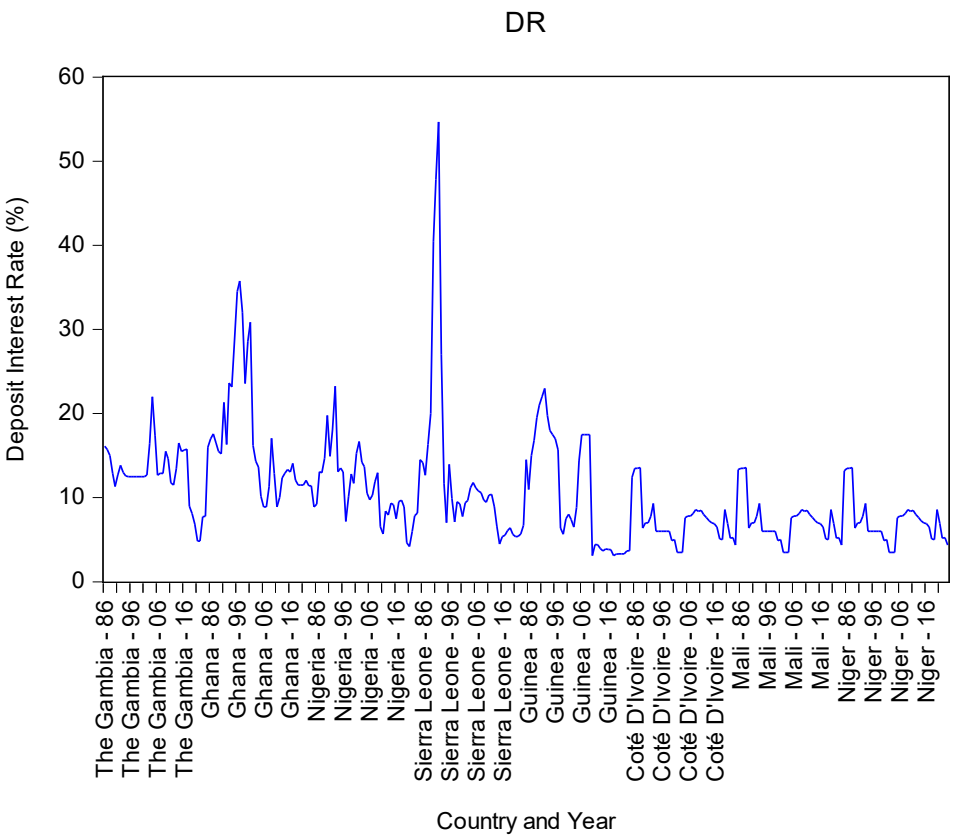
Figure 1. Savings in West Africa



Source: authors' work.

Figure 2 shows the deposit interest rate (DR) of the sampled countries in the region in a similar way. According to the figure, Guinea had the lowest DR (3.1%), Sierra Leone the highest (over 50%), and Ghana the second-highest (over 35%). The average deposit interest rate in the region hovered around 10%, which is relatively high compared to some developed countries. It can be observed that Sierra Leone, which had the highest deposit interest rate, failed to stimulate savings (the country had the lowest savings rate).

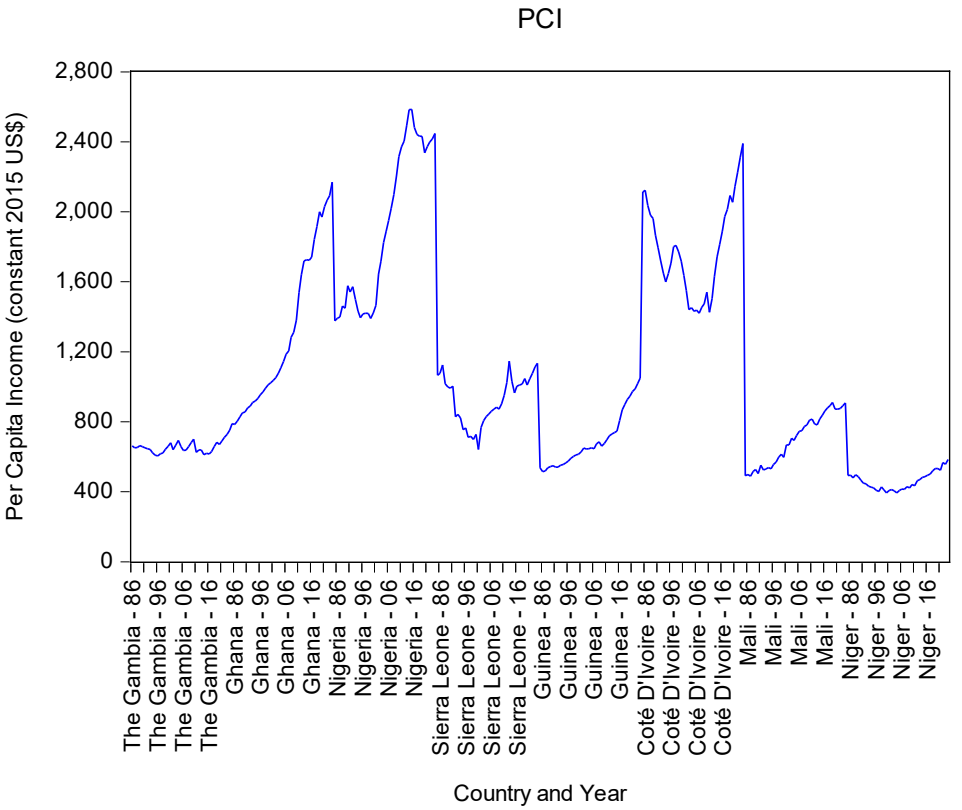
Figure 2. Deposit interest rate in West Africa



Source: authors' work.

Figure 3 presents the income of the studied countries assessed by means of the *per capita* income.

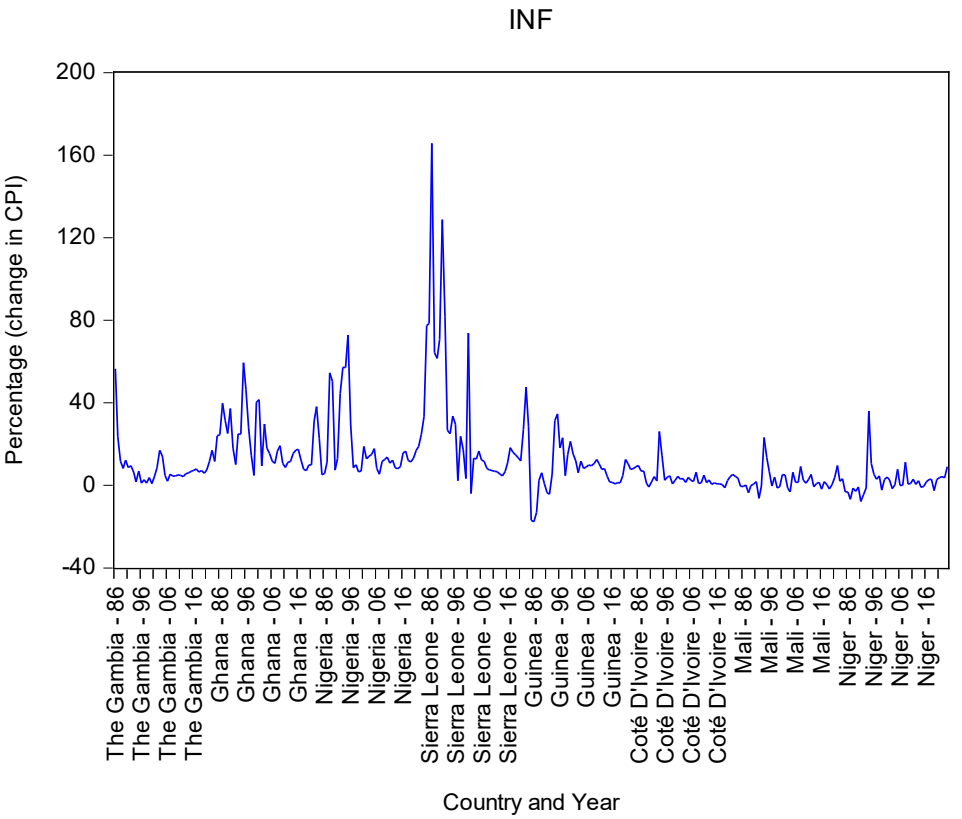
Figure 3. *Per capita* income in West Africa



Source: authors' work.

Figure 3 shows that Niger had the lowest annual income *per capita* of about 393 USD, followed by Mali (489 USD) and the Gambia (605 USD). The Gambia had a more stable PCI variation within the period – in other words, the country had a very low deviation of the PCI over the years. Other countries have a high rate of PCI deviation, with Nigeria taking the lead in this respect and being followed by Niger and Ghana.

Figure 4. Inflation in West Africa



Source: authors' work.

The inflation (INF) in the region over the studied period is presented in Figure 4. The average inflation totalled around 11% with Guinea experiencing negative inflation (–6.2%), which indicates a deflationary period. The highest inflation (over 160%) was observed in Sierra Leone. Ghana and Nigeria also saw challenging inflation rates. On the other hand, the Gambia, Guinea, Cote D'Ivoire, Mali and Niger had relatively low inflation rates.

4.2. Correlation matrix

We used a correlation matrix to show the nature and degree of correlation between the dependent and independent variables of the model. Here, although it expressed the relationship between the dependent and independent variables, more emphasis is placed on the relationship within the independent variables to detect the presence of multicollinearity in the specified model.

Table 2. Results of correlation matrix

Variable	SAV	DR	PCI	INF
SAV	1			
DR	0.0693	1		
PCI	0.3279	–0.0645	1	
INF	0.0977	0.5058	0.1090	1

Source: authors' computation.

Table 2 shows the results of the correlation matrix. The DR, PCI and INF have a positive association of 0.069, 0.328 and 0.098 with the dependent variable, SAV (savings). More importantly, the correlation coefficients (-0.065, 0.0506 and 0.109) within the independent variables are not up to the threshold of 0.8, which indicates that the explanatory variables are not highly correlated and hence free from the problem of multicollinearity.

4.3. Unit Root Test

For the stationarity test, two test statistics were adopted to test for the presence of unit root, and the results are presented in Table 3.

Table 3. Results of the panel stationarity test

Variable	Levin, Lin & Chu t*				Im, Pesaran and Shin W-stat				Remark
	Level		1st Diff		Level		1st Diff		
	Stat	Prob	Stat	Prob	Stat	Prob	Stat	Prob	
SAV	–0.0546	0.4782	–9.7846	0.000	–0.4818	0.3150	–10.7456	0.000	I(1)
DR	–2.8912	0.0019	–	–	–3.6081	0.0002	–	–	I(0)
LPCI	3.4238	0.9997	–6.6136	0.000	4.9433	1.0000	–5.9061	0.000	I(1)
INF	–5.9760	0.000	–	–	–7.5281	0.000	–	–	I(0)

Source: authors' computation.

The results in Table 3 indicate that DR and INF are stationary at level (I(0)), while SAV and PCI are stationary at the first difference (I(1)). This means the variables are integrated into both order zero and order one. As a result, this study used a technique that can differentiate lag selection for optimal model selection, specifically the panel autoregressive distributed lag (PARDL) model. This model applies to both short-run and long-run analyses, depending on the results of the cointegration test, which assesses the long-run relationship between the variables.

4.4. Panel Cointegration Test

To check for the presence of the long-run impact, the panel cointegration test is employed using both within-dimension (eight test statistics) and between-dimension (three test statistics), and these results are presented in Table 4, with deterministic intercept and trend and an automatic lag length selection based on the Schwartz information criteria with a maximum lag of eight.

