Improving sociometric measurements through modeling and statistical instruments to support the coherence and effectiveness of legislation

ALKIDA LLAKAJ[1], ELMIRA KUSHTA[2], DODË PRENGA[3]
[1]Department of Economy,
Mediterranean University of Tirana, Tirana,
ALBANIA

[2]Department of Physics,
University of Tirana, Tirana,
ALBANIA

[3]Department of Mathematics and Physics,
University of Vlora, Vlora,
ALBANIA

Abstract: - Mathematical modelling, exhaustive techniques adopted for measuring latent behaviour alongside standard statistical analysis are being expanded continuously in studying legislative effectiveness, auditing policies, pre-post- and ex-post evaluations and several other assessments in behavioural and political sciences. The spreads of numerical approaches in socio-political features analysis are aligned with the necessities for quantitative answering of standard problems regarding the effectiveness of legislative packages and normative acts that might change frequently, their expediency to achieve target objectives, in establishing thought standards, in addressing society problems, for ensuring sustainable development, measuring social supportiveness and many other important indicators. In addition to the standard protocols that regulate measurement procedures for those features, the extraction of telling evidence from typically heterogenous and un-relaxed systems requires careful and well-adopted methodologies. This feature is boldened for measurement on systems representing transiting democracies features where frequent legislative revisions and legal challenges of normative acts infuriate the stability of the parameters and impede time dynamics on them. In this context, we have discussed some improvements in the generalised measurement procedures by contemplating physics' instruments features, to diminish the effect of heterogeneity and dynamics. Consequently, non-objective answering in common questionnaires, lack of homogeneity and relaxation's assumption, are balanced and corrected in the framework of the assessment and certification for instruments' reliability, accuracy, resolution power etc. The result consists of enhanced measurement and reliable quantitative assessment for low scale measurement for the systems considered in this methodical analysis. It can be suggested for use in scientific measurement and research but also, it assist entities that operate in the field of measurement by conducting evidence-based analyses, to help boost the validity and generalizability of research findings, and to support governmental institutions in implementing cost-efficient legislative procedures. Also, improvements in classical sociometric techniques can be seen as adequate response of the rapid advancement of artificial intelligence, intensification of the network-based interactions, and to count for complexity of opinion and belief formation within contemporary societies. Within this framework, the present work outlines several thematic considerations related to the implementation of direct and indirect measurement techniques in heterogeneous and non-relaxed environments. Particular attention is given to modelling approaches, the identification of characteristic statistical distributions and network properties, as well as the application of sequential measurement procedures in small communities where conventional sampling strategies are either impractical or prohibitively costly

Key-Words: - Statistical analysis; Legislation; Modeling; Sociometric Measurement

Received: March 5, 2026. Revised: April 4, 2026. Accepted: May 13, 2026. Published: June 29, 2026.

