Chilean Journal of Statistics
Vol. 16, No. 2, December 2025, 217–233
DOI: 10.32372/ChJS.16-02-04

TIME-SERIES MODELS
RESEARCH ARTICLE

On the use of the concentration function to compare predictive distributions in ARMA models

CRISTIAN CRUZ-TORRES[1,*] and FABRIZIO RUGGERI[2]

[1]Department of Statistics, National Autonomous University of Honduras, Tegucigalpa, Honduras
[2]Institute of Applied Mathematics and Information Technology, National Research Council, Milano, Italy

(Received: 24 October 2025 · Accepted: 28 November 2025)

Abstract

In this study, we propose the use of the concentration function as a novel tool for assessing the adequacy of predictive distributions in a Bayesian framework. The methodology relies on a rolling-origin evaluation strategy, in which the available time series is partitioned into two segments: the first is used for model estimation, while the second is used for out-of-sample prediction. The concentration function is then constructed for each set of forecasts, providing a distributional assessment of predictive performance across competing models. In this framework, the model with the least inequality in its concentration function is interpreted as offering superior predictive fit. To complement this curve-based analysis, we compute two well-established inequality measures—the Gini concentration coefficient and the Pietra index— which provide further evidence and consistency regarding model performance. These measures allow for a comprehensive evaluation of how closely predictive distributions align with observed outcomes. The proposed methodology is validated through a simulation study, in which its ability to discriminate between models is assessed under controlled conditions. We illustrate its practical applicability by analyzing real-world data, thus demonstrating the usefulness of concentration functions and associated inequality measures as diagnostic tools for Bayesian predictive evaluation.

Keywords: ARMA models · Concentration function · Gini coefficient · Predictive distribution.

Mathematics Subject Classification: Primary 62M10 · Secondary 62F15.

*Corresponding author. Email: cristian.cruz@unah.edu.hn (C. Cruz-Torres)

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

1. Introduction

The concentration function generalizes the concept of the Lorenz curve [1], a graphical tool designed to illustrate the distribution of a variable, most commonly income or wealth, within a given population. The concept was first introduced by the American economist Max O. Lorenz in 1905 and has since become a standard tool in the analysis of inequality and resource allocation. Its main strength lies in its ability to provide an intuitive visualization of the degree of inequality present in a distribution.

The construction of the Lorenz curve begins with a dataset representing the distribution of the variable of interest (such as household income). The observations are arranged in ascending order, from the lowest to the highest values. For each ordered group, the cumulative share of the variable is calculated as the proportion of the total sum accounted for up to that point. Simultaneously, the cumulative proportion of the population is determined, corresponding to the share of individuals or households represented in each group.

By plotting the cumulative share of the population on the horizontal and this share of the variable on the vertical axis, the Lorenz curve is obtained. Each point along the curve represents the cumulative distribution at a specific percentile of the population. A perfectly equal distribution corresponds to a 45-degree line (the line of perfect equality), where each cumulative share of the population holds an equivalent cumulative share of the variable. In practice, Lorenz curves typically lie below this line, signaling the presence of inequality. The magnitude of this deviation can be summarized by the Gini coefficient [2], a scalar measure of inequality derived from the Lorenz curve. The Gini coefficient takes values between 0, representing perfect equality, and 1 in the population limit, corresponding to the extreme case in which all of the variable is concentrated in a single individual.

Autoregressive Moving Average (ARMA) models are classical statistical models widely used for analyzing and forecasting time series data. They combine two components: the autoregressive (AR) part, which captures the dependence of the current observation on its past values, and the moving average (MA) part, which models the dependence on past forecast errors or shocks. Formally, the autoregressive component assumes that the current value of a series can be expressed as a linear function of its lagged observations, while the moving average component represents the current value as a linear combination of present and past innovations. ARMA models have been widely applied in various fields, including economics, finance, engineering, and environmental sciences, due to their ability to parsimoniously capture temporal dependence. Importantly, ARMA models require the underlying time series to be stationary, meaning that its mean, variance, and autocovariance structure remain constant over time. In cases where non-stationarity is present, transformations such as differencing or variance-stabilizing methods are typically applied before model fitting.

In this study, we propose using the concentration function to evaluate the predictive distributions generated by ARMA models. Specifically, we construct a concentration function based on the predictive likelihood ratio of two competing ARMA models. Deviations from the line of perfect equality in this curve indicate the presence of inequality, which, in turn, reveals that a given observation is more likely to be forecast by one model rather than the other. This approach allows us to compare models not only in terms of point forecasts but also with respect to the distributional properties of their predictive performance.

The remainder of the article is organized as follows. Section 2 provides a literature review on concentration functions and related measures. In Section 3, we present the formal definition of the concentration function. In Section 4, the general methodology is introduced for applying the concentration function to predictive distributions. Section 5 offers a brief overview of ARMA models, while Section 6 describes the simulation study designed to validate the proposed procedure. In Section 7, we illustrate the methodology with an application to gas prices and Section 8 summarizes our findings.

2. Review of Literature

The Lorenz curve, originally developed to measure income inequality, has been widely adapted in various fields as a general tool for assessing distributional inequality. In health economics, it forms the basis of the concentration curve and concentration index, which are used to assess the distribution of health outcomes or healthcare utilization among socioeconomic groups [3]. These methods allow researchers to capture inequalities beyond income by showing whether poor health disproportionately affects disadvantaged populations. Similarly, in environmental justice research, Lorenz-type dominance methods have been applied to analyze inequalities in the distribution of pollution and environmental risks. Empirical studies have demonstrated unequal exposure to air pollution among population subgroups by comparing cumulative population shares with cumulative environmental burdens [4, 5].

In credit risk modeling, Lorenz-type curves appear in the form of cumulative accuracy profiles (CAP), closely related to the Gini coefficient and the ROC–AUC measure. These curves provide a graphical means of assessing the discriminatory power of scoring models, with stronger curvature indicating greater predictive accuracy [6]. In ecology, Lorenz curves are used to quantify inequality in species abundances, with direct connections to concepts such as dominance, diversity, and entropy [7]. Beyond the natural sciences, bibliometrics employs Lorenz-type methods to describe the unequal distribution of scientific citations and journal productivity [8]. Theoretical developments have further expanded these ideas: multivariate Lorenz curves generalize the approach to multidimensional attributes, while in probability theory, the Lévy concentration function provides a related but distinct concept that measures the clustering of random variables within intervals [9]. These studies highlight that Lorenz-type curves are not confined to analyzing income inequality; instead, they serve as a unifying framework for representing unevenness, concentration, and dominance across a wide range of domains. Whether applied to health outcomes, environmental burdens, model performance, biodiversity, or scientific outputs, these curves provide a powerful geometric tool for visualizing and comparing distributions across contexts.