Table 4. Results of panel cointegration test

Alternative hypothesis: common AR coefs. (within-dimension)					Alternative hypothesis: individual AR coefs. (between-dimension)		
Test	Statistic	Prob.	Weighted Statistic	Prob.	Test	Statistic	Prob.
Panel v-Stat	0.015	0.4939	0.0271	0.4892	Group rho-Stat	–0.4001	0.3445
Panel rho-Stat	–0.440	0.3299	–1.3304	0.0917	Group PP-Stat	–2.6117**	0.0045
Panel PP-Stat	–1.991**	0.0232	–3.0834**	0.0010	Group ADF-Stat	–2.7112**	0.0034
Panel ADF-Stat	–2.2465*	0.0123	–3.0504**	0.0011			

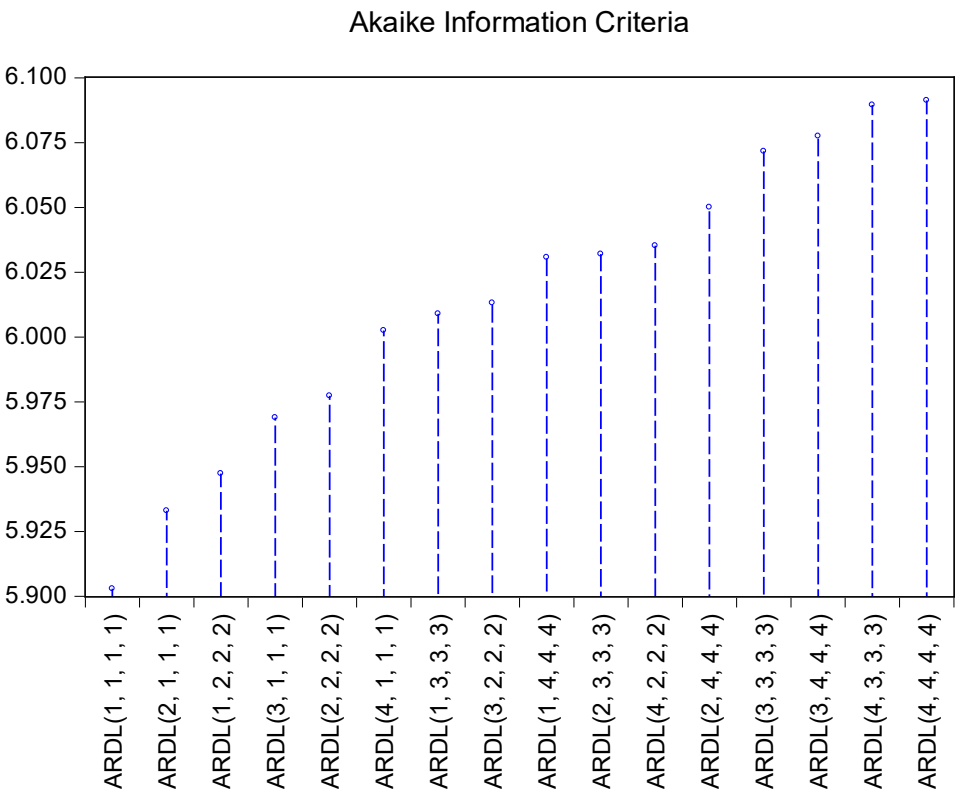
Note. ** and * indicate significance at 1% and 5% levels, respectively. V-stat is rescaled variance statistics, rho-stat is Rho statistics, PP-stat is Phillips-Perron statistics, ADF-stat is Augmented Dickey-Fuller stat. They are all test statistics for stationarity.

Source: authors' computation.

Table 4 showcases the overall eleven test statistics to ascertain the presence of cointegration in the long run. Of all the eleven test statistics, six identify cointegration in the model at the 1%- and 5%-level of significance, while five express otherwise (no long-run cointegration). Therefore, it is concluded that there is a long-run relationship in the model, which is in line with the findings of Obeh and Brotoboth (2021), Umoru and Tedunjaiye (2023), Fundji (2024) and Idi and Jabil (2024). Further analysis of the model is presented in Tables 5 and 6, which present the error correction, long-run and short-run analysis for the clarity of the impact analysis.

To continue the analysis, we need to select the best lag order for the model. The study used the Akaike information criteria (AIC), and the result is presented in Figure 5.

Figure 5. Model selection order using AIC



Source: authors' work.

From the figure, the PARDL (1,1,1,1) has been chosen as the best lag among sixteen models showcased, because it has the lowest AIC coefficient (5.900). Therefore, these lags were used for the long-run and short-run impact analysis, as presented in Tables 5 and 6.

Table 5. Results of panel ARDL analysis (long-run analysis)

Selected model: ARDL (1, 1, 1, 1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.*
DR	–0.0345	0.1674	–0.2061	0.8368
LPCI	15.9722**	3.6271	4.4035	0.0000
INF	0.1617*	0.0821	1.9696	0.0499

Note. ** and * indicate significance at 1% and 5% levels, respectively. This is the result of Equation 5.
Source: authors' computation.

In the long run, as evident in Table 5, on average the deposit rate (DR) does not have any significant impact on the savings (SAV) in the region, as demonstrated by its probability of 83.68%. This implies that the interest rate on savings in the region could not stimulate customers to deposit their funds as savings, which is in line with the study of Babalola and Abdul (2022), but different from the findings of Umoru and Tedunjaiye (2023), Fundji (2024) and Idi and Jabil (2024). This discrepancy is caused by the measurement of the interest rate. The *per capita* income (PCI), which is the main cause of saving, has a significant and direct effect on savings in the region, as indicated by the probability of 0.000%. This conforms with most theories that explain the impact of income on savings, such as the classical, Keynesian and the monetarist theories. This result confirms that, in the long run, income has a significant and direct impact on savings. Let us check, then, what happens in the short run.

Table 6 presents the results of the error correction and the short-run impact analysis, after the automatic lag selection of PARDL (1,1,1,1).

The deposit interest rate (DIR) is negative and has an insignificant effect on SAV in the region in the short run at one lag period, as indicated by the probability (29.19%). By implication, in the short run, the interest rate (deposit rate) does not significantly stimulate savings.

Table 6. Results of panel ARDL analysis (short-run analysis)

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
ECM(-1)	–0.3094	0.0733	–4.2239	0.0000
D(DR)	–0.0975	0.0923	–1.0561	0.2919
D(LPCI)	20.1150	21.6894	0.9274	0.3545
D(INF)	0.0466	0.0561	0.8301	0.4072
C	–30.7928	7.2081	–4.2720	0.0000

Note. This is the result of Equation 5.

Source: authors' computation.

The two control variables (PCI and INF) are positive sign but also insignificant in the short run, as indicated by their p-values (35.45% and 40.72%). This implies that on average, in the short run, an increase in the *per capita* income does not affect savings immediately. Inflation plays a significant role in shaping saving behaviour, as fluctuations in the rate of price level do not immediately influence savings in the short term. Initially, when prices rise, individuals may not immediately adjust their savings habits. This delay can be attributed to various factors, such as a lack of awareness about how inflation erodes purchasing power or a tendency to prioritise current consumption over future savings. As a result, the impact of inflation on savings tends to manifest over a longer period, ultimately affecting individuals' financial decisions and their ability to maintain or grow their savings in real terms.

The ECM coefficient has the correct negative sign (-0.3094), and it is significant with a probability of 0.00%, meaning that yearly, a 30.9% disequilibrium in the model is corrected. Although the speed is low, it is quite significant in making the necessary corrections.

5. Conclusions

The study explores low saving levels in West Africa and the impact of interest rates on the saving behaviour using the life-cycle hypothesis (LCH). We found that while interest rates did not significantly affect savings, the *per capita* income did. The relationship between interest rates and savings is complex, influenced by factors beyond the economic theory. This research contributes to the existing literature by testing several theories, including the classical theory of interest rate, the loanable fund theory of interest rate and the Keynesian theory of interest rate in the West African context, using the deposit interest rate to measure its effect on savings. An econometric model was developed to analyse the practical application of these theories.

The study recommends that, since the deposit interest rates do not have a significant effect on savings, policies focused on other incentives rather than interest rates should be developed to promote long-term investment strategies in order to balance immediate investments with savings and encourage firms to set aside funds for future use.

References

African Development Bank Group. (2021). *West Africa Economic Outlook 2021*. African Development Bank. https://www.afdb.org/en/documents/west-africa-economic-outlook-2021.

African Development Bank Group. (2025). *Country Focus Report 2025 Sierra Leone*. African Development Bank. https://www.afdb.org/sites/default/files/documents/publications/sierra_leone_cfr_2025.pdf.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press. https://doi.org/10.1093/0199245282.001.0001.

Babalola, A. (2021). Impact of Interest Rates on Exchange Rates in Nigeria. An Analytical Investigation. *Timisoara Journal of Economics and Business*, *14*(2), 107–124. https://doi.org/10.2478/tjeb-2021-0007.

Babalola, A., & Abdul, A. I. (2022). Does Interest Rate Really Stimulate Savings in Nigeria?. *Folia Oeconomica Stetinensia*, *22*(2), 18–37. https://doi.org/10.2478/foli-2022-0017.

Babalola, A., Yelwa, M., & Olaniyi, O. (2023). Monetary Policy and Unemployment Rate in Developing Economies: Evidence from Nigeria. *Annals of Spiru Haret University. Economic Series*, *23*(3), 39–62. https://doi.org/10.26458/2332.

Bank of Ghana. (2022). *Monetary Policy Report*. https://www.bog.gov.gh/wp-content/uploads/2022/11/Monetary-Policy-Report-October-2022.pdf.

Bismans, F. J., & Damette, O. (2025). *Dynamic Econometrics. Models and Applications*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-72910-2.

Blanchard, O., Amighini, A., & Giavazzi, F. (2017). *Macroeconomics. A European Perspective* (3rd ed.). Pearson.

Central Bank of Liberia. (2023). *Annual Report and Financial Statements 31 December 2023*. https://www.cbl.org.lr/sites/default/files/documents/Central%20bank%20of%20Liberia%20signed%20fs-2023._240927_151505.pdf.

Central Bank of Nigeria. (2023). *CBN Update*, *5*(7), 1–18. https://www.cbn.gov.ng/Out/2023/CCD/CBN%20UPDATE%20JULY%20CURVED%202023.pdf.

Diop, S., & Diaw, A. (2023). Transmission channels of the COVID-19 pandemic effects in West African Economic and Monetary Union countries and BCEAO responses. In S. Olawoye (Ed.), *COVID-19 and the Response of Central Banks* (pp. 37–52). Edward Elgar Publishing. https://doi.org/10.4337/9781802205374.00013.

Feldstein, M. (1974). Social Security, Induced Retirement, and Aggregate Capital Accumulation. *Journal of Political Economy*, *82*(5), 905–926. https://doi.org/10.1086/260246.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton University Press.

Fundji, O. J. (2024). The Impact of Financial Inclusion on Economic Growth based on East, West and Southern Africa. *International Journal of Economics and Financial Issues*, *14*(5), 203–209. https://doi.org/10.32479/ijefi.16404.

Gertler, M., & Kiyotaki, N. (2010). Financial Intermediation and Credit Policy in Business Cycle Analysis. In B. M. Friedman & M. Woodford (Eds.), *Handbook of Monetary Economics* (vol. 3, pp. 547–599). North-Holland. https://doi.org/10.1016/B978-0-444-53238-1.00011-9.

Idi, R. Z., & Jabil, Y. I. (2024). Impact of Interest Rate on Domestic Savings And Investment In Nigeria. *Journal of Management Science and Entrepreneurship*, *3*(7), 98–114. https://berkeleypublications.com/bjmse/article/view/142.

IHS Global. (2016). *EViews 9 User's Guide I*.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. Palgrave Macmillan.

Loaba, S. (2022). The Impact of Mobile Banking Services on Savings Behavior in West Africa. *Global Finance Journal*, *53*, 10–20. https://doi.org/10.1016/j.gfj.2021.100620.