1 Introduction

The research practice in contemporary sociometry, political sciences and modern sociology, are relying extensively on quantitative analyses, improved measurement and modelling techniques based on rigor statistical data processing and analysis. Thought political sciences have stipulated in principles research techniques and methodologies with typologies like econometric or sociometric studies of all natures and levels, see [1], [2], there are room for some extra adjustment of general protocols in practice, especially for the systems with heterogeneity presence, or non-relaxed time dynamics of key variables. Also, the final interpretation and alignment of results with the intended purpose by domain experts require skills and statistical expertise regarding the correct statement, comparison of quantities embodied on several categories as reforms efficacity, legislative improvements, economic advance or enhanced activities. The increased interests and general understanding for high-quality legislative and normative frameworks, productive public policies, systematic auditing, and the continuous assessment of institutional performance, combined with the spread of laws and the high frequency of amendments in post-communist societies as in our country, have encouraged the need for integrated ex-ant and ex post evaluations. Meanwhile, asquai commentaire as common political approaches nowadays necessity a concise review of reports, that must be based on numerical evidence and measurable significance. In this regard, a generalised measuring process including data analysis, data gathering procedure or direct measurement, regression, hypothesis test and modelling are not easy, nor straightforward standard procedure. In fact, even direct measurement of characteristic quantities would result too noisy for simplistic consideration, whereas modelling with them or measuring indirectly some latent variables, or extending the analysis toward numerical calculation for those fields results in crispy job not only for young scholars and senior researchers. In support of resolving problems related to the numerical approaches, modern political science and related disciplines have established research and analytical methodologies which mimic, compete and parallel the econometric and sociometric studies across various domains and levels, see [3], [2], [4], [5] etc. Measurement in these fields brings additional layer of complexity because it involves behavioural variables, attitudes, perceptions, and similar structures which are by nature not strictly quantizable. Dealing with their analysis requires caution and a dynamic perspective both during the implementation of protocols and procedures and in the accurate interpretation of findings, processes for which interdisciplinary approaches are frequently indispensable. The degree of analytical rigor and accuracy required ultimately depends on the nature of the analysis and the system under examination. For instance, assessing community support for a particular issue may be conducted through direct measurement, whereas the preliminary evaluation of a legal act or a specific public policy can be significantly more challenging, and so on. To highlight the motive and importance of this work we can illustrate the idea from a less scientific perspective, but a more important for society, such as historical facts. It is well known that misunderstanding is related to the failure of correctly measuring a typical parameter in the system. For example, politicians can insist that their idea is the best or at least the better one, without any proof! A fatal failure of state agencies to assess the impossibility of some financial agencies during 1996-1997 based on Ponzy scheme for guaranteeing the promised returns has brought Albania on the edge of catastrophe the year. We believe also that the failure to provide correct measurements and rigor balance on social alternatives is one of the most basic causes for many crises in modern times. In short, this work is based on an incentive to elaborate some practical calculation elements that are helpful for analysing non-standard system of our country's' reality, which also can be useful for similar systems or circumstances.

2 Sociometric Measurement

Quantitative methods in sociometric and similar analysis and their required accurateness follow the usability of their outputs outlined by the analyzing body and the purposes of the investigation. In this regard research work and monitoring measurement might differ considerably. So, a political survey might be happy even for biased measurement to promote electorate partition in election for example, some government agencies can accept non rigor values from the informality sector to create a general idea, and this can be easily evidenced in discrepancies or recognized error effect in the official databases, as seen in several categories in our statistics (https://www.bankofalbania) and for sure, in similar one around the region or more. It is well known that official measurements of many types are managed by rigor protocols, as for example, the European Statistics Code of Practice (https://ec.europa.eu/eurostat/). Other monitoring aspects are driven by purpose-oriented research guidelines in addition to statistical protocols. Nonetheless, when disusing very specific measurement suchlike the legislative effectiveness or ex post evaluation for example, more caution and mathematical rigor, or even dedicated methodologies are needed. For example, there exists a set of good practices and regulatory acts issued by institutions of the European Community, [4] that can be used from auditing agencies, particularly for posterior evaluations. When analyses are conceptually or theoretically more complicated, interdisciplinary approaches are often preferable, as discussed in [2] and references therein. For more complicated cases such as the ex-ante analyses, wherein a legal norm or policy is examined as a projection of a desired future outcome that can be temporally transposed to anticipate the intended effect, quantitative techniques become broader and more complex and cannot be described by a general protocol. In this framework, even non-research activities are very interested in accurate measurement in social sciences. Evaluations and assessments performed by stakeholders or interest groups also require typical advanced data processing in order to persuade

decision-making, that is regarded herein as an indirect and high complexity measurement. Some agencies are interested constantly in getting accurate data or reports aimed at considering options or to prevent the introduction of poorly regulated policy products. In response to growing demands for methodological efficiency, sociometry has evolved toward interdisciplinary approaches and complex systems analysis. This transformation has been affected by exchanges with network science initiated by [6], [7], and socio-physics disciplines pioneered by [8]. Nowadays, sociologists and sociometrist scientists are progressively using experiments as research tools for hypothesis testing [10], for analyzing specialized structures. Recently, computational political science has gained notoriety [11]. From a basic perspective in this consideration, several measurements have been carried out for our system. In [18], [20] different nonstationary parameters for distributions of the votes for the left and right political wings, and for small and major parties have been reported. In [19] interesting parameters of the COVID – 19 data have been analyzed by using crossed approaches, etc. Both references reveal the importance of system-matching methodologies for accurate results. Within this context, improving measurement is not a research objective or interest, but an important procedural step for warranting the reliability of the policy or regulatory product that is based on the measurement outcomes and predictions. Considering the social impact of legislation efficiency, we recognize that incremental enhancements in measurement accuracy can vintage substantial and tangible results. They can contribute to reducing political distortions in determining legal norms, relieve shifted interpretation from the implementing bodies, and foster a convincing culture of legality. In this context, this paper aims also to evoke measurement quality and methodological rigor to promote the primacy of facts over assumptions in social and political media. From technical point of view this paper focuses on descriptive elements related to generalized instruments and their calibration, methodological consistency in sociometric evaluations.