Evaluating forecasts and comparing predictive models in time series are central activities in econometrics and statistics. The classic foundation lies in measures of forecast accuracy, such as mean squared error, mean absolute error, and mean absolute percentage error. Early surveys, such as [10], established systematic frameworks for benchmarking predictive methods. A key advance was introduced in [11], who developed a widely used statistical test for comparing predictive accuracy between models. Their contribution formalized the idea that forecast errors can be interpreted as loss functions and evaluated according to asymptotic theory, thereby providing a rigorous basis for model comparison.

Subsequent research expanded this framework to address the limitations of standard procedures. For instance, in [12, 13, 14], forecast comparison methods were extended to the case of nested models, where conventional tests tend to be biased in favor of larger specifications. To address the challenges in multi-model comparisons, in [15], it was proposed the reality check, and in [16], it was developed the superior predictive ability test, both of which can mitigate data snooping problems. Complementary approaches such as the model confidence set presented in [17] provide formal procedures for identifying subsets of models that cannot be statistically distinguished in terms of predictive performance.

Choice of the evaluation method depends critically on the loss function and the decision context. While symmetric quadratic losses remain standard, alternative loss functions, including quantile-based, asymmetric, and utility- or value-based measures, have been developed for applications in risk management, finance, and decision making [18]. In addition to point forecasts, recent advances emphasize the evaluation of density and probabilistic forecasts. Tools such as the probability integral transform, logarithmic scoring rules, and other scoring rules have become essential for assessing predictive distributions [19, 20].

In parallel, Bayesian econometrics has developed a complementary perspective on forecast evaluation by directly focusing on predictive distributions. Central to this approach is the log predictive density, which provides a natural scoring rule consistent with Bayesian decision theory [21].

Related measures include Bayes factors and the log predictive score, which allow formal comparisons between competing models on the basis of their predictive likelihoods [22]. Bayesian forecast evaluation also incorporates posterior predictive checks, which assess the calibration of predictive distributions relative to observed data. More recent contributions propose simulation-based tools, such as predictive Bayes factors and cross-validation criteria (for example, LOO and WAIC), that offer flexible methods for model comparison even in high-dimensional or nonlinear settings [23, 24]. Related Bayesian approaches to detecting structural changes in autoregressive time series models are discussed in [25]. Both the frequentist and Bayesian literatures provide a rich and complementary set of tools for evaluating time series forecasts, with Bayesian methods offering particular advantages when the full distribution of uncertainty is of interest.

Although existing methods offer powerful tools for assessing forecast accuracy, they are not without limitations. Classical approaches often focus on point forecasts or rely heavily on specific loss functions, while Bayesian measures such as log predictive density or Bayes factors, though theoretically grounded, can be sensitive to prior choices and may lack intuitive interpretability in applied contexts. To complement these approaches, we propose the use of the concentration function as a new tool for comparing predictive models in a Bayesian setting. By representing predictive likelihood ratios through Lorenz-type curves, the concentration function provides a geometric and distributional perspective on model performance. This framework allows researchers to assess not only average predictive fit but also the extent of inequality in how competing models distribute probability mass across outcomes. Thus, it bridges the literature on inequality measurement with Bayesian forecast evaluation, offering a new perspective that is both rigorous and intuitively interpretable.

3. Concentration function and Lorenz curve

The Lorenz curve is a graphical tool used to describe the discrepancy between a discrete probability measure Π and a discrete uniform measure Π_0. Its classical application concerns the comparison between the actual income distribution in a population (Π) and a hypothetical scenario in which income is evenly distributed across individuals (Π_0). Formally, the Lorenz curve is obtained by plotting the cumulative wealth of the poorest fraction of the population against its population share.

Consider a population of n individuals, where x_i denotes the wealth of individual i, for $i = 1, \ldots, n$. Ordering individuals from poorest to richest yields $x_{(1)}, \ldots, x_{(n)}$. Define $S_0 = 0$ and $S_k = \sum_{i=1}^{k} x_{(i)}$, the cumulative wealth of the poorest k individuals. Then, the Lorenz curve is represented by the sequence of points stated as $(k/n, S_k/S_n)$ for $k = 0, \ldots, n$, which are connected by line segments to form a continuous curve joining the origin $(0, 0)$ with the point $(1, 1)$. If wealth is uniformly distributed, the Lorenz curve coincides with the 45° line of equality. Otherwise, the curve is convex and lies strictly below the line of equality, reflecting inequality in the distribution of wealth [26].

To summarize inequality, we employ two classical measures from the inequality literature. The Gini concentration area [2], which is equal to one half of the usual Gini coefficient, measures the area between the Lorenz curve and the line of equality given by

$$A_G = \frac{n+1}{2n} - \frac{1}{n} \sum_{k=1}^{n} \frac{S_k}{S_n},$$

where S_k is the cumulative sum of the ordered wealth levels.

The corresponding Gini coefficient is $G = 2A_G$, which ranges between 0, representing perfect equality, and 1, representing maximal inequality in the population limit. The Pietra index [27] captures the maximum vertical deviation from the line of equality formulated as

$$P = \sup_{1\le k\le n-1} \left(\frac{k}{n} - \frac{S_k}{S_n} \right).$$

These indices can be used to order income distributions in terms of inequality: smaller values indicate that a distribution is closer to perfect equality (with respect to the uniform reference), whereas larger values reflect stronger concentration of wealth.

In a more general probability space $(\Omega, \mathcal{A})$, the concentration function of Π relative to Π_0 is defined for any $x \in [0,1]$ as $\varphi(x) = \inf\{\Pi(A)\colon A \in \mathcal{A} \text{ and } \Pi_0(A) \ge x\}$. Intuitively, $\varphi(x)$ provides a lower bound on set probabilities, capturing the minimum proportion of Π's mass that can be allocated to subsets with Π_0-measure at least x. When Π_0 is nonatomic, this construction yields a function φ that is defined for all $x \in [0,1]$ and has additional regularity properties; a more general but equivalent formulation is provided in [28].