Mankiw, N. G. (2014). *Principles of Economics*. Cengage Learning.

Modigliani, F., & Brumberg, R. (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In K. K. Kurihara (Ed.), *Post Keynesian Economics* (pp. 388–436). Rutgers University Press.

Muntanga, M. S. (2020). *The impact of interest rates on savings and investment in Zambia (1980–2018)*. School of Social Sciences Cavendish University Zambia.

Muse, B. O. (2024). Differential Impact of Interest Rate Regimes on Savings in Nigeria: New Empirical Evidence. *African Journal of Business & Economic Research*, *19*(2), 145–164. https://doi.org/10.31920/1750-4562/2024/v19n2a7.

Obeh, H. O., & Brotoboh, D. E. (2021). Impact of Interest Rate Spread on Savings in Nigeria. An Empirical Investigation. *ESUT Journal of Social Sciences*, *6*(3), 283–294. https://www.esutjss.com/index.php/ESUTJSS/article/view/89.

Obi, C. I. (2022). Impact of Interest Rate on Savings and Investment in Nigeria. *African Journal of Business and Economic Development*, *2*(6), 1–18. https://www.ijaar.org/articles/ajbed/v2n6/ajbed2526.pdf.

Pesaran, M. H., & Shin, Y. (2003). An Autoregressive Distributed-Lag Modelling Approach to Cointegration Analysis. In S. Strom (Ed.), *Econometrics and Economic Theory in the 20th century. The Ragnar Frisch centennial symposium* (pp. 371–413). Cambridge University Press. https://doi.org/10.1017/CCOL521633230.011.

Umoru, D., & Tedunjaiye, O. D. (2023). Saving in Presence of Volatilities in Interest Rate and Exchange Rate Devaluation. *Asian Journal of Economics, Business and Accounting*, *23*(21), 111–128. https://doi.org/10.9734/ajeba/2023/v23i211121.

World Bank. (2018). *Poverty and Shared Prosperity 2018. Piecing Together the Poverty Puzzle*. https://doi.org/10.1596/978-1-4648-1330-6.

World Bank Group. (2025). *World Development Indicators 2025* [data set]. Retrieved October 19, 2025, from https://databank.worldbank.org/source/world-development-indicators.

Przegląd Statystyczny. Statistical Review, 2025, vol. 72, 2, 54–83 https://doi.org/10.59139/ps.2025.02.3
Received: 4.04.2025, revised: 20.10.2025, accepted: 21.10.2025

Application of tree ensemble methods to the two-asset portfolio selection problem – a case study

Krzysztof Kaczmarek,[a] Aleksandra Rutkowska[b]

Abstract. The aim of the study was to construct a two-asset optimal investment portfolio using machine learning and macroeconomic data at monthly and quarterly intervals. The auxiliary objective was to identify which macroeconomic variables significantly impact the estimation of the S&P 500 stock index and the USD/GBP currency pair. The framework included two steps: firstly, time series forecasts were conducted using tree ensemble methods, namely the random forest and XGBoost, and secondly, the forecasts were used as expected values to construct the portfolios. We analyze the extent to which the structure of a portfolio based on the estimated data differs from the one built using historical data. The results of the research showed that it was possible to use the macroeconomic data to efficiently forecast the considered time series and construct an optimal investment portfolio.

Keywords: random forest, ensemble model, XGBoost, portfolio optimization

JEL: G11, G17

1. Introduction

Investing requires skillful risk management and returns optimization. In a dynamic market environment, where asset volatility can be both an opportunity and a threat, constructing a well-balanced investment portfolio is crucial. This study explores the strategy of building a two-component portfolio consisting of a currency and a stock index. We focus on a long-term investment and simple diversification, in contrast to numerous publications that emphasize highly active investing and trading. A two-asset portfolio is easier to monitor, rebalance, and understand over the long term. With fewer components, there is less complexity in tracking performance and making strategic decisions. While equity/bond strategies are the most analyzed (Pham, 2025), we explore a slightly riskier yet potentially more rewarding combination, namely an equity index and a currency. This portfolio structure offers several key benefits. First, combining a currency with an index can enhance diversification: when the currency appreciates, stock indexes may react differently, helping to mitigate the overall risk.

[a] Student at Informatics and Econometrics, Poznan University of Economics and Business, al. Niepodległości 10, 61–875 Poznań, Poland, e-mail: 84890@student.ue.poznan.pl, ORCID: https://orcid.org/0000-0003-1469-646X.

[b] Poznan University of Economics and Business, Institute of Informatics and Quantitative Economics, Department of Applied Mathematics, al. Niepodległości 10, 61–875 Poznań, Poland, e-mail: aleksandra.rutkowska@ue.poznan.pl, ORCID: https://orcid.org/0000-0002-2111-7764.

Second, monetary policy and global economic trends often influence the relationship between exchange rates and stock markets, allowing investors to leverage correlations (or the lack thereof) to improve portfolio efficiency. Finally, selecting the right combination of these two components can enhance the risk-return profile, making this approach attractive for individual and institutional investors.

Forecasting financial markets remains a central concern for both investors and researchers, despite the challenges posed by the Efficient Market Hypothesis (EMH), which suggests that financial markets are largely unpredictable, as asset prices already reflect all the available information. According to Țițan (2015), who provides a comprehensive review of the empirical studies testing the EMH, the dilemma of whether the market is efficient or not remains unresolved. Recent advances in artificial intelligence and machine learning have reignited interest in improving financial forecasting. Modern computational techniques offer new possibilities for extracting patterns from complex and high-dimensional financial data, potentially enhancing the predictive power of models even in markets traditionally considered efficient. A particularly interesting overview of the capabilities of different machine learning models for time series forecasting can be found in Ahmed et al. (2010), who compare the performance of several models applied to the M3 competition data. Additional comparative studies include the review by Tang et al. (2022) and Maung and Swanson (2025), which focus specifically on machine learning approaches for financial time series forecasting. From a wide range of methods, we selected ensemble tree-based machine learning methods for our study. Tree-based ensemble methods, such as random forest (RF) and Extreme Gradient Boosting (XGBoost) have gained popularity in forecasting due to their robustness and accuracy. Moreover, they capture complex, non-linear relationships in time series data, which makes them suitable for real-world forecasting applications (Wong et al., 2023). These methods demonstrated competitive performance (they are fast and more efficient although with a slightly higher number of forecast errors) for more complex recurrent neural networks and LSTMs (Nabipour et al., 2020a). Moreover, RFs can be trained with a relatively small amount of data (Roßbach, 2018).

This case study aims to shed light on several aspects related to the broader research objective. In particular, it explores: (1) which macroeconomic variables may have a notable impact on the estimation of a stock market index and a currency pair; (2) how the structure of a portfolio based on estimated data compares with the one constructed using complete information on current returns; and (3) what implications these differences may have for the portfolio's value, return, and risk profile. By examining these areas, the study seeks to offer a more detailed view of portfolio behavior under model-based estimation, contributing to the ongoing discussion in the area of investment strategy analysis.

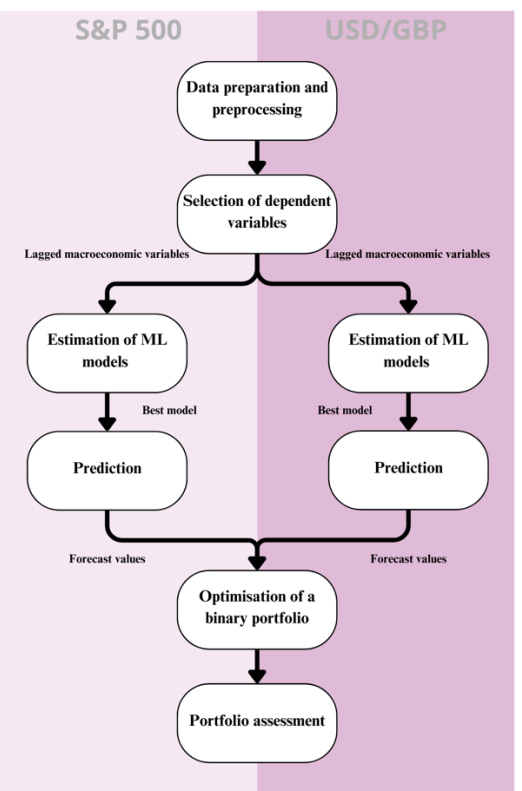
In the current literature, researchers have analyzed a variety of financial and macroeconomic indicators to model capital market behavior. The following variables have been examined in studies by Leippold et al. (2022): dividend yield, price-to-earnings (P/E) ratio, book-to-market ratio, net equity expansion, stock variance, term spread, and inflation. Additionally, money supply aggregates such as M2, trading volume, and monthly turnover were also considered. In the context of derivatives markets, futures contracts on silver, platinum, crude oil, and gold were analyzed by Shen et al. (2012) and Zhong and Enke (2019). They also used foreign exchange rates and stock indexes. Technical indicators such as the momentum indicator were studied by Choudhry and Garg (2008) and Reddy (2018), while %R Williams and the price volume trend were explored by Choudhry and Garg (2008). The stochastic oscillator was analyzed by Choudhry and Garg (2008), Patel et al. (2015), and Hegazy et al. (2014). Furthermore, technical indicators such as the relative strength index and moving average convergence divergence (Hegazy et al., 2014; Patel et al., 2015), moving averages (Hegazy et al., 2014; Patel et al., 2015; Zhong & Enke, 2019) were also investigated, as were indicators such as the Accumulation /Distribution (A/D) line and the Commodity Channel Index (CCI) (Patel et al., 2015); and finally, closing prices were described by Zhong & Enke (2019).

In the case of the currency pair, in articles where the authors applied ML models, the researchers used macroeconomic variables such as inflation or interest rates (Boyoukliev et al., 2022; Kaushik & Giri, 2020), money supply aggregates, government reserves, trade balance, and IIP (Kaushik & Giri, 2020). Boyoukliev et al. (2022) also used unemployment and, as Matuszewska-Janica & Witkowska (2008), interest rates on treasury bills. The reviewed studies also used precious metals and oil derivatives futures and stock indexes (Matuszewska-Janica & Witkowska, 2008). Other frequently used variables included technical analysis indicators, such as the Relative Strength Index and Rate of Change, which were examined by Abreu et al. (2018), Loh et al. (2022), Qi et al. (2020), and Baasher and Fakhr (2011). Moving averages (including WMA, EMA, SMA) were analyzed by Loh et al. (2022), Mabrouk et al. (2022) and Qi et al. (2020). Similarly, the moving average convergence/divergence (MACD) indicator was assessed by Baasher and Fakhr (2011), Mabrouk et al. (2022), and Loh et al. (2022). Other indicators such as the commodity channel index (Baasher & Fakhr, 2011) and the stochastic oscillator (Abreu et al., 2018; Baasher & Fakhr, 2011) were also considered, as were opening prices, highest and lowest prices during a session, closing prices, and trading volumes (Loh et al., 2022; Mabrouk et al., 2022).

2. Methodology

The proposed solution combines machine learning with standard investment portfolio estimation methods. The proposed framework is presented in Figure 1. In the first step, the data used in the study were downloaded and prepared. The second step consisted in a two-stage selection of the variables used in the models. The first stage involved the literature review depicted in the introduction, aiming to identify the variables that should, in theory, significantly affect the estimated series. The second stage involved the selection of these variables considering their correlation with the dependent variable. In the third step, the lagged macroeconomic variables were put in the selected machine-learning models and estimated. Then, those models that achieved the smallest root mean squared errors were selected. Using these models, one-period estimations of the levels of the S&P 500 index and the USD/GBP currency pair were made. The resulting time series were applied to construct minimum variance and maximum Sharpe ratio portfolios.

Figure 1. Research flow



Source: authors' work.