2.1 Calibration of Sociometric Instruments.

Classical instruments for direct measurement in social systems are interviews, group discussion, tests, surveys and questionnaires, while for indirect measurements models are used. Considering those tools as homologues of physical instruments or physical devices, an easier understanding for their required features can be achieved by direct analogy. In this regard, following the established methodological standards for specific measurement, the researcher must convey attention to ensure that the fulfillment of the conditions under which those standards are applicable. Also, like physical measuring devices, the measuring instrument in social systems must be calibrated to guarantee trustworthy measurement, the needed sensitivity, measurement validity and reliability. These properties may be paired with specialized guiding principles, such as those drawn in [12]. It is important in this regard to outline the degree to which a concept can be associated with a numerical variable. Noticing that specifically the systems of perceptions are dynamic over time and exhibit spatial heterogeneity, the instruments mentioned herein need to be matched and adapted with appropriate statistical procedures. For example, when using the Concept Inventory test proposed by [32] and elaborated next by [33], [34] etc., as measuring instruments for knowledge, we perform a Rasch calibration step, see [22], [21], etc., which consists of an iterative statistical procedure that fits consequently the results of the measurement with the statistics proposed (logistic probability). This differs from physical device collaboration that is performed separately from the scrutiny measurement. Also, other challenges originate from the specifics of socio-political measurements, such as local social inference and individuality of entities being examined, and exterior effects as the compliance with ethical principles such as voluntariness, anonymity, and confidentiality that affect the quality of the measurement regardless of the instrument's proficiency. For example, voluntary participation in surveys impedes the randomness of the sampling, the presence of the groups could impose non IID and therefore compromise the sample size determination etc. Based on our observations from surveys concerning legal issues, political preferences, and similar topics, it appears that these issues are often not clearly addressed. In the following, we will discuss a powerful technique that facilitates the calibration of a measuring instrument in symmetry and psychometry. Before exploring capacities of the method, we are evoking some measured failures of direct and indirect measurement in a very simple procedure, the analysis of knowledge for our high school students. A relevant example derives from the administration of the FCI (Force Concept Inventory) test consisting of 30 inquiries and demanding roughly 45 minutes to complete. Additional to some well-known responses shortcomings, such as respondent fatigue, poor reaching, data biased and others described in [26],[27], we observed that for our systems, some elements are more persistent and must be addressed with priority in various measurements for our systems. So, in the FCI measurement, about 22% of participants selected answers probably without thinking, imposing confusion in classifying them as bad conduct, guessing behavior, or irresponsible responses for further analysis. Also, many respondents have completed the test in only 10–15 minutes (which may help explain the anomalously low performance observed in PISA results). Also, additional 16–18% appeared to engage in guessing when completing the test, but this is identifiable and do not complicate the calibration procedure. Attempts to rely solely on voluntary participation also resulted in an anomalously low number of respondents. Another observation originates from a preparation stage a workshop on education-related topics. In the survey two formats were used: Google Forms, which does not conceal the respondent's email address, and filling forms in www.survenut.org which guarantees anonymity. Among others, the teachers were asked 'how much time do they needed to prepare a new lesson for the following day' that is basically a common question without any consequence or other impact on their pedagogical or professional issues. Nevertheless, differences were apparent: the average time reported was 40 minutes for the anonymous survey, 20 minutes for non-anonymous version. It was also observed that more than 27% of participants contacted did not trust that the survey was anonymous and therefore did not respond at all. Interestingly, the survey's underlying aimed to propose that time spent on lesson preparation should be formally recognized as working time, which was conveyed to them, but as observed, the results were remarkably inconsistent. Such examples highlight the importance of improved measurement techniques in compliance with systems specifics, that are not included in standard protocols. In those cases, and similar circumstances, ensuring the fulfillment of statistical assumptions upon which estimation formulas are based, are of high relevance in order to obtain trustworthy results. Careful attention on guaranteeing the randomness in probabilistic sampling and use advanced calibration techniques such as Rasch model could be harmonized with seeking for latent traits instead of direct measurement which could be more sensitive to ethical or other influences.