The concentration function possesses several useful properties: it is nondecreasing and convex. Moreover, when Π_0 is nonatomic, φ is continuous on $[0,1]$. In addition, we have that

$$\varphi(x) = 0,\ \forall x \in [0,1] \Leftrightarrow \Pi \perp \Pi_0, \qquad \varphi(x) = x,\ \forall x \in [0,1] \Leftrightarrow \Pi = \Pi_0.$$

The concentration is equal to zero on the entire interval $[0,1]$ if and only if Π is singular with respect to Π_0. An example is the concentration function of a discrete distribution with respect to a continuous one. Two illustrative cases highlight the interpretive value of the concentration function. First, if $\varphi(1) = 1$, then Π is absolutely continuous with respect to Π_0. Conversely, if $\varphi(x) = 0$ for all $0 \le x \le a$, then Π assigns zero mass to a subset $A \in \mathcal{A}$ such that $\Pi_0(A) = a$.

Properties of the coefficients of divergence in this framework were studied in [29] and given by

$$G_{\Pi_0}(\Pi) = 2\int_0^1 (x - \varphi(x))\,\mathrm{d}x, \qquad C_{\Pi_0}(\Pi) = \sup_{x\in[0,1]} (x - \varphi(x)),$$

which correspond, respectively, to the Gini concentration coefficient (twice the concentration area) and the Pietra index. The associated concentration area is $G_{\Pi_0}(\Pi)/2$, taking values in the interval $[0, 0.5]$. The latter coefficient $C_{\Pi_0}(\Pi)$ has been shown to be equal to the total variation norm [28].

The divergence coefficients vanish when $\Pi = \Pi_0$ and attain their maximum value when Π and Π_0 are mutually singular, that is, when

$$C_{\Pi_0}(\Pi) = 0 \Leftrightarrow G_{\Pi_0}(\Pi) = 0 \Leftrightarrow \Pi = \Pi_0, \quad C_{\Pi_0}(\Pi) = 1 \Leftrightarrow G_{\Pi_0}(\Pi) = 1 \Leftrightarrow \Pi \perp \Pi_0.$$

In Figure 1 we show an example of the Lorenz curve and the concentration function. Panel (a) shows the construction of the Lorenz curve from cumulative population and measure shares Π. Panel (b) illustrates inequality measures: the shaded area corresponds to the Gini concentration area, while the maximal vertical distance $x - \varphi(x)$ over $x \in [0,1]$ corresponds to the Pietra index. Panel (c) depicts the concentration function $\varphi(x)$ directly, highlighting its convex, nondecreasing nature and its interpretation as a lower bound on set probabilities. The line of perfect equality ($y = x$) corresponds to the case $\Pi = \Pi_0$, while deviations below the line indicate increasing divergence between Π and Π_0.

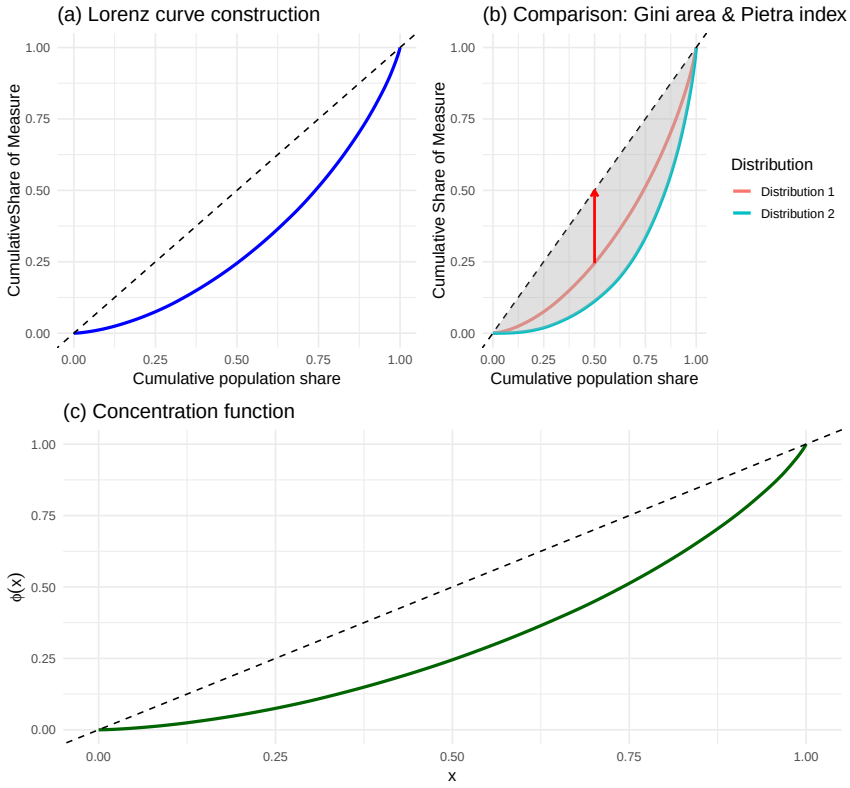

Figure 1.: Lorenz curves and concentration function. Panel (a) shows the construction of the Lorenz curve from cumulative population and measure shares Π. Panel (b) illustrates inequality measures: the shaded area corresponds to the Gini concentration area, while the maximal vertical distance $x - \varphi(x)$ over $x \in [0, 1]$ corresponds to the Pietra index. Panel (c) depicts the concentration function $\varphi(x)$ directly. This figure uses the classic income inequality context for illustrative purposes only.

4. General approach

The proposed tool is based on the concentration function [28], which generalizes the classical Lorenz curve. Unlike the traditional Lorenz curve, which compares a distribution to a uniform reference, the concentration function allows the comparison of any pair of probability measures Π_0 and Π_1 defined on the same measurable space. In our context, Π_0 does not need to be uniform.

Consider the predictive likelihood ratio expressed as

$$\Lambda(y) = \frac{d\Pi_1}{d\Pi_0}(y) = \frac{p_1(y \mid x_{1:t})}{p_0(y \mid x_{1:t})} = \frac{\int p_1(y \mid \theta_1, x_{1:t})\, p_1(\theta_1 \mid x_{1:t})\, \mathrm{d}\theta_1}{\int p_0(y \mid \theta_0, x_{1:t})\, p_0(\theta_0 \mid x_{1:t})\, \mathrm{d}\theta_0}, \tag{1}$$

where $p_0(y \mid x_{1:t})$ and $p_1(y \mid x_{1:t})$ denote the predictive density functions of two competing models, M_0 and M_1, respectively, and $p_0(\theta_0 \mid x_{1:t})$ and $p_1(\theta_1 \mid x_{1:t})$ are the corresponding posterior distributions. The posterior distributions are obtained via Bayes' theorem, which in this context can be written as

$$p_j(\theta_j \mid x_{1:t}) \propto p_j(x_{1:t} \mid \theta_j)\, p_j(\theta_j) = l_j(\theta_j; x_{1:t})\, p_j(\theta_j), \quad j = 0, 1,$$

where $l_j(\theta_j; x_{1:t})$ and $p_j(\theta_j)$ denote the likelihood function and prior distribution for the parameters θ_j of model M_j.