The final step was the portfolio assessment. First, the structure of the portfolios built using predictions from machine learning models was compared to that of the portfolios based on the realized data, with the current return rate known. Then, to assess the optimal portfolios, a visualization of the capital changes over time was created and the portfolio evaluation metrics were calculated.

The portfolio optimization issue involves determining the optimal weights for different assets. This problem was confronted by the theory presented by Markowitz (1952) – a groundbreaking idea in the area of portfolio construction that has become the foundation of modern approaches to this issue. The key idea is to condition the selection of the assets' weight on the expected return and risk, which means maximizing return for a given level of risk, or minimizing risk for a given return. According to Markowitz, the return rate on an investment represents the income earned from it, with investors knowing the probability distribution of returns. Investors' risk estimation is proportional to the expected return distribution. They make decisions based solely on two parameters of the probability distribution of returns. Investors prefer to minimize risk for a given rate of return, and for a given level of risk, they choose the investment offering the highest return. In Markowitz's theory, asset returns are identified with random variables. If the distribution of such a variable is known, then it is possible to determine the parameters of this distribution from it. Otherwise, investors are forced to estimate the expected value and the variance on the basis of historical data. The estimation of these parameters may cause problems related to the selection of an appropriate estimator, as well as to the quality of the obtained result. Moreover, the choice of the period from which the data used for the estimation is drawn can significantly affect the results. Considering the classical approach, the distributions should be normal or close to normal, but this may not be the case (Kaszuba, 2011).

Therefore, assets with the smallest possible variance, the highest return and the lowest correlation should ultimately be selected for the portfolio. For this reason, it is worth considering portfolio diversification, which enables the reduction of portfolio risk (Łuniewska, 2012).

2.1. Time series forecasting

RF, which is a machine learning algorithm based on decision trees, is a popular method used in relation to classification and regression problems (Król-Nowak & Kotabra, 2022). Multiple decision trees are involved when creating an RF model. This results in the reduction of the negative effects of overfitting some of the trees that make up the forest. Classification and regression in this algorithm then consist of comparing the obtained results by the individual independent trees in the forest.

When comparing the results from the trees, most of the same outcomes finally shape the classification or regression value of the forest (Basak et al., 2019; Géron, 2022).

XGBoost is also a decision tree-based algorithm capable of solving regression and classification problems. It was developed for better performance compared to other tree-based models (Nabipour et al., 2020b). The approach differs significantly from that used in the RF, where the result is supposed to be the best possible partitioning verified by impurity coefficients. In this case, however, the method uses gradient boosting, which involves combining several weak predictors or classifiers into one. This is done by sequentially training the successive models, with each successive model attempting to correct the errors of its predecessor (Basak et al., 2019; Géron, 2022). Table 1 presents a comparison of both methods. One undoubted advantage of these methods is that tree-based ensemble models, such as RF and XGBoost, are generally robust to multicollinearity in terms of predictive accuracy (Cabrera Malik, 2024; Gregorutti et al., 2017; Roy & Larocque, 2012). However, it should not be forgotten that in the case of multicollinearity, measures of trait importance obtained from these models may become unreliable when strong correlations occur between the predictors.

Table 1. Comparison of RF and XGBoost features

Feature	RF	XGBoost
Ensemble type	Bagging	Boosting
Tree building	Parallel	Sequential
Focus	Diversity of trees	Correcting errors of the previous trees
Accuracy	Generally good, but may be slightly lower than that of XGBoost	High accuracy, often of a state-of-the-art level
Training time	Generally faster	Can be slower, but in many cases more accurate
Overfitting	Relatively robust to overfitting	Can be prone to overfitting, but regularization helps overcome its effects
Interpretability	More interpretable	Less interpretable

Source: authors' work.

In the forecasting stage, the explanatory variables are the closing prices of the S&P 500 and the USD/GBP currency pair. We use quarterly and monthly data downloaded from stooq.pl. The independent variables are from Q1 1985 to Q2 2023 for quarterly frequency and from January 1990 to September 2023 for the monthly interval. The total length of the time series is 154 and 405 periods, respectively. 20% of the observations are a test sample, not used for model training. Most of the

downloaded data were complete, but any missing values were filled by adjusting the next or previous value using the average growth rate of the nearby observations.

2.2. Portfolio construction

This paper compares two portfolios: one optimized for minimal volatility (further designated as Portfolio 1: Min variance) and the other for the Sharpe ratio (Portfolio 2: Max Sharpe ratio).

In the case of a two-asset portfolio, the risk is determined by the variance of the portfolio return and is given by the formula:

$$V_p = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2cov_{12}, \tag{1}$$

where:

w_i – share of the *i*-th asset in the portfolio,

σ_i – standard deviation of return on the *i*-th asset,

cov_{ij} – covariance of the *i*-th and *j*-th return rates.

The Sharpe ratio is one of the possible indicators for assessing investment performance in terms of risk and was first presented by Sharpe in his work on the asset pricing model (Sharpe, 1966). It considers both return and risk, combining these two key factors into a single measure, as shown in Equation (2) (Sharpe, 1998):

$$S = \left(R_p - R_f\right)/\sigma_p, \tag{2}$$

where

R_p – average return of the portfolio,

R_f – risk-free rate of return,

σ_p – standard deviation of the portfolio, so $\sqrt{V_p}$, according to (1).

The higher the value, the better rate of return on investment to the risk taken. The outcome indicates how many units of return can be obtained for each of the incurred unit of risk that the investor is potentially exposed to.

Besides the metrics used to create the portfolio, other ones were used to evaluate it, i.e. the average return, tracking error and maximum drawdown. The tracking error measures how much an investment differs from its benchmark. It is often used for ETFs, which aim to closely follow an index or make certain performance assumptions (Charteris & McCullough, 2020). The maximum drawdown is a popular risk measure commonly used in the financial sector. It measures how severe a single investment loss can be for an investor. This indicator has been defined as the maximum cumulative loss occurring, beginning at a price peak and ending at a bottom (Choi, 2021; Magdon-Ismail & Atiya, 2004).

3. Empirical results

3.1. Prediction results

To estimate the expected value of the return rate, two different ML models were estimated with the one with a lower root mean square error (RMSE) chosen for the next step. The libraries used in the forecasts, along with the models' hyperparameters are presented in Table A1 in the Appendix. The independent variables for the models were chosen based on their correlation with the dependent variable. From this set, we selected those that also aligned with the macroeconomic variables identified in the literature review. For the results to truly reflect the predictive capabilities, the independent variables lagged when building the models, so the data from the *i*-th period were used to estimate the dependent variable in the i+1 period. The results are presented in Table 2.

Table 2. RMSE of return predictions of the RF and XGBoost with the naive method as a benchmark

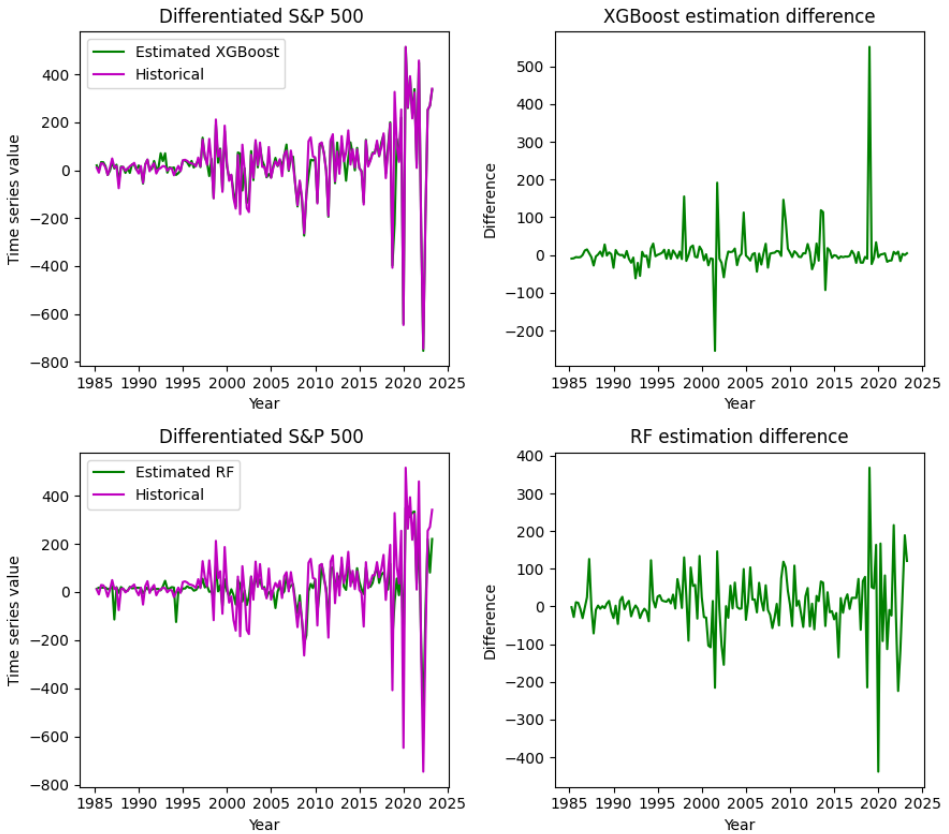
Assets	RF	XGBoost	Naive
Quarterly data			
S&P 500	103.07	130.49	204.51
USD/GBP	0.0377	0.0378	0.0429
Monthly data			
S&P 500	98.12	116.52	129.72
USD/GBP	0.0146	0.0170	0.0232

Note. The results from the test sample cover 20% of the observations.
Source: authors' work.

3.1.1. Quarterly interval

8 out of the 1,037 sampled variables were used to estimate the quarterly values of the S&P 500 index (see Figure A2 in the Appendix). Figure 2 shows the historical differentiated values of the index with the estimated values and the differences between the historical (real) and estimated differentiated values. Both models outperform the naive forecast: the XGBoost reduced the forecast error by 36.19%, and the RF by 49.60%.

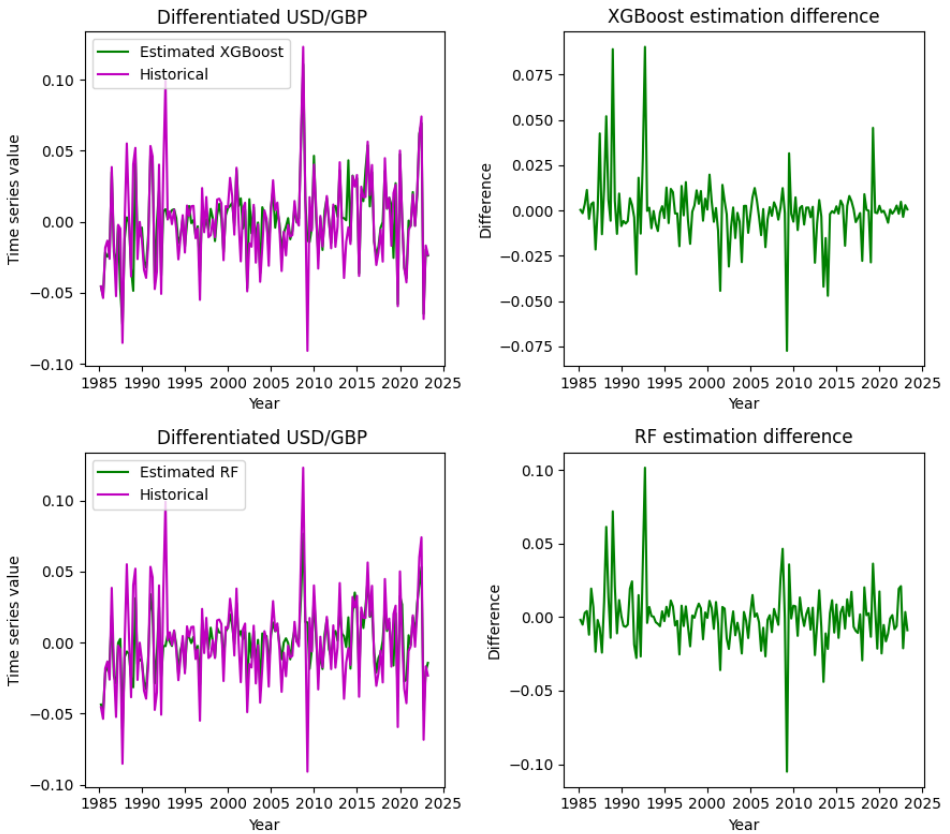
Figure 2. Estimated and historical quarterly differences of the S&P 500 index (left column) and estimation difference (right column)



Note. The results relate to the whole sample.
Source: authors' work.