2.2 Measuring Latent variables. The Rasch model

Measuring latent variables or traits by a modeling approach rather than conducting direct measurement is proven fruitful in studying knowledge or academic proficiency for several of our systems. Accordingly, in [35] and in two works under publishing, we have proposed extending the very consolidated Concept Inventory analysis in studying legislation efficiency, auditing performance and other investigation of this category, that routinely have been measured directly. Considering that the error nature is more complicated for indirect measurement, we proposed improving the calibration techniques for the instruments that consist of the whole survey procedures, that is, drafting the questionary, selecting the appropriate answering scale (e.g., binary evaluation or multiple scale of Likert type, see [37] or [15] for this model), evidencing for symptoms, etc. In this case, Rasch analysis has proven to overcome several problems of such instruments flaws, but after an exploration observation, we obtained that this method has not been implemented yet for our systems of interests. However, several research and other analyses focused on high school education, [38], etc., constitute a based experience so far. Let us assume that we aim to measure numerically a given performance or effectiveness by constructing a questionnaire composed of items with binary (yes/no) responses, which may be approximated as "correct/incorrect." In such specific cases, the measurement is typically carried out within expert communities, for example, through their judgments regarding a legal act based on its specific amendments and other elements of its structure. The quantity to be assessed behaves as a latent variable, unobserved directly, and determines the response pattern to each item. Accordingly, the responses may be denoted as $V\ (i,j) = 1{,}0$. We can assume those responses as probabilities of binary outcomes based on a given utility function. So far, the respondent who expresses support for a given issue by the probability p, can be assumed to employ a certain utility function regarding the issues

$$U = \log\left(\frac{p}{1-p}\right);\ p(x=1) = \frac{e^U}{1+e^U} \qquad (1)$$

and U is a function of several hidden variables. Therefore, the direct measurement (V) can be seen related to the indirect measurement of U. It followed that binary outcomes gathered by direct measurement could be related to continuous variables that express more than raw probabilistic behavior: it involves judgment, factors' trade off, etc., hence, a deterministic and numerical approach to the behavioral activities. In this approach, the sum of all values $v\ (i,j)$ across the questionnaire divided by the total number of responses $N_i x N_p$ where N_i is the number of respondents and N_p the number of items, is also estimator of the latent quantity: $V = \frac{1}{N_p x N_i}\sum_{i=1 j=1}^{N_p;N_i} v\ (i,j)$. For example, if the values $v\ (i,j)$ represent the experts' support or assessment of specific amendments of a law, then the quantity V serves as an overall estimator of the law itself, regardless the fact that in their raw form, those quantities represent the probability of the support. Before going to some useful details, let discuss the consequence of this equivalence: what degree of trustworthiness is that quantity V expresses the hidden utility U? It relies directly on physical instrument's analogies requirements, that must be verified in a similar way that physical devices are certified for measuring purpose. In this case, we cannot anticipate the needed guarantee, therefore, only a specialized procedure suchlike Rasch model can help, and in this case, we would realize a post factum certification only. However, this is not sufficient for overcoming problems of high error ana uncertainty in direct measurement. So, while answering a questionnaire with Np items may be straightforward for a qualified respondent, but some individuals might lack sufficient clarity regarding the issue and inevitably could inject "noise" into the measurement process, by circling alternatives mostly at the midpoint or selecting Yes/No by random! For addressing this problem, two conditions must be met:

(a) the questionnaire itself must be certified and standardized, or certified by rigor procedures and piloting

(b) the testing procedure must be calibrated and demised to filter out random or non-informative responses.