In practice, the predictive likelihood ratio $\Lambda(y)$ can be approximated using posterior Monte Carlo samples defined as

$$\hat{\Lambda}(y) = \frac{\hat{p}_1(y \mid x_{1:t})}{\hat{p}_0(y \mid x_{1:t})} = \frac{\frac{1}{M}\sum_{m=1}^{M} p_1(y \mid \theta_1^{(m)}, x_{1:t})}{\frac{1}{M}\sum_{m=1}^{M} p_0(y \mid \theta_0^{(m)}, x_{1:t})} = \hat{r},$$

where $(\theta_j^{(1)}, \ldots, \theta_j^{(M)})$ are posterior draws from model M_j. Given a forecast observation y, this ratio $\hat{r}$ quantifies which model assigns a higher predictive probability to the realized outcome.

5. Autoregressive moving average models

ARMA models are central to time series econometrics, providing a flexible framework for capturing serial dependence in stationary stochastic processes. Formally, an ARMA(p, q) model is defined as

$$y_t = \mu + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \theta_j u_{t-j} + u_t,$$

where $\{u_t\}$ is a white noise process with zero mean and constant variance σ^2. The AR component of order p captures linear dependence on past observations, while the MA component of order q accounts for dependence on past innovations. Stationarity and invertibility conditions, expressed in terms of the roots of the AR and MA polynomials, ensure that the process is well-defined and that parameters are uniquely identified. Related developments for periodic conditional heteroscedastic models can be found in [30].

Classical estimation of ARMA models is typically performed by maximum likelihood, although least squares and conditional sum of squares estimators are also used in practice. The Gaussian likelihood is particularly convenient, yielding efficient estimators under normally distributed innovations and consistent, asymptotically normal estimators under more general conditions. Model order selection is guided by information criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which balance in-sample fit with parsimony.

In the Bayesian framework, prior distributions are specified for the parameter vectors $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$, and the variance σ^2. Common choices include Gaussian priors for the AR and MA coefficients, constrained to satisfy stationarity and invertibility, with inverse-gamma or Jeffreys priors for σ^2. Given the observed data $Y = (y_1, \ldots, y_T)^\top$, inference proceeds by evaluating the posterior distribution defined as

$$p(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2 \mid Y) \propto L(Y \mid \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)\, p(\boldsymbol{\phi})\, p(\boldsymbol{\theta})\, p(\sigma^2),$$

where L denotes the likelihood implied by the ARMA model. Because this posterior distribution is generally intractable, simulation-based methods such as Markov Chain Monte Carlo are employed. Gibbs sampling and Metropolis–Hastings algorithms are standard approaches, with tailored proposal mechanisms used to enforce stationarity and invertibility.

The Bayesian formulation offers several advantages over classical maximum likelihood estimation. First, it delivers full posterior distributions of parameters, allowing for direct probabilistic statements about uncertainty rather than relying solely on asymptotic approximations. Second, Bayesian model comparison can be performed via marginal likelihoods or Bayes factors, offering an alternative to information-criteria-based selection. Lastly, posterior predictive distributions naturally extend the framework to forecasting, producing not only point forecasts but also predictive densities that incorporate parameter uncertainty.

6. Simulation study

Suppose we observe a time series $\{x_1, \ldots, x_n\}$ and consider two competing models, M_0 and M_1, fitted to these data. Given a forecast observation y, we compute the predictive ratio r to determine which model provides the better forecast. Specifically, if $r < 1$, model M_0 is favored, whereas if $r > 1$, model M_1 is favored.

To construct concentration functions, we simulate N replications of time series $\{x_1, x_2, \ldots, x_n\}_i$, for $i = 1, \ldots, N$, from a neutral model M^*. For each replication, the first two-thirds of the sample are used for training, while the remaining third is reserved for forecast evaluation. Using a rolling-origin forecast procedure, we fit M_0 and M_1 sequentially at each origin t and compute the predictive ratio r_t at $y = x_{t+1}$. This results in a matrix of predictive ratios of dimension $(n - \lfloor 2n/3 \rfloor) \times N$, where $T = \lfloor 2n/3 \rfloor$, and entry (i, j) represents the ratio $r_{i,j}$ for the forecast $y = x_{T+i}$ in replication j.

Following [31], we express the ratios as $r_k = p_k/q_k$, $k = 1, \ldots, n - T$, where $p = (p_1, \ldots, p_{n-T})$ and $q = (q_1, \ldots, q_{n-T})$ denote the predictive densities under M_1 and M_0, respectively, for the outcomes $y_k = x_{T+k}$. Normalizing to form discrete probability measures over the forecast points, we define

$$\tilde{p}_k = \frac{p_k}{\sum_{i=1}^{n-T} p_i}, \quad \tilde{q}_k = \frac{q_k}{\sum_{i=1}^{n-T} q_i}.$$

For each replication j, we treat $\tilde{q}$ as the reference distribution and measure the divergence of $\tilde{p}$ from $\tilde{q}$.

To construct the concentration function, we rank the outcomes by increasing values of r_k, thereby ordering them from those where M_1 assigns substantially less probability mass than M_0 to those where it assigns more. Denoting the reordered outcomes as $y_{(1)}, \ldots, y_{(n-T)}$, with corresponding reordered probabilities $\tilde{p}_{(1)}, \ldots, \tilde{p}_{(n-T)}$ and $\tilde{q}_{(1)}, \ldots, \tilde{q}_{(n-T)}$, we define cumulative sums as

$$P_k = \sum_{i=1}^{k} \tilde{p}_{(i)}, \qquad Q_k = \sum_{i=1}^{k} \tilde{q}_{(i)}, \qquad P_0 = Q_0 = 0.$$

The concentration curve of Π_1 relative to Π_0 is the polygonal line joining given by

$$(0, 0) \to (Q_1, P_1) \to \cdots \to (1, 1),$$

with consecutive points connected by line segments. These constructions yield convex, nondecreasing functions from $(0, 0)$ to $(1, 1)$. This framework enables direct comparison of M_0 and M_1 in terms of predictive performance, highlighting which model achieves a closer fit to the observed forecasts.