The estimation of the quarterly differentiated closing prices of the USD/GBP currency pair used 8 variables from the 1,938 considered, of which both the US and the UK variables were among the sampled macroeconomic variables (for details, see Table A2 and Figure A2 in the Appendix) The results of the estimations, together with the differentiated historical values, are presented in Figure 3. Both models are characterized by a good fit to the data. Once again, both models corrected the forecast errors compared to the naive forecast by about 12% although there is no longer such a clear difference between the RF and the XGBoost.

Figure 3. Estimated and historical quarterly differences of the USD/GBP currency pair (left column) and estimation difference (right column)



Note. The results concern the whole sample.
Source: authors' work.

According to the results presented in Table 2 for the US stock exchange, the smallest error was observed in the RF regression model, and the same model proved to be the best for the currency pair.

The significance of the individual parameters was checked (see Figure A5 in the Appendix) for those models that achieved the best results (i.e. the RF for both estimated series). The most important variables for the stock market index are: the M3 aggregate money supply in the US, the US consumer price index (CPI), with the 2015 outcome as the baseline, and the net trade balance (the time series are presented in Figure A1 in the Appendix).

In the case of the model built to estimate the currency pair, the characteristics that have brought the greatest improvement in this case are: the UK public debt, the M1

aggregate money supply in the UK, the UK CPI and the US GDP (see Figure A2 in the Appendix). Despite the above distinction of features, it should be noted that the M1 monetary aggregate has the largest average contribution exceeding 20%, while the rest of the mentioned features exceed the threshold of 10% of the average total contribution to the model improvement.

Notably, three of the four most important variables used to estimate this currency pair are associated with the UK economy. This suggests that the US economy has a relatively minor impact on the UK.

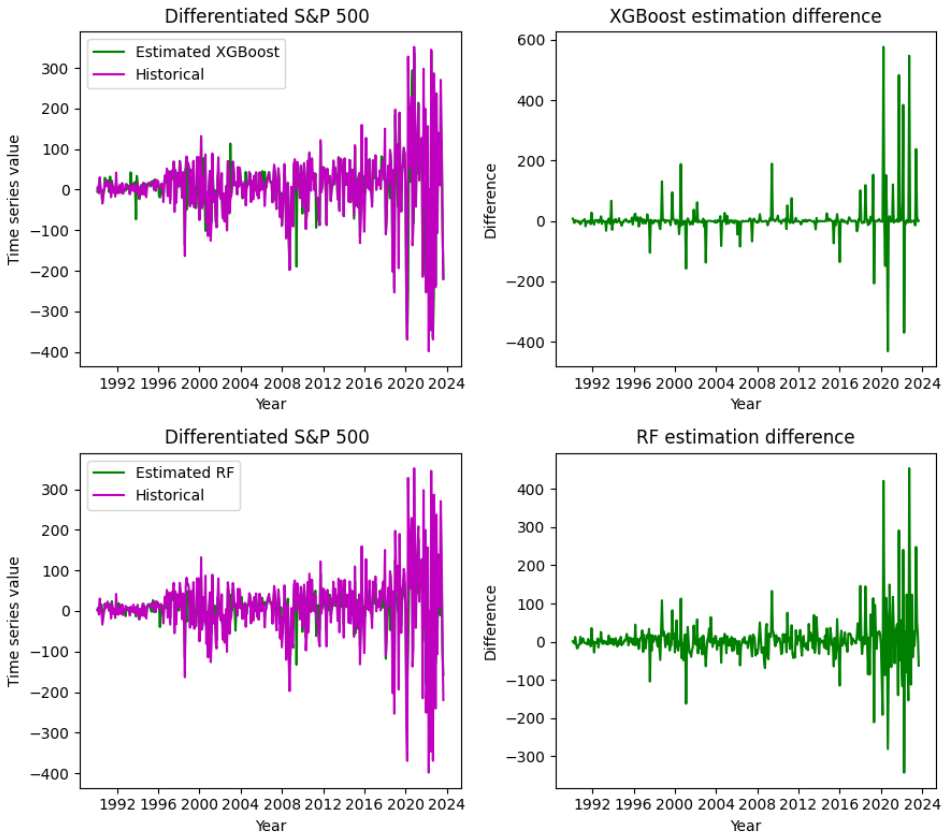
3.1.2. Monthly interval

In constructing models to estimate the differentiated stock market for the monthly interval, 6 out of the 780 independent variables were included (see Figure A3 in the Appendix). The significant difference in the number of the available data results from the fact that not all macroeconomic data are published monthly. The results are presented in Figure 4 and, once again, the models had a very good match. In the test sample, the XGBoost forecasts were better than the naive ones by 10% and RF by almost 25%.

From among a total of 1,379 variables considered, 11 independent variables were used in the construction of the models for the currency pair. Figure 5 illustrates the obtained results. The XGBoost improved the forecast results by 27% compared to the naive forecast, and the RF by 37%.

The RF models achieve the smallest RMSE for both assets, so the structure of the monthly portfolio is determined by their results.

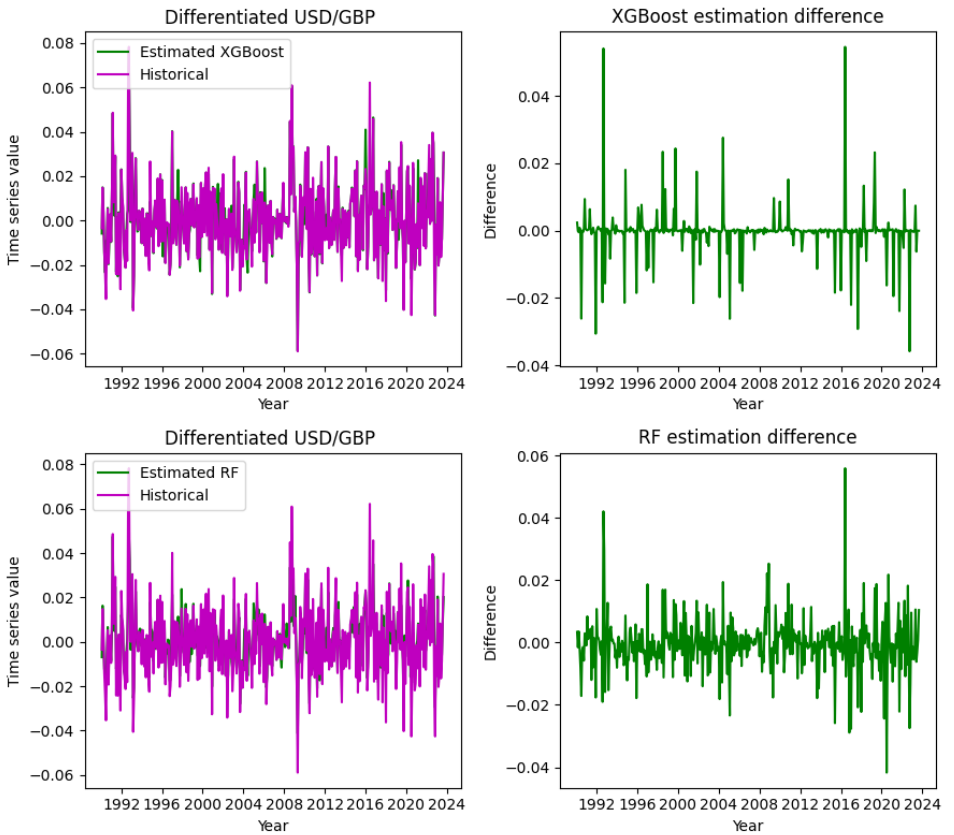
Figure 4. Estimated and historical monthly differences of the S&P 500 index and estimation difference



Note. The results relate to the whole sample.

Source: authors' work.

Figure 5. Estimated and historical monthly differences of the USD/GBP currency pair and estimation difference



Note. The results relate to the whole sample.
Source: authors' work.

For the stock market index, again three of the five variables played a significant role in improving the quality of the model, and these were: the M3 aggregate money supply in the US, the US CPI (with the 2015 outcome as the baseline) and the net trade value. It is worth noting that both the CPI and the M3 aggregate played a crucial role in both time intervals. Although all the variables put in the model should, according to the literature review, play an important role in the closing prices, the relevance of the interest rates was below 10% in the conducted study.

The second model shows a lower diversification in the average impact of the variables, which may be indicative of their inferior selection for the model. The variables that brought the greatest improvement were the US and UK real effective exchange rates calculated as weighted average two-sided exchange rates adjusted for relative consumer prices with 2020 as the base period (see Figure A4 in the

Appendix). The first variable played twice as an important role in the model as the second variable, and in addition, the rest of the variables did not exceed 10% of the average contribution (for importance, see Figure A6 in the Appendix).

3.2. Investment portfolio results

To build the portfolio, the expected value of each instrument was based on the return rate from the one-period prediction of the selected model. Asset shares were calculated using an algorithm that either minimized variance (1) or maximized the Sharpe ratio (2). A minimum asset weighting of 10% was set to maintain portfolio stability.

These initial portfolio weights were considered 'raw' since constantly adjusting them for minor changes would have been impractical. To account for transaction costs and frequent fluctuations, weight adjustments were only made when an asset's share changed by more than 2%. Additionally, portfolios calculated using ex-post data were used as benchmarks (here: xxx_hist). The deviation from these benchmarks was measured by calculating the root of the sum of the squared differences between the estimated and benchmark weights. We calculated the portfolio for the following cases:

A – raw and recalculated weights (here: xxx_new) of the minimum variance portfolio,

B – raw and recalculated weights (here: xxx_new) of the maximum Sharpe ratio portfolio,

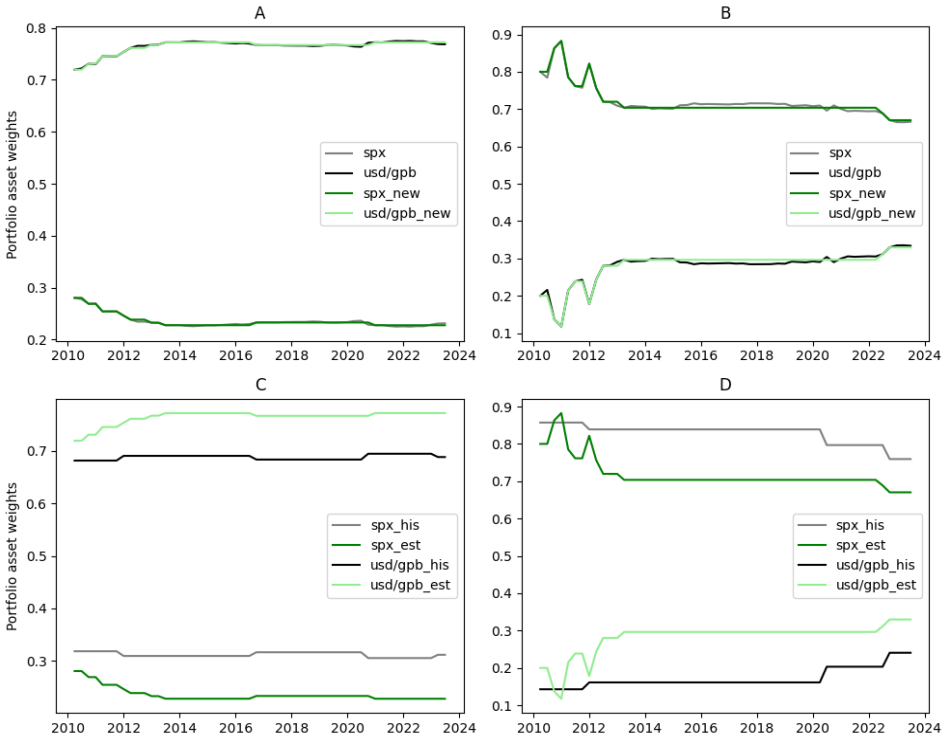
C – recalculated weights (here: xxx_est) and weights based on the realized values (here: xxx_hist) of the minimum variance,

D – recalculated weights (here: xxx_est) and weights based on the realized values (here: xxx_hist) of the maximum Sharpe ratio portfolio.