Issue (a) can be presumed by employing indexes starting from CVR index after experts' reviews, see [28], [29] etc., for details, and issues (b) can be resolved by adopting the Rasch model, and additional methodological adaptations provided in [25], [14], [15]. In this context, for certain types of questionnaires where the ability–difficulty compromise is meaningful or can be appropriately adapted, the Rasch model can be applied effectively, by slight adaption of original model quantities such as difficult and ability related to the Concept Inventory analysis. For practical purposes, when using this approximation, precautions must be taken to ensure that the items relate to a coherent, integral attitude that behaves as a latent factor, and that the questions meaningfully complement one another. The items should also require domain-specific knowledge meaning that the method is most suitably applied within communities of relevant specialists, such as experts evaluating legislative amendments, legal norms, financial regulatory acts, or other areas requiring specialized expertise. Through the proposed technique, it becomes possible to:

a. Calibrate the questionnaire as a measurement instrument.

b. Identify events or response patterns that degrade measurement quality, such as guessing.

c. Detect individuals who lack competence in the topic addressed by the questionnaire, as well as poorly constructed items.

d. Assess the satisfaction of the model's statistical assumptions through the evaluation of deviations.

For further understanding of the Rasch technique we address the reader to the relevant and dedicated literature for Rash model such as [39], [20], [21]., [22], etc. For clearances of reading, we are providing here a few adaptations we proposed in this work. Starting from the original model of evaluating the probability of correct answer if the respondent has the knowledge α_i whereas the item to be answered represents a certain understanding difficulty β_j,

$$P(V(i,j) = 1|\alpha_i, \beta_j) = \frac{e^{\alpha_i - \beta_j}}{1 + e^{\alpha_i - \beta_j}} (2)$$

we propose to acknowledge its variables by the following respectively $\alpha_i = ln\frac{a_i}{1-a_i}$; $\beta_j = ln\frac{1-b_j}{b_j}$

where intermediate quantities a_i and b_j are calculated directly form the measurements via $a_i = \frac{1}{n}\sum_{j=1}^{n} V(i,j)$, $b_j = \frac{1}{N}\sum_{i=1}^{N} V(i,j)$. Additionally, if a multilevel scale evaluation test like Likert variables is used for assessing a certain attitude, the probability for attributing a value $h = (1,2, \dots m)$ is given by the Polytomous Rasch model [22]

$$P(V(i,j) = 1|\alpha_i, \beta_j, h\,) = \frac{e^{\alpha_i - \beta_j - \Sigma_{k=1}^{h} m_k}}{1 + e^{\alpha_i - \beta_j \Sigma_{k=1}^{h} m_k}} (3)$$

This last, involves a little more complexity for calculating and for matching features with the original Concept Inventory test where this model is used purposely, but, at the same time, it provides substantial enhanced analytical capacity, and greater discriminative power. Notice that after those adaptations, computing is a straightforward procedure. So, starting from the first raw data $V(i,j) \equiv P^0(i,j)$ that are considered also as initial probabilities, one replaces iteratively ${P^{old}}_{estimate}(i,j)\ with\ {P^{new}}_{estimate}(i,j)$ where P^{new} is estimating by using current step calculation for α_i and β_j. The iteration stops when a residual estimator $R(\ P_{estimate}(i,j) - V(i,j))$ falls below a specified threshold. Besides original references [21],[22] we suggest to the reader to visit (www.Rasch.org) and (www.real-statistics.com) for step-by-step algorithms and detailed procedure. Also dedicated software's such as WINSTEP can be very fruitful. Finally, ${P^{final}}_{estimate}$, and quantities $\alpha_i^{final} end\ \beta_j^{final}$ constitute improved estimators relative to the initial ones in (1). In this regard, the test or questionnaire can be filtered by removing elements that include bad statistics and pathological events, (Planinic, 2010). For example, if the test is ideally well constructed, the spectra of item difficulties and respondent abilities will align; any misalignment is then examined and interpreted. Pathological cases can also be detected, for example when $V(i,j) = 1,\ P(i,j) < 0.5$ and vice versa. In this way, the Rasch model provides a genuine "radiography" of testing, surveying, or questionnaire-based measurement, bringing to light latent elements that remain invisible to conventional measurement approaches. A very intriguing analysis consists of discussing the threshold parameter calculated when using Polytomous Rash model formula (3). Here parameter $\tau = \{\tau_0, \tau_1, \dots \tau_m)$ estimate the observed difference between Likert values of the difficulty like quantity. Based on this exposition, we believe that this model is highly suitable for multi category surveys, particularly when the categories represent an ordered and classifiable property for example, point–based scoring systems or Likert, type scales. With the aid of this model, it becomes possible to estimate the underlying latent variable, which may correspond to knowledge of a law or regulatory framework within an executive body, support for amendments, item–by–item assessment of the effectiveness of a legislative package, and similar constructs. The model also enables the identification of the structure of the latent variable, the removal of noise from the results, and the detection of numerous features that cannot be revealed through direct measurement or conventional statistical modelling.