Classical inequality measures extend naturally to this predictive setting. The Pietra index and the Gini concentration coefficient become

$$C_{\Pi_0}(\Pi_1) = \max_k (Q_k - P_k), \quad G_{\Pi_0}(\Pi_1) = 1 - 2 \sum_{k=1}^{n-T} \frac{P_{k-1} + P_k}{2} (Q_k - Q_{k-1}), \tag{2}$$

respectively. Smaller values of these indices indicate closer agreement between the predictive distributions, whereas larger values reflect greater divergence.

Next, we compare the AR(1) model versus the MA(1) model. We consider $N = 100$ replications of $n = 100$ observations $\{x_t\}$ generated from the data-generating process

$$M^*\text{: ARMA}(1, 1), \qquad x_t = 0.2x_{t-1} + 0.5u_{t-1} + u_t.$$

For each replication, the sample of length 100 is divided into two parts: the first two-thirds ($T = \lfloor 2n/3 \rfloor = 66$) are used as the training set, while the last third is reserved for forecast evaluation using a rolling-origin procedure. At each forecast origin t, two competing models are estimated: M_0, an AR(1) model, and M_1, an MA(1) model. Forecasts are then generated for $y = x_{t+1}$, iterating until the full evaluation window is exhausted.

Posterior inference is carried out separately for both models using Hamiltonian Monte Carlo with 15,000 iterations, of which 7,500 are discarded as warm-up. Let M denote the number of posterior draws, producing two sets of posterior samples,

$$\theta_0^{(1)}, \ldots, \theta_0^{(M)} \quad \text{for } M_0, \qquad \theta_1^{(1)}, \ldots, \theta_1^{(M)} \quad \text{for } M_1.$$

For each forecast point $y = x_{t+1}$ with $t = T, \ldots, n-1$, the predictive densities $p_0(y \mid x_{1:t})$ and $p_1(y \mid x_{1:t})$ are approximated via posterior predictive averaging as in (1). Repeating this process across all replications $j = 1, \ldots, N$ yields, for each replication, a sequence of predictive ratios $r_{1,j}, \ldots, r_{n-T,j}$. For each j, these ratios are then used to construct a concentration function, resulting in a collection of concentration curves across the simulation study.

Figure 2 displays the mean concentration functions for both models. Panel (a) compares the concentration function of each model with respect to the uniform distribution. The concentration curve of M_0 lies closer to the diagonal line $y = x$ than that of M_1, indicating that the AR(1) model exhibits less inequality in predictive allocation. In predictive terms, this implies that M_0 provides systematically better forecasts for $y = x_{t+1}$ across replications. Since the data were generated from an ARMA(1,1) process, these results suggest that AR(1) captures the dominant predictive structure more effectively than MA(1). The figure also includes 95% credible intervals, computed from the 2.5% and 97.5% quantiles across simulations. In the vast majority of cases, the concentration function of M_0 dominates that of M_1.

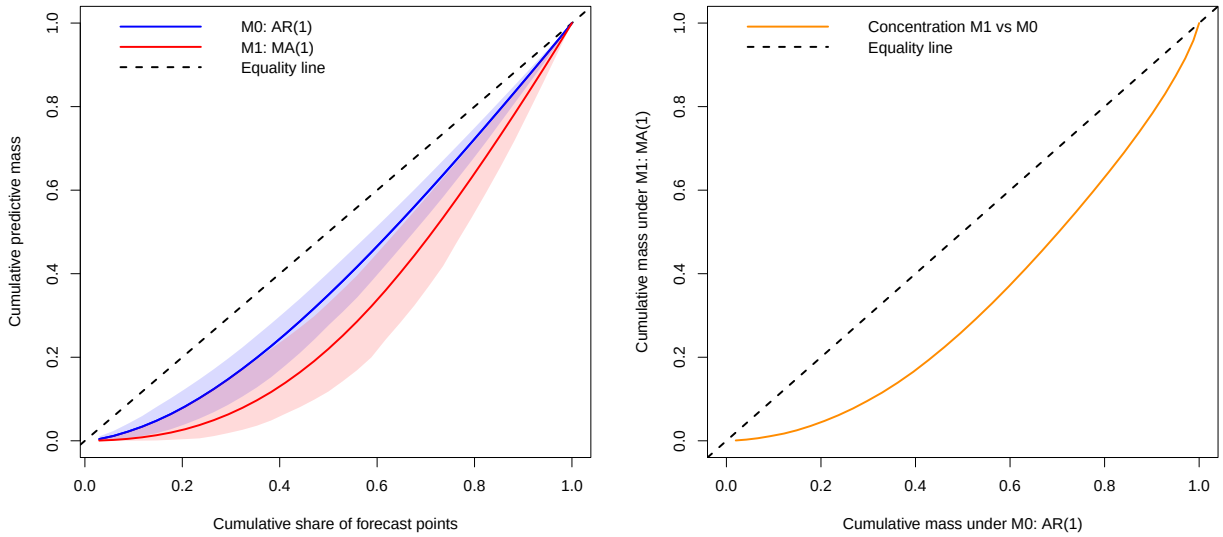

Figure 2.: Comparison of concentration functions between AR(1) and MA(1). Panel (a): concentration function of each model with respect to the uniform distribution. Panel (b): concentration function of MA(1) relative to AR(1).

Panel (b) of Figure 2 shows the concentration function of M_1 relative to M_0. The curve is constructed by joining the points $(0,0),(Q_1,P_1),\ldots,(1,1)$, with the line $y=x$ shown for reference. When the two models exhibit similar predictive behavior, their concentration functions lie close to the diagonal. Conversely, the farther the curve lies below the diagonal, the stronger the evidence in favor of M_0 relative to M_1.

For further comparison, inequality measures were computed for each replication. Specifically, the Pietra index and the Gini concentration coefficient were calculated from the concentration functions, yielding a distribution of indices for each model. Figure 3 shows the histograms of both indices. On average, the indices are systematically lower for the AR(1) model than for the MA(1) model, consistent with the graphical evidence from the concentration functions.