The graphs in the first row of Figure 6 illustrate the structure of the portfolios determined by variance minimization (A) and Sharpe ratio maximization (B) for the quarterly interval. Portfolio (A) maintains a quite stable structure remaining in a similar ratio of 2:8 for the currency pair. A similar stable structure was obtained for (C), where the ratio was 3:7.

For portfolio (B), the structure is slightly more unstable at the beginning, and stabilizes at a ratio of 3:7, but the asset weights reversed. The weights in (D) exhibit a similar behavior. The obtained results indicate a higher volatility of the stock index.

Figure 6. Structure of optimal portfolios for the quarterly interval



Note. Minimum variance portfolio – left panels, maximum Sharpe ratio portfolio – right panels, raw weights – upper panels, recalculated weights – lower panels. The results relate to the whole sample.

Source: authors' work.

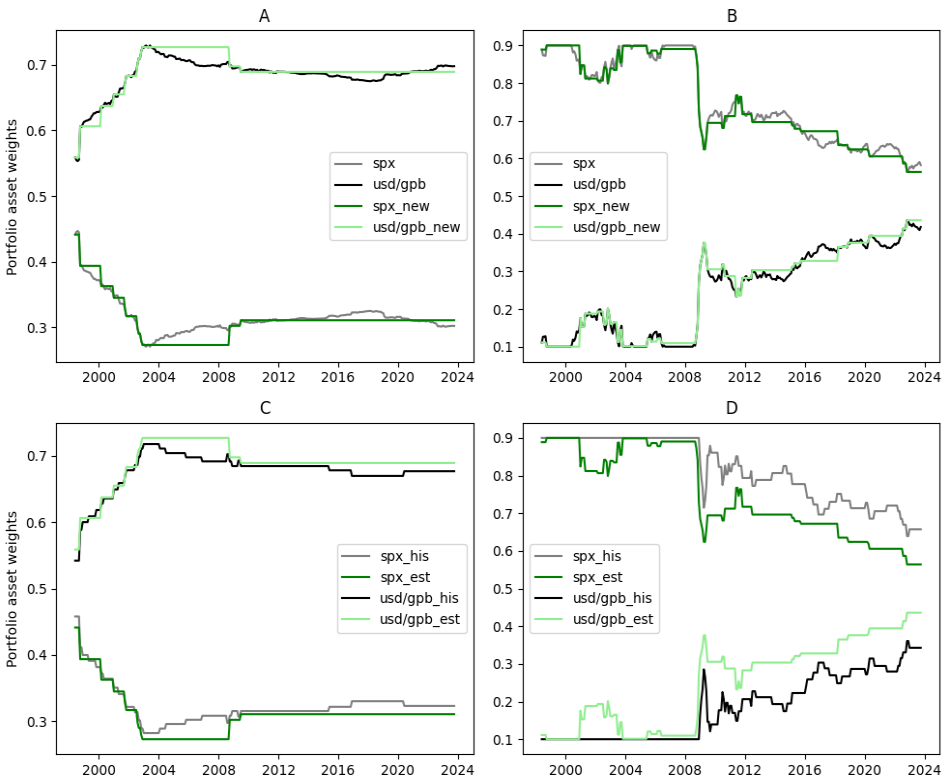
Furthermore, the graphs in the second row of Figure 6 show the adjusted portfolio structure based on the estimated closing prices, overlaid with the structure based on historical data. Due to the criteria applied, the weights of the optimal portfolios largely overlap with the weights resulting from the historical values. The root sum of the squares of the differences in the weights calculated based on price prediction and historical values is 8% and 11%, respectively. These are low values, which indicate correct estimation results and fairly small deviations from the benchmark.

Figure 7 illustrates the structure of the portfolios established for monthly intervals. As the interval is shortened, a significantly higher volatility of the structure is observable in both portfolios compared to the quarterly portfolios.

Minimum variance portfolio A was at first characterized by an almost equal distribution of assets which tended to assign more weight to the currency pair and maintain that level until the end of the considered period. In this case, the portfolio

is slightly less diversified and the weights move in ranges close to the 7:3 ratio for the currency pair.

Figure 7. Structure of optimal portfolios for the monthly interval



Note. Minimum variance portfolio – left panels, maximum Sharpe ratio portfolio – right panels, raw weights – upper panels, recalculated weights – lower panels. The results relate to the whole sample.

Source: authors' work.

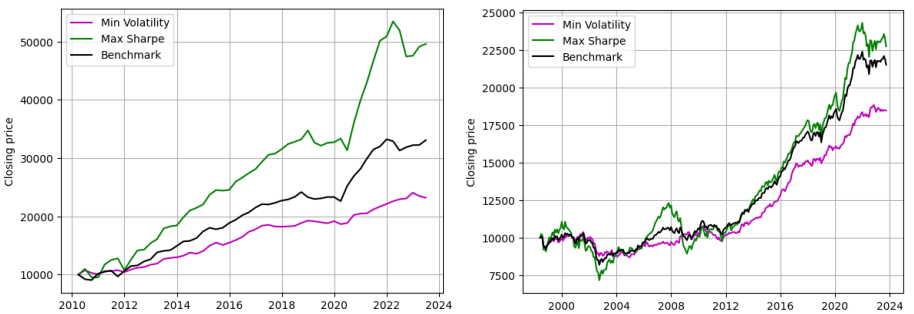
For portfolio (B), and similarly for (D), a higher stock market index weight was observed again, which held the maximum possible weight for almost a quarter of the examined period. In the monthly time interval, there was much more variance in the portfolio structures than in the longer period. This underestimation is noticeable in the minimum variance portfolio and it occurs throughout the whole studied period. In the case of portfolio B, however, the significant changes in the structure are more likely to be temporary and occur towards the end of the period.

In this case, the root sum of the squares of the differences between the weights calculated based on price prediction and historical prices was 2% and 8%. The minimum variance portfolio shows a significant improvement of this parameter.

In contrast, the portfolio maximizing the Sharpe ratio, despite the more abrupt changes in the weights, slightly adjusts the level of variance compared to the previous interval.

The resulting portfolios were also compared to a benchmark, which was assumed to be a portfolio with equal instrument weights throughout the period under study. Therefore, a kind of simulation was performed assuming the investment of 10,000 monetary units in both portfolios and the benchmark. The results are presented in Figure 8.

Figure 8. Capital value over time



Note. Quarterly – left panel, monthly – right panel. The results concern the whole sample.
Source: authors' work.

The portfolio results are shown in Table 3. The smallest portfolio variance occurred in Portfolio 1 and the largest in Portfolio 2. Portfolio 3 represents a kind of arithmetic average in this respect, which is related to the structure of this portfolio. Surprisingly, the largest Sharpe ratio was achieved by Portfolio 1, which should theoretically be the third-best portfolio in this metric. Given that this metric combines a portfolio's return with its risk, there can be some kind of anomaly or discrepancy between the expected and received result when making comparisons and in terms of extreme values. In this case, this is probably due to the three times lower variance in relation to Portfolio 2, which was in the denominator when calculating this ratio.

The highest overall rate of return of approximately 500% was achieved by Portfolio 2, thus achieving double the performance of Portfolio 1. The Sharpe ratio, maximizing the portfolio also recorded the highest average return of 3.27% from period to period. This seems a logical consequence of achieving such a high overall return. This portfolio, however, did not perform as well as the others in terms of portfolio variance. This is where the variance minimizing portfolio performed best, while the benchmark underperformed by just 7 percentage points.

Table 3. Portfolio assessment metrics

Metrics	Portfolio 1 Min variance	Portfolio 2 Max Sharpe ratio	Portfolio 3 Benchmark (50:50)
quarterly			
Portfolio variance	9.43%	27.43%	16.17%
Sharpe ratio	59.62	49.73	56.77
Rate of return	231.96%	496.18%	330.89%
Average rate of return (quarterly)	1.64%	3.27%	2.37%
Maximum drawdown	6.63%	15.87%	9.47%
Tracking error	0.36%	0.37%	–
monthly			
Portfolio variance	1.77%	2.95%	2.36%
Sharpe ratio	14.87	12.12	15.91
Rate of return	184.80%	232.42%	215.01%
Average rate of return (monthly)	0.21%	0.30%	0.26%
Maximum drawdown	16.13%	35.25%	20.61%
Tracking error	0.04%	0.6%	–

Note. The results relate to the whole sample.

Source: authors' work.

The smallest maximum drawdown was achieved by Portfolio 1, but it should be noted that there were no large divergences in relation to Portfolio 3. The variance minimizing portfolio (Portfolio 1) is the closest to the benchmark, where on average its one-period returns deviated from the benchmark rates by 0.36%. In contrast, Portfolio 2 outperforms both the benchmark and the variance minimizing portfolio.

Comparing the graphs in Figure 8 and focusing on the period from 2010, it becomes apparent that there was a definite smoothing of any price peaks and lows by the quarterly portfolios and a greater divergence between the portfolios compared to the shorter interval.

In contrast to the previous interval, in this case, the portfolio maximizing the Sharpe ratio significantly outperformed the other two portfolios only near the years 2000 and 2008. The higher capital value in these periods was due to the large price fluctuations of both assets. Around 2002, these were caused by the 9/11 attacks and the bursting of the speculative internet bubble (involving the overvaluation of IT companies). In the second period mentioned, stock market falls and the weakening of the dollar after the subprime crisis were the reasons. During the recession following these events, only two periods occurred where Portfolio 2 was outperformed by both the minimizing variance portfolio and the benchmark.

In the monthly time interval, there were no major differences in terms of which portfolio performed best on a given metric compared to the quarterly interval. On the other hand, despite the shorter intervals, portfolio variance improved significantly

by up to nine times the variance of the quarterly portfolios. Due to the shortening of the time interval, it is natural for the one-period portfolio return to decrease as well. The largest change between the considered intervals was in the metric of the average rate of return, where a decrease of up to ten times was observed for Portfolio 2. Despite the lower volatility, the tracking error of the optimized portfolios increased. In addition, Figure 8 indicates much smaller divergences between portfolios in the quarterly interval than in the monthly one.

The above graphs of capital change over time and the calculated evaluation metrics show that the portfolio maximizing the Sharpe ratio offers the highest return but at the price of the highest volatility. The minimizing variance portfolio instead offers more stable but slower capital growth. However, the benchmark set places itself between the two portfolios, combining and averaging the advantages and disadvantages of both investment portfolios. Ultimately, the benchmark in the monthly interval considerably outperformed the minimizing variance portfolio.

4. Conclusions

In this study, we proposed a framework to build a two-asset portfolio. For this purpose, we combined machine learning algorithms based on trees (RF and XGBoost) with methods optimizing portfolio performance. The framework was illustrated in the periods from Q1 1985 to Q2 2023 (quarterly intervals) and from January 1990 to September 2023 (monthly intervals), showing promising results.