2.3 Importance of Survey Piloting

In the case of surveys intended for complex measurements such as those involving the expression of attitudes or individual viewpoints, the validity, reproducibility, and reliability of the instrument are related with verification of the method and with the calibration of the survey itself. When the behavior under investigation exhibits features of complexity, considerations must be given to the effects of nonlinear dynamics and related phenomena. The need for calibration is thus linked to the targeted measurement accuracy and to the alignment of the methodological approach with the system being studied; in practice, these two components are interdependent. Accordingly, when research focuses on organizing basic methodological rules and steps, the measurement instrument is presumed to be calibrated, [42], and therefore, the scientific interest must be shifted to the system features, e.g., evidence of relaxation of the state and achieving sufficient homogenization. For example, if one aims to obtain accurate measures using the PISA test, selecting students randomly from the whole population in low-income countries is not adequate, because of the effect of hidden variables. So, one must select randomly from among students who have a computer at home and who are familiar with this mode of testing, (Prenga, 2025). This ensures homogeneity and equal testing conditions for all participants, even though the test itself is certified. However, when the questionnaire is neither standardized nor certified, pilot testing becomes an essential step particularly to avoid impractical equivalences or selective numerical approximations discussed in the research literature [41], or when working with more advanced techniques such as bootstrapping (sampling – resembling) [11], and other methodologies referred in [25], [27], 29] etc, etc. Through piloting the survey aimed at measuring latent variables or hidden effects in our sociometric mediums, we consider verification of the instrument features in the current medium, to especially ensure:

Validity, which encompasses: (1) content validity, achieved through the determination of the extent to which the questionnaire items adequately represent the domain of interest; (2) construct validity, meaning that the response corresponds directly to the intended construct of the question and is not influenced by extraneous elements; and (3) face validity, whereby the respondent clearly understands what the question is asking. Validity is assessed by obtaining expert judgment, typically quantified through an index CVR (content Validity Index) $CVR = \frac{N_e - 0.5N}{0.5N}$, where N denotes the total number of evaluating experts, and N_e represents the number of experts who have judged a given item as content-appropriate.

Reliability. Through re-testing during the pilot phase, the consistency of responses is assessed, including both internal consistency and external consistency (i.e., the extent to which different evaluators obtain the same results). In this analysis, Cronbach's alpha is one of the most widely used reliability indices. The coefficient is defined as follows: $= \frac{n}{n-1}(1 - \frac{\sum_{i=1}^{n}\sigma_{x_i}^2}{\sigma_x^2})$, where n denotes the number of items in the questionnaire, $\sigma^2(x_i)$ represents the variance associated with item i, and σ^2 denotes the total variance. The sufficiently high value of Cronbach's alpha at the pilot stage constitutes a strong indication that the measurement instrument is functioning adequately.

After considering CVR analyses for a draft survey that we piloted in a small sample of experts, we realized that when using odd numbers in the Likert Scale, more than 28% of voluntary participants have chosen pragmatically the mid-point (the value 3 for the 5-scale model), though they had the needed expertise on the field. Also, in a dedicated questionary for the quality of the schoolbooks at high education system in the country, several items appear infected by this syndrome, even though the organizers were confident that respondents acknowledge the problematics therein. When using 4 scale evaluation in a later address, it resulted that physics textbooks do have major shortcomings in the formulation of very basic concepts such as the force! The results have been reported in the framework of master's degree thesis in physics, but it reveals the relevance of piloting for our systems that again, cannot be assumed as a direct application of general rules. We also believe that this feature is of a more general nature, almost for online surveys.