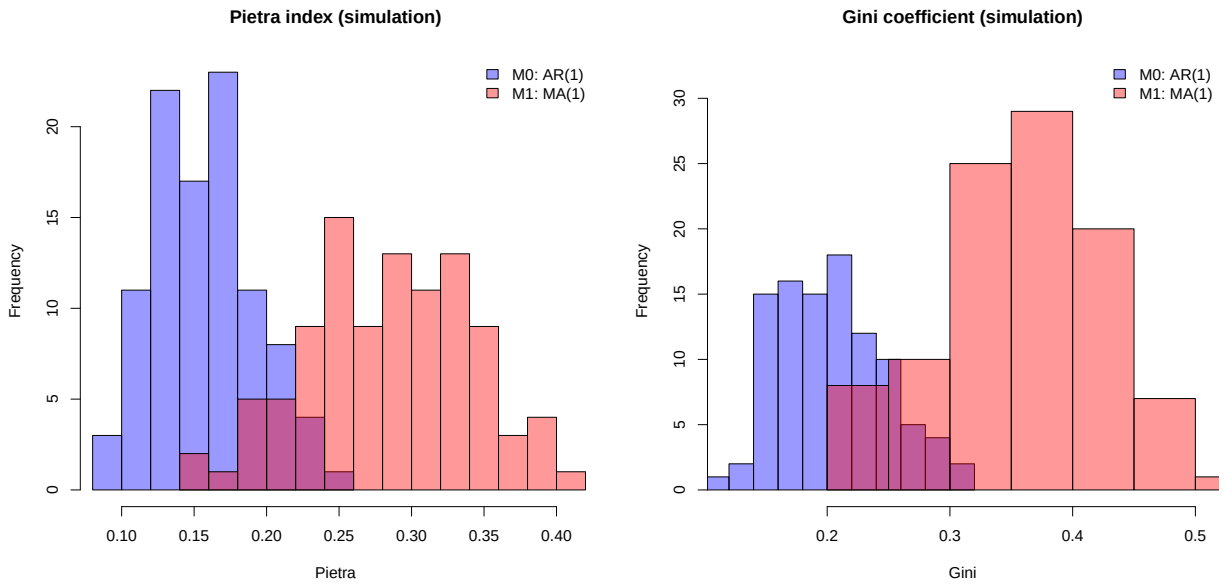

Figure 3.: Histograms of the Pietra index and the Gini concentration coefficient for AR(1) and MA(1) models.

To assess consistency with standard predictive criteria, we also computed the negative log predictive density (NLPD). Table 1 summarizes the average values across replications. For the Pietra index, the Gini concentration coefficient, and the NLPD, smaller values indicate a better fit. The results confirm consistency across all criteria: the AR(1) model is systematically favored.

Table 1.: Comparison of indices for M_0: AR(1) and M_1: MA(1)

Model	Pietra index	Gini concentration coefficient	NLPD
M_0: AR(1)	0.157	0.204	0.540
M_1: MA(1)	0.284	0.359	1.189

Table 2 reports the agreement rates between the concentration-function-based decision and the standard criteria. Note that the Pietra index, the Gini concentration coefficient, and the NLPD all agree 100 percent of the time in favor of the AR(1) model. This provides strong evidence that the concentration function is consistent with established predictive criteria, while also offering an intuitive, distributional perspective on predictive performance.

Table 2.: Agreement rates between the concentration-function-based decision and standard criteria

	Pietra index	Gini concentration coefficient	NLPD
M_0 better than M_1	100%	100%	100%

7. Real data application: U.S. natural gas prices

We illustrate the use of the concentration function by analyzing a publicly available dataset of average U.S. natural gas industrial prices (measured in dollars per thousand cubic feet) obtained from the Energy Information Administration. The sample covers the period 2010M01–2025M06, for a total of 186 monthly observations. The data are available from the U.S. Energy Information Administration website.

To build the concentration function, we proceed as follows. First, we fit the complete dataset using the function `auto.sarima` from the R package `bayesforecast`, which automatically selects the best SARIMA model in a Bayesian framework. Specifically, `auto.sarima` chooses the optimal specification according to the Bayesian Information Criterion (BIC), which is the default model selection criterion in the package, and then estimates the model via Hamiltonian Monte Carlo using default SARIMA priors. For the autoregressive and moving-average coefficients, we use truncated normal priors with baseline distribution $\mathrm{N}(0, 0.5)$, restricted to the parameter region that satisfies the stationarity and invertibility conditions. This automatic procedure provides an initial benchmark model for Bayesian analysis.

Next, we split the sample into an initial two-thirds training set ($\lfloor 2n/3 \rfloor = 124$ observations) and a final one-third evaluation set. Using a rolling-origin scheme, we estimate and compare two models at each step:

- M_0 – The model proposed by `auto.sarima` (an AR(1));
- M_1 – An alternative ARMA specification.

At each iteration t, we expand the training sample by adding one new observation x_t and re-estimate both models. Forecast evaluation is based on the next-period realization $y = x_{t+1}$. For each forecast, we compute the predictive likelihood ratio and update the concentration function. Repeating this procedure until the full sample is used yields a complete concentration curve for each model comparison.

For the complete dataset, `auto.sarima` selects an AR(1) model as the best specification. Figure 4 displays the concentration functions comparing M_0: AR(1) against various alternatives.

When comparing AR(1) with AR(2) or AR(3), the concentration functions are nearly indistinguishable from one another, with curves that lie very close to the 45-degree diagonal line. This indicates that their predictive performances are almost identical. This result is expected, since the reference model selected was an AR(1); the higher-order autoregressions essentially overfit but still capture the same dynamics. In contrast, differences become more pronounced when AR(1) is compared with ARMA(1,1), ARMA(1,2), ARMA(2,1), or MA(1). In these cases, the concentration functions diverge more clearly, indicating inferior predictive performance relative to AR(1).

Table 3 summarizes additional fit metrics for the competing models, including the Pietra index, the Gini concentration coefficient, the negative log predictive density (NLPD), the continuous ranked probability score (CRPS), and the marginal log-likelihood. For the Pietra index, the Gini concentration coefficient, the NLPD, and the CRPS, smaller values indicate a better fit, whereas for the marginal log-likelihood, larger values indicate a better fit.