Based on our case study, for the selected assets and time frame, the analysis of the research questions led to several conclusions.

The macroeconomic variables used in the modeling allowed the models to estimate the time series efficiently. The assessment of the significance of these variables confirms that some of the variables derived from the economic theory and the literature review have a particularly significant influence on the estimation results, especially money supply aggregates and inflation rates.

The structure of the constructed portfolio based on the estimated data does not differ significantly from the structure of the portfolio based on the realized return rates. These deviations were measured by the mean square error between the received and historical structure. The obtained result involved deviations ranging from 2 to 12 percentage points.

The return of the resulting portfolios, which is also the most important portfolio evaluation metric, ranged from around 200% to just over 500% depending on the optimized values in the portfolio. The portfolio maximizing the Sharpe ratio obtained, on average, a better return than the portfolio minimizing variances by about one-third of the portfolio results. Of course, this was associated with

a significantly higher risk, as measured by the portfolio variance, which was on average about three times higher. It came as a surprise that in the conducted study, the portfolio maximizing the Sharpe ratio did not score the best values in this metric. This is related to the higher contribution of capital to a riskier asset which outweighed the achieved high return and resulted in lower scores in this metric. It is also worth mentioning that, despite the highest return, this portfolio did not always produce the best results. Indeed, in the monthly interval, there were two periods in which the value of the evenly spread assets portfolio was the highest.

Finally, summarizing the obtained results, this case study shows that it is possible to efficiently construct a two-component portfolio using macroeconomic data and machine learning methods. The research problem explored in this study should be further developed in future work by expanding the scope to include a greater number of markets and investment portfolio components, as well as by employing a broader range of ML methods.

References

Abreu, G., Neves, R., & Horta, N. (2018). *Currency exchange prediction using machine learning, genetic algorithms and technical analysis*. https://doi.org/10.48550/arXiv.1805.11232.

Ahmed, N. K., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5–6), 594–621. https://doi.org/10.1080/07474938.2010.481556.

Baasher, A. A., & Fakhr, M. W. (2011). FOREX Trend Classification using Machine Learning Techniques. In A. Zaharim, K. Sopian, N. Mastorakis & V. Mladenov (Eds.), *ACS'11: Proceedings of the 11th WSEAS International Conference on Applied Computer Science* (pp. 41–47). World Scientific and Engineering Academy and Society. https://www.wseas.us/e-library/conferences/2011/Penang/ACRE/ACRE-05.pdf.

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013.

Boyoukliev, I. V., Kulina, H. N., & Gocheva-Ilieva, S. G. (2022). Modelling and Forecasting of EUR/USD Exchange Rate Using Ensemble Learning Approach. *Cybernetics and Information Technologies*, *22*(4), 142–151. https://doi.org/10.2478/cait-2022-0044.

Cabrera Malik, S. (2024). *Is xgboost immune to multicolinearity?*. https://medium.com/@sebastian.cabrera-malik/is-xgboost-inmune-to-multicolinearity-4dd9978605b7.

Charteris, A., & McCullough, K. (2020). Tracking error vs tracking difference: Does it matter? *Investment Analysts Journal*, *49*(3), 269–287. https://doi.org/10.1080/10293523.2020.1806480.

Choi, J. (2021). Maximum Drawdown, Recovery, and Momentum. *Journal of Risk and Financial Management*, *14*(11), 1–25. https://doi.org/10.3390/jrfm14110542.

Choudhry, R., & Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *Proceedings of World Academy of Science, Engineering and Technology*, *29*, 315–318. https://par.cse.nsysu.edu.tw/resource/paper/2008/080820/A%20Hybrid%20Machine%20Learning%20System%20for%20Stock%20Market%20Forecasting.pdf.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly Media.

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1.

Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). *A Machine Learning Model for Stock Market Prediction*. https://doi.org/10.48550/arXiv.1402.7351.

Kaszuba, B. (2011). Praktyczne problemy w teorii portfela. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (174), 381–393.

Kaushik, M., & Giri, A. K. (2020). *Forecasting Foreign Exchange Rate: A Multivariate Comparative Analysis between Traditional Econometric, Contemporary Machine Learning & Deep Learning Techniques*. https://doi.org/10.48550/arXiv.2002.10247.

Król-Nowak, A., & Kotarba, K. (2022). *Podstawy uczenia maszynowego*. Wydawnictwa Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. https://winntbg.bg.agh.edu.pl/skrypty4/0599/podstawy_uczenia.pdf.

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2A), 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017.

Liu, Y. (2023). Government debt and risk premia. *Journal of Monetary Economics*, *136*, 18–34. https://doi.org/10.1016/j.jmoneco.2023.01.009.

Loh, L. K. Y., Kueh, H. K., Parikh, N. J., Chan, H., Ho, N. J. H., & Chua, M. C. H. (2022). An Ensembling Architecture Incorporating Machine Learning Models and Genetic Algorithm Optimization for Forex Trading. *FinTech*, *1*(2), 100–124. https://doi.org/10.3390/fintech1020008.

Łuniewska, M. (2012). *Ekonometria finansowa. Analiza rynku kapitałowego*. Wydawnictwo Naukowe PWN.

Mabrouk, N., Chihab, M., & Chihab, Y. (2022). *A Trading Strategy in the Forex Market based on Linear and Non-linear Machine Learning Algorithms*. https://www.scitepress.org/Papers/2021/107288/107288.pdf.

Magdon-Ismail, M., & Atiya, A. F. (2004). Maximum Drawdown. *Risk Magazine*, *17*(10), 99–102.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.2307/2975974.

Matuszewska-Janica, A., & Witkowska, D. (2008). Modelowanie kursu euro/dolar: dynamiczne modele ekonometryczne i sztuczne sieci neuronowe. *Zeszyty Naukowe SGGW – Ekonomika i Organizacja Gospodarki Żywnościowej*, (69), 55–75. https://doi.org/10.22630/EIOGZ.2008.69.92.

Maung, K., & Swanson, N. R. (2025). A survey of models and methods used for forecasting when investing in financial markets. *International Journal of Forecasting*, *41*(4), 1355–1382. https://doi.org/10.1016/j.ijforecast.2025.03.002.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & Shahab, S. (2020a). Deep Learning for Stock Market Prediction. *Entropy*, *22*(8), 1–23. https://doi.org/10.3390/e22080840.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020b). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, *8*, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031.

Pham, N., Cui, B., & Ruthbah, U. (2025). *The Performance of the 60/40 Portfolio: A Historical Perspective*. https://rpc.cfainstitute.org/research/reports/2025/performance-of-the-60-40-portfolio.

Qi, L., Khushi, M., & Poon, J. (2020). Event-Driven LSTM for Forex Price Prediction. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Australia, 1–6. https://doi.org/10.1109/CSDE50874.2020.9411540.

Reddy, V. K. S. (2018). Stock Market Prediction Using Machine Learning. *International Research Journal of Engineering and Technology*, *5*(10), 1032–1035. https://www.irjet.net/archives/V5/i10/IRJET-V5I10193.pdf.

Roßbach, P. (2018). *Neural Networks vs. Random Forests – Does it always have to be Deep Learning*. https://blog.frankfurt-school.de/wp-content/uploads/2018/10/Neural-Networks-vs-Random-Forests.pdf.

Roy, M.-H., & Larocque, D. (2012). Robustness of random forests for regression. *Journal of Nonparametric Statistics*, *24*(4), 993–1006. https://doi.org/10.1080/10485252.2012.715161.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, *39*(1), 119–138.

Sharpe, W. F. (1998). The Sharpe ratio. In P. L. Bernstein & F. J. Fabozzi (Eds.), *Streetwise. The Best of The Journal of Portfolio Management* (pp. 169–178). Princeton University Press. https://doi.org/10.2307/j.ctv1mjqtwg.24.

Shen, S., Jiang, H., & Zhang, T. (2012). *Stock Market Forecasting Using Machine Learning Algorithms*. Stanford University.

Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tnag, C., & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing*, *512*, 363–380. https://doi.org/10.1016/j.neucom.2022.09.003.

Țițan, A. G. (2015). The Efficient Market Hypothesis: review of specialized literature and empirical research. *Procedia Economics and Finance*, *32*, 442–449. https://doi.org/10.1016/S2212-5671(15)01416-1.

Wong, A., Whang, S., Sagre, E., Sachin, N., Dutra, G., Lim, Y. W., Hains, G, Khmelevsky, Y., & Zhang, F. C. (2023, December). *Short-term stock price forecasting using exogenous variables and machine learning algorithms* [paper presentation]. 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Denpasar, Indonesia. https://doi.org/10.1109/ICICyTA60173.2023.10428814.

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, *5*(1), 1–20. https://doi.org/10.1186/s40854-019-0138-0.

Appendix

Table A1. Models' hyperparameters

	Model	Monthly interval		Quarterly interval	
		Parameter	Value	Parameter	Value
SPX	XGBoost	Objective	Reg: squarederror	Objective	Reg: squarederror
		max_depth learning_rate n_estimators num_boost_round	10 0.66 100 10	max_depth learning_rate n_estimators	4 1 100
	RF	n_estimators random_state	100 24	max_depth min_samples_leaf min_samples_split n_estimators random_state	10 1 8 300 24
USD/GBP	XGBoost	Objective	Reg:squarederror	Objective	Reg:squarederror
		max_depth learning_rate n_estimators	10 0.9 100	max_depth learning_rate n_estimators	4 0.6 100
	RF	n_estimators random_state	14 25	n_estimators random_state	135 24

Note. The algorithm implementations used in this study are provided by the following Python libraries: XGBoost, sklearn.ensemble, sklearn.metrics, sklearn.preprocessing, and sklearn.model_selection.

Source: authors' work.

Table A2. List of independent variables

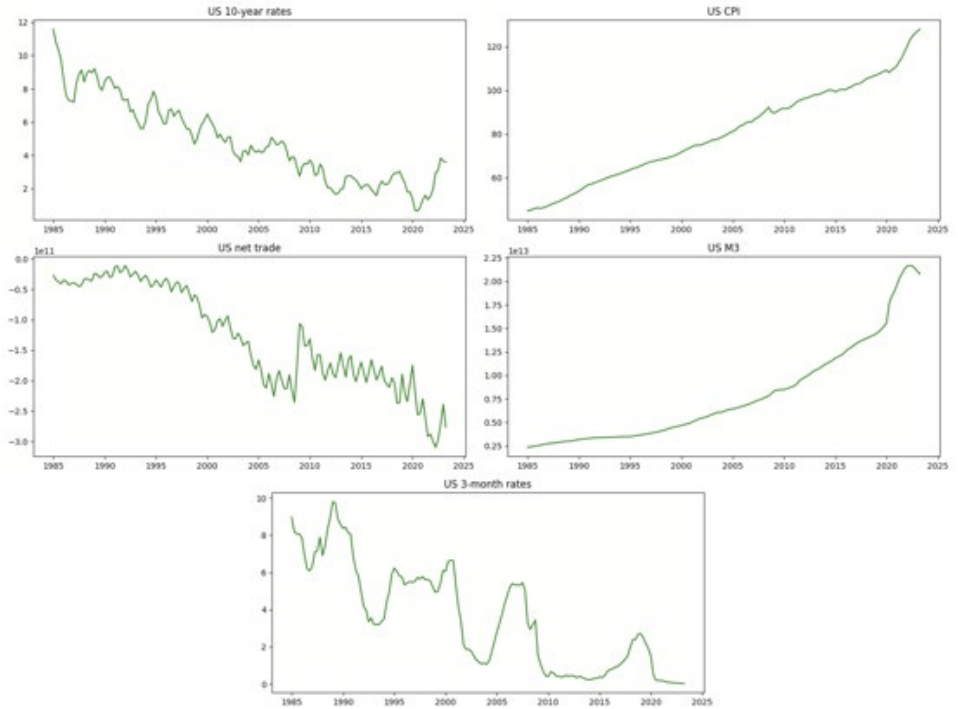
Short name	Description	Source study
	Quarterly S&P 500	
US 10-year rates	10-year government bond yield for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US net trade	Commodities Trade Balance for United States	(Kaushik & Giri, 2020)
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
US 3-month rates	3-month interest rate for United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
	Quarterly USD/GBP	
US labor inactivity rate	Labor inactivity rate for persons aged 15 and older in the United States	(Boyoukliev et al., 2022)
US import	Real Imports of Goods and Services for the United States	(Kaushik & Giri, 2020)
US export	Real Exports of Goods and Services for the United States	(Kaushik & Giri, 2020)
US GBP	Real Gross Domestic Product for the United States	(Liu, 2023)
WB labor inactivity rate	Labor inactivity rate for persons aged between 25 and 54 in the United Kingdom	(Boyoukliev et al., 2022)
WB CPI	Consumer price index for the United Kingdom, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
WB M1	M1 aggregate money supply in the United Kingdom	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB public debt	Total credit to general government for the United Kingdom (credit covers loans and debt securities)	(Liu, 2023)
	Monthly S&P 500	
US 10-year rates	10-year government bond yield for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US net trade	Commodities Trade Balance for the United States	(Kaushik & Giri, 2020)

Table A2. List of independent variables (cont.)