2.4 Use of other indices to improve measuring instruments

As noted earlier, the researcher can improve the test and thereby serve the measurement process, but cannot influence the system itself, which may be non-homogeneous, anisotropic, and non-relaxed.

Table 1. Indexes and some comments

Index name, Desired range	*Formula/*	Our explanation if out range	Comments
Difficulty 0.3-0.9	$\delta_i = \frac{n_{Correct}^{answer}}{N}$	Confusing question	Biased supportiveness
Discrimination index >0.3	$d = \frac{(N_{highestScore}^{Succses} - N_i}{N}$	Not suitable for distinguishing e attitudes	Binary variable is in-adequate
Cronbach alpha >0.2	$r_i = \frac{\tilde{x_i} - \tilde{x}}{\sigma(x)}\sqrt{\frac{\delta_i}{1-\delta_i}}$;	No solid knowledge of the subject's	Possible random selection
Discriminatory power >0.7	$\rho = \frac{n_i}{n_i - 1}(1 - \frac{\sum_{i=1}^{n_i}\sigma(x_i)}{\sigma^2(x)}$	Not clearly distinguish between attitudes.	Binary variable is in-adequate
Self-consistency >0.7	$\Delta = \frac{N^2 - \sum_i^{n_i} f_i^2}{N^2 - \frac{N^2}{n_i+1}}$	Questionnaire is not consistent.	Heterogeneity issues are present

Likewise, respondents may provide biased judgments, offering evaluations driven by political perception rather than rational assessment. Others may give general evaluations without having sufficient knowledge of the issue. Given these considerations, the analytical treatment of questionnaires dealing with sensitive topics such as the assessment of legal performance, where multiple factors intervene requires an evaluation of the procedure itself, including the administration of the questionnaire and the collection of responses. In such cases, the use of general-purpose measurement tools is difficult, as each questionnaire and interviewing procedure has its own specificities. However, in similar contexts where validated measurement instruments exist such as Concept Inventory tests used to assess conceptual learning see [36], we have used recently indexes broader s methods have been applied successfully Notice that based on the post factum nature of this analysis, it can be used as confirmatory or rejection tools for the measuring instrument used in a concrete measurement. It competes with the Rashc analysis but has a more general nature. These adaptations are currently undergoing revision and are included here merely as an illustration of the practical feasibility of refining ordinary questionnaires through index-based approaches. If the questionnaire items are carefully adapted so that index theory, as described above, becomes applicable, then the indices may serve as useful indicators of the questionnaire's validity, completing the CVR initial evaluation. Based on this premise, we conducted a pilot measurement for a questionnaire addressing the performance of several policy measures and legislative reforms, analogous in structure to Concept Inventory instruments used to measure scientific understanding within disciplinary contexts.

To approximate the construct of "evaluative judgment" regarding policy implementation, agreement with a policy statement was coded as 1 and disagreement as 0. It should be noted that the criterion of "adequate knowledge" was not speculative but based on verified familiarity with the subject matter. The primary aim was to evaluate the instrument itself, a procedure that is often overlooked in the development of questionnaires. The questionnaire was administered through purposive sampling, i.e., to a reduced group in which all participants held a master's degree and possessed solid knowledge of the issues addressed. This was censured by selecting civil service employees with at least 10 years of experience (most respondents had over 15 consecutive years of work in the civil administration or in teaching). In this pilot test, 25 responses from administrators and 25 from teachers were qualified for analysis. The results indicated that most items in the questionnaire are outside the desirable range; in other words, our test did not effectively elicit stable or meaningful judgments. Interestingly, although index verification revealed structural weaknesses, the reliability and validity check performed through the Content Validity Ratio (CVR) based on evaluations from several domain experts produced an acceptable value (CVR ≈ 0.67). The questionnaire is currently undergoing corrective revision within a broader research project. Here, we observed that while the formal formulation of the test may sound well-grounded, post-pilot index verification can reveal how respondents react to questions that involve both knowledge and beliefs associated with the latent variable in this case, administrative performance. In table 1 we have resumed our idea of commenting on the finding by indices and subsequently considering significant review the instrument used. We observed that by several attempts to enhance indices of a low scale questionnaire the use of polytomous variable has been fruitful in the case of purposive sampling, and mostly in cases where enquiries have been directed to the expert community only. When using random sampling we observed that a total improvement of the indexes has not been observed. By this procedure, we have identified indicators that judiciary's support for major reforms in the country was not stable among different samplers, and also, there is no common belief on the performance of administrative directors in public and non-public entities. Based on those preliminary findings we have considered using Likert Scale questionnaire and Rasch Polytomous model approach for better measurement. It seems contradictory, but first results theta measuring directedly involve larger error and indeterminacy than indirect measurement.