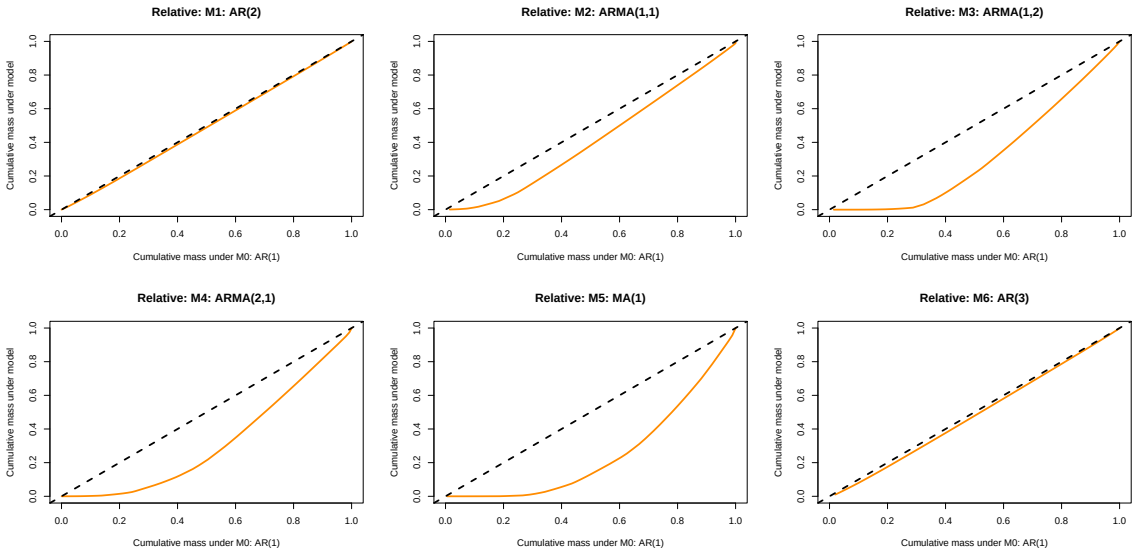

Figure 4.: Concentration functions comparing M_0: AR(1) with alternative ARMA models (M_1): (a) AR(2), (b) ARMA(1,1), (c) ARMA(1,2), (d) ARMA(2,1), (e) MA(1), and (f) AR(3).

The results confirm that AR(1) provides the best overall fit, with AR(2) and AR(3) yielding very similar measures, while the alternative ARMA and MA specifications perform substantially worse. It is worth noting that, according to the marginal log-likelihood criterion, the ARMA(1,2) and ARMA(2,1) model outperforms the AR(1) specification. However, this superiority is limited to that particular criterion, as the AR(1) model exhibits a better fit according to the remaining evaluation measures. A plausible explanation is that, since the marginal log-likelihood is estimated rather than computed exactly, the use of the simplified version shifted gamma estimator proposed in [32] may introduce instability. Alternative estimation approaches, such as the bridge sampling method proposed in [33] or thermodynamic integration [34], could also be used to estimate the marginal likelihood.

Table 3.: Pietra index, Gini concentration coefficient, NLPD, CRPS, and marginal log-likelihood for ARMA models

Model	Pietra index	Gini concentration coefficient	NLPD	CRPS	Marginal log-likelihood
M_0: AR(1)	0.188	0.252	1.785	0.569	-190.27
M_1: AR(2)	0.192	0.256	1.814	0.582	-191.67
M_2: ARMA(1,1)	0.264	0.345	2.141	0.712	-189.73
M_3: ARMA(1,2)	0.382	0.479	3.920	1.210	-184.49
M_4: ARMA(2,1)	0.397	0.491	3.057	1.111	-185.38
M_5: MA(1)	0.468	0.583	5.214	1.672	-201.94
M_6: AR(3)	0.197	0.261	1.828	0.589	-193.42

The Pietra index and the Gini concentration coefficient reported in Table 3 are computed with respect to the reference line $y = x$. Given that the model M_0: AR(1) demonstrates superior predictive performance relative to the other specifications, the measures defined in (2) can be used to compute the Pietra index and the Gini concentration coefficient of the competing models $M_1, \ldots, M_6$ relative to M_0. Table 4 presents these coefficients. The results are consistent with those reported in Table 3, reinforcing the comparative conclusions previously discussed.

Table 4.: Pietra index and Gini concentration coefficient of models $M_1, \ldots, M_6$ relative to model M_0: AR(1)

Model	Pietra index	Gini concentration coefficient
M_1: AR(2)	0.0135	0.0191
M_2: ARMA(1,1)	0.1460	0.1824
M_3: ARMA(1,2)	0.2873	0.3504
M_4: ARMA(2,1)	0.2897	0.3594
M_5: MA(1)	0.3728	0.4921
M_6: AR(3)	0.0248	0.0356

The concentration function proves to be a powerful diagnostic tool for comparing the predictive performance of competing time series models. It offers an intuitive, visual method that complements more traditional information criteria and Bayesian measures such as the NLPD, the CRPS, and marginal likelihoods. Importantly, it can be computed directly from posterior samples, making it especially suitable in Bayesian settings.

8. Conclusions

The concentration function can be applied to compare predictive distributions across competing models. A model that provides superior predictive performance yields a concentration function that lies closer to the line of equality ($y = x$), while a model with poorer predictive fit generates a concentration function that lies farther below this line. This visual and quantitative comparison allows for an intuitive assessment of predictive adequacy.

In a Bayesian context, the concentration function serves as a useful measure of model fit. It can be complemented with inequality-based indices such as the Pietra index and the Gini concentration coefficient, which quantify the degree of deviation from the line of equality. These measures can be readily computed from predictive distributions by leveraging posterior samples of the model parameters obtained through Markov Chain Monte Carlo methods.

The concentration function and its associated indices provide an alternative framework for model evaluation that complements traditional Bayesian criteria, such as the NLPD, the CRPS, marginal likelihoods, and Bayes factors. This approach highlights the distributional properties of predictive performance and offers a flexible, simulation-based tool for model comparison.

Future research should apply the concentration function to broader model classes, including stochastic volatility and GARCH-type models, heavy-tailed innovations, nonlinear autoregressive or threshold models, structural-break and regime-switching models. This would test the robustness and generality of the measure beyond ARMA dynamics.

Statements

Acknowledgements

The authors would like to thank the Editors-in-Chief and the two anonymous reviewers for their valuable comments and suggestions, which substantially improved the quality of this article.

Author contributions

Conceptualization, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, writing original draft, writing revision and editing: C. Cruz-Torres, F. Ruggeri.

Conflict of interest

The authors declare that they have no conflicts of interest.

Data and code availability

The data and complete source code is publicly available on GitHub at https://github.com/cact560/CF_ARMA

Declaration on the use of artificial intelligence (AI) technologies

The author used a generative AI tool only for language polishing.