Short name	Description	Source study
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
US annual rates	Annual interest rate for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
Monthly USD/GBP		
US effective exchange rate	Real effective exchange rates are calculated as weighted averages of bilateral exchange rates adjusted by the relative consumer prices	(Shen et al., 2012), (Zhong & Enke, 2019)
US labor participation rate	Labor force participation rate for persons aged 15 and older in the United States	(Boyoukliev et al., 2022)
US reserves, excluding gold	Total reserves, excluding gold for the United States	(Kaushik & Giri, 2020)
US retail sales	Total retail trade value in the United States	(Kaushik & Giri, 2020)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB M1	M1 aggregate money supply in the United Kingdom	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB reserves, excluding gold	Total reserves, excluding gold for the United Kingdom	(Kaushik & Giri, 2020)
WB effective exchange rate	Real effective exchange rates are calculated as weighted averages of bilateral exchange rates adjusted by the relative consumer prices	(Shen et al., 2012), (Zhong & Enke, 2019)
WB CPI	Consumer price index for the United Kingdom, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
WB net trade	Commodities trade balance for the United Kingdom	(Kaushik & Giri, 2020)

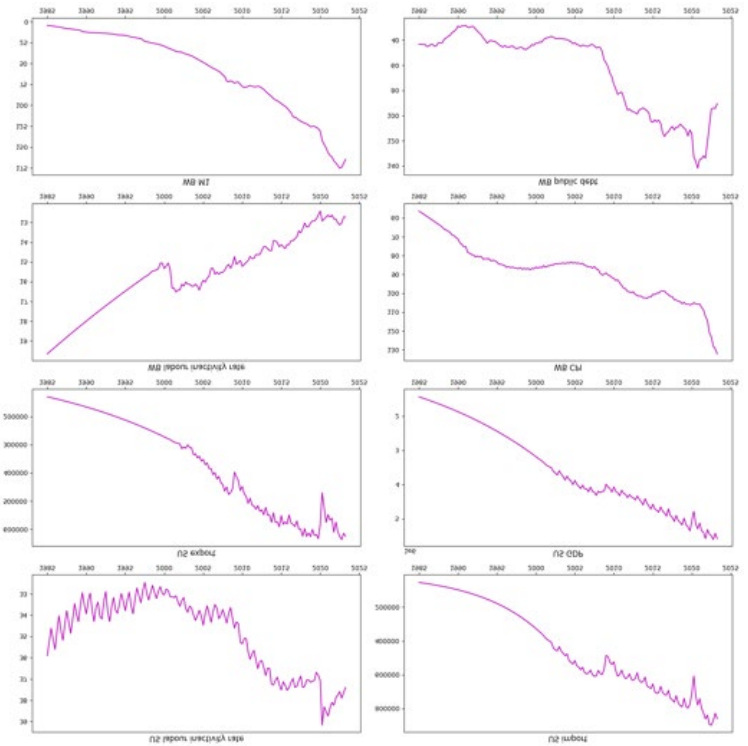
Source: authors' work.

Figure A1. Independent variables used in quarterly S&P 500 estimation



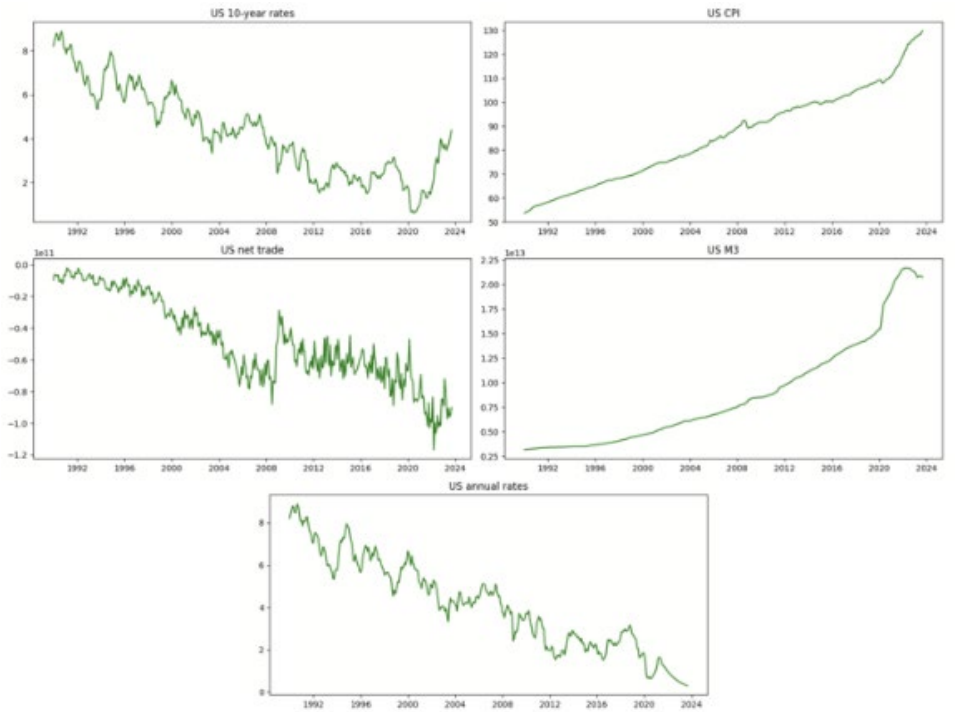
Source: authors' work.

Figure A2. Independent variables used in quarterly USD/GBP estimation



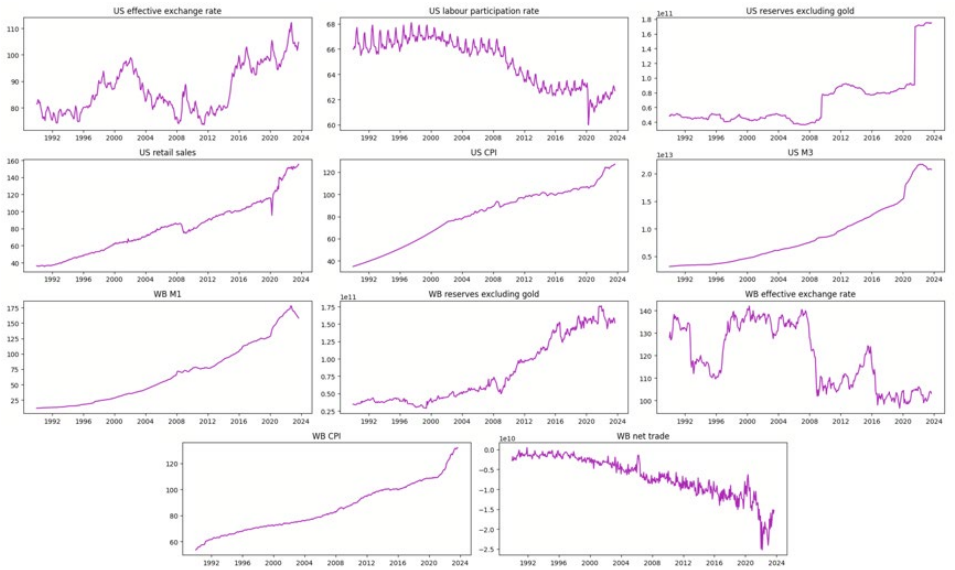
Source: authors' work.

Figure A3. Independent variables used in monthly S&P 500 estimation



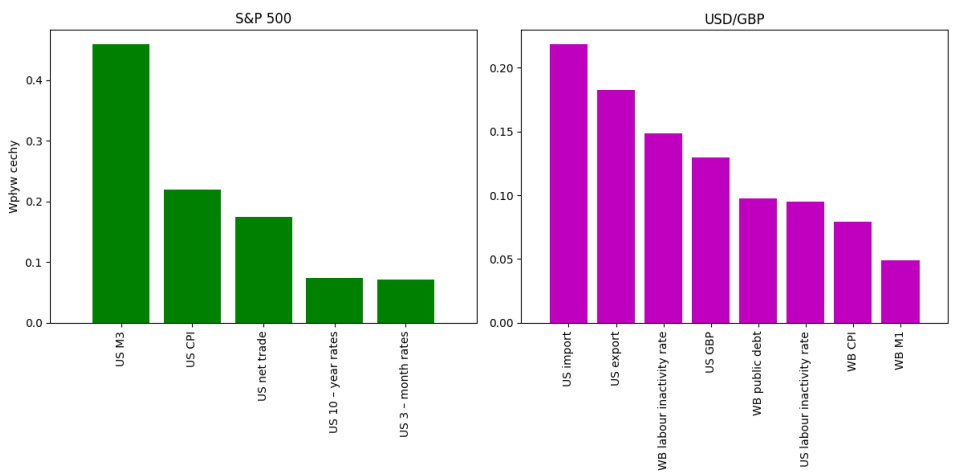
Source: authors' work.

Figure A4. Independent variables used in monthly USD/GBP estimation



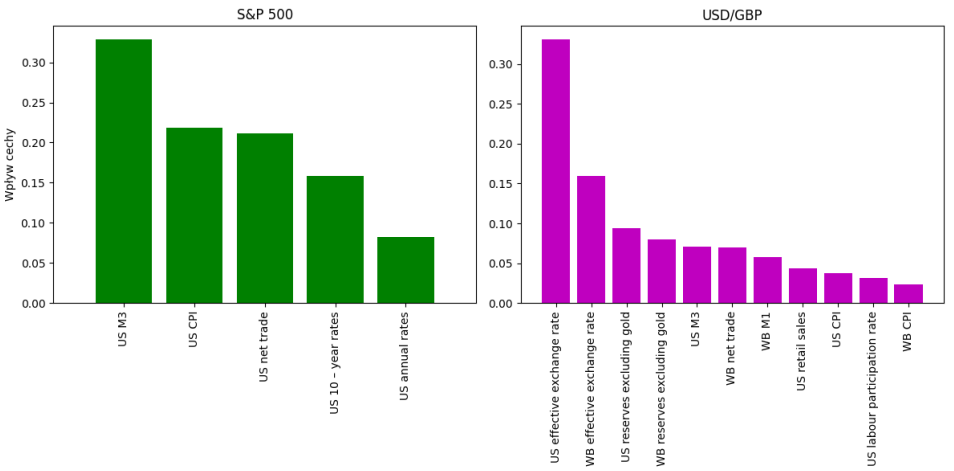
Source: authors' work.

Figure A5. Feature importance of variables used in the quarterly estimation



Source: authors' work.

Figure A6. Feature importance of variables used in the monthly estimation



Source: authors' work.