Funding

The work started during the visit of F. Ruggeri to Universidad Nacional Autónoma de Honduras to teach a course supported by the Volunteer Lecturer Programme of the Commission for Developing Countries of the International Mathematical Union, whose contribution is kindly acknowledged.

References

[1] Lorenz, M.O., 1905. Methods of measuring the concentration of wealth. Publications of the American Statistical Association, 9, 209–219. DOI: 10.1080/15225437.1905.10503443.

[2] Gini, C., 1914. Sulla misura della concentrazione della variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Anno Accademico 1913-1914, Tomo LXXIII, Parte II, 1203–1248. URL: https://www.hetwebsite.net/het/texts/gini/gini_1914.pdf.

[3] Wagstaff, A., Paci, P., and van Doorslaer, E., 1991. On the measurement of inequalities in health. Social Science and Medicine, 33, 545–557. DOI: 10.1016/0277-9536(91)90212-U.

[4] Boyce, J.K., Zwickl, K., and Ash, M., 2016. Measuring environmental inequality. Ecological Economics, 124, 114–123. DOI: 10.1016/j.ecolecon.2016.01.014.

[5] Tessum, C., Apte, J., Goodkind, A., Muller, N., Mullins, K., Paolella, D., Polasky, S., Springer, N., Thakrar, S., Marshall, J., and Hill, J., 2019. Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. Proceedings of the National Academy of Sciences of the United States of America, 116, 6001–6006. DOI: 10.1073/pnas.1818859116.

[6] Engelmann, B., Hayden, E., and Tasche, D., 2003. Measuring the discriminatory power of rating systems. Deutsche Bundesbank Discussion Paper 01/2003. DOI: 10.2139/ssrn.2793951.

[7] Jost, L., 2006. Entropy and diversity. Oikos, 113, 363-375. DOI: 10.1111/j.2006.0030-1299.14714.x.

[8] Albarrán, P., Crespo, J.A., Ortuño, I., and Ruiz-Castillo, J., 2011. The skewness of science in 219 sub-fields and a number of aggregates. Scientometrics, 88, 385–397. DOI: 10.1007/s11192-011-0407-9.

[9] Ledoux, M., 2001. The concentration of measure phenomenon. American Mathematical Society, Providence, RI, USA. DOI: 10.1090/surv/089.

[10] Armstrong, J.S. and Collopy, F., 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting, 8, 69–80. DOI: 10.1016/0169-2070(92)90008-W.

[11] Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. Journal of Business and Economic Statistics, 20, 134-144. DOI: 10.1198/073500102753410444.

[12] West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica, 64, 1067–1084. DOI: 10.2307/2171956.

[13] Clark, T.E. and McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics, 105, 85–110. DOI: 10.1016/S0304-4076(01)00071-9.

[14] Clark, T.E. and McCracken, M.W., 2005. Evaluating direct multistep forecasts. Econometric Reviews, 24, 369–404. DOI: 10.1080/07474930500405683.

[15] White, H., 2000. A reality check for data snooping. Econometrica, 68, 1097–1126. DOI: 10.1111/1468-0262.00152.

[16] Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics, 23, 365–380. DOI: 10.1198/073500105000000063.

[17] Hansen, P.R., Lunde, A., and Nason, J.M., 2011. The model confidence set. Econometrica, 79, 453–497. DOI: 10.3982/ECTA5771.

[18] Elliott, G. and Timmermann, A., 2008. Economic forecasting. Journal of Economic Literature, 46, 3–56. DOI: 10.1257/jel.46.1.3.

[19] Diebold, F.X., Gunther, T.A., and Tay, A.S. 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review, 39, 863–883. DOI: 10.2307/2527342.

[20] Gneiting, T. and Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102, 359–378. DOI: 10.1198/016214506000001437.

[21] Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication. Econometric Reviews, 18, 1–73. DOI: 10.1080/07474939908800428.

[22] Geweke, J., 2001. Bayesian econometrics and forecasting. Journal of Econometrics, 100, 11–15. DOI: 10.1016/S0304-4076(00)00046-4.

[23] Gelman, A., Hwang, J., and Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing, 24, 997–1016. DOI: 10.1007/s11222-013-9416-2.

[24] Vehtari, A., Gelman, and A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27, 1413–1432. DOI: 10.1007/s11222-016-9696-4.

[25] Slama, A., 2022. A Bayesian detection of structural changes in autoregressive time series models. Chilean Journal of Statistics, 13, 47-66. DOI: 10.32372/ChJS.13-01-03.

[26] Ruggeri, F., 2025. Data analysis using the concentration function. Wiley Reviews, 17, e70048. DOI: 10.1002/wics.70048.

[27] Pietra, G., 1915. Delle relazioni tra gli indici di variabilità. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, Anno Accademico 1914-1915, Tomo LXXIV, Parte II, 74, 775–792. URL: https://preserver.beic.it/delivery/DeliveryManagerServlet?dps_pid=IE3581190.

[28] Cifarelli, D. and Regazzini, E., 1987. On a general definition of concentration function. Sankhya, 49, 307–319. URL: https://www.jstor.org/stable/25052510.

[29] Regazzini, E., 1992. Concentration comparisons between probability measures. Sankhya, 54, 129–149. URL: https://www.jstor.org/stable/25052733.

[30] Bibi, A., 2024. Statistical inference for first-order periodic autoregressive conditional heteroscedasticity models. Chilean Journal of Statistics, 15, 169–191. DOI: 10.32372/ChJS.15-02-04.

[31] Ekin, T., Ieva, F., Ruggeri, F., and Soyer, R., 2017. On the use of the concentration function in medical fraud assessment. The American Statistician, 71, 236–241. DOI: 10.1080/00031305.2017.1292955.

[32] Raftery, A., Newton, M., Satagopan, J., and Krivitsky, P., 2005. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., and West, M. (eds.) Bayesian Statistics 8: Proceedings of the Eighth Valencia International Meeting, Valencia, Spain. Oxford Academic, Oxford, UK, pp. 381–426. DOI: 10.1093/oso/9780199214655.003.0015.

[33] Chib, S. and Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association, 96, 270–281. DOI: 10.1198/016214501750332848.

[34] Calderhead, B. and Girolami, M., 2009. Estimating Bayes factors via thermodynamic integration and population MCMC. Computational Statistics and Data Analysis, 53, 4028–4045. DOI: 10.1016/j.csda.2009.07.025.