3 Improving hypothesis testing procedures

Based on observations of surveys and questionnaires conducted locally using probabilistic sampling (random sampling), it appears that organizers typically provide information on formal elements such as the final results, margins of error, and similar indicators. However, as a rule, they do not present any justification that the variable measured by the questionnaire is IID (independent and identically distributed). Only when this condition is satisfied do sample means obtained from different random selections follow a normal distribution, which in turn allows the application of familiar statistical formulas, including estimation procedures and hypothesis testing. The distributional behavior of the variable under investigation is crucial for determining the appropriate sample size, which itself must also be drawn randomly.

3.1 Evidence on distributions features

In general, the estimators of interest in this work relate to individual attitudes, prompting the preliminary question: "Are the distributions of attitudes of the same functional nature?" If the answer is yes, then according to the Central Limit Theorem, the distribution of sample means of IID random variables is Gaussian (and for small samples, t-Student), and the problem is solved we can directly apply standard statistical formulas for sampling, deviation estimation, and determining Type I and Type II errors. Thus, the first step is to verify the ID quality of the variable under consideration namely the attitude, preference, support, representativeness, etc. To illustrate the existence of this issue, we bring several theoretical arguments and examples. Consider a measurement for the average attitude $m = \frac{1}{N} \sum_{i=1}^{N} x_i$ in a large population, providing that $x_i = 1{,}0$. If we assume that real supportiveness

$X^{real} \sim N(m_0, \sigma)$, a distinctive approach is to use aggregate binary data contemplating the fraction of the answers $p = m$ and using z-score $z = \frac{m-m_0}{\sqrt{\frac{m_0(1-m_0)}{n}}}$. Also, we might use e logit transformation $ln\left(\frac{m}{1-m}\right)$ to get continuous distributions which are pleasurable for general analysis. In this case, if we assume that all individuals have the same distribution of theory decision hidden variable, which concludes onto Accept/Reject answer, the z-score variable is normally distributed and therefore, we might proceed with a rigor statistical analysis, finalizing correct statement regarding the measurement. For example, we might assume that this hidden variable is normally distributed, or has a unform distribution in [0,1] etc. It follows that in general, the assumption of the IID is not questioned when making statistical observations, but if this is not the case, the resulting measurement or the regressions, modeling etc., are not correct. To clarify the idea that distribution of a certain variable could be specific and not unique in a community, we might consider the relationship between the mechanism that generates a connection in a social graph and the corresponding distribution. So, for electoral graphs, in [16], [17], it is shown that the distribution of candidates receiving a given proportion of votes follows a power (Zipf law, $k = \frac{1}{v^a}$) which is supported by theoretical works of [6], [9] **for** random networks with the preferential attachment mechanism, i.e. $P_{i-j} = \frac{l_j}{\sum_{k=1}^{j} l_k}$ where l denotes the number of links of node (j) at the moment when node (i) enters the system. Also, in the presence of strong political polarization, such as majoritarian elections, the distribution is approached to a Gaussian, [24] and more complex cases are observed in [27]. Therefore, we expect that the distributions of attitudes, preferences, etc., are not necessarily unique functions across the population. This implies that variables of this category are not IID. Specifically for our systems, in [20] it has been shown that the votes of candidates from major parties follow a bell-shaped but non-Gaussian distribution, and the votes of candidates from small parties follow a power-law distribution. It is suggested also that other preferences are not IID [18], and consequently, the sampling for measuring such aspects in our concrete must be reconsidered. Through this discussion, we conclude that when measuring in a large population we must proceed with a check for IID. In this regard, we have administered the raw data from a set answer provided for 20 questions survey on the supportiveness for government reforms ($which\ we\ are\ asigning\ by\ S_i$) that we have analyzed in our recent work [35]. Some procedures and activities were repeated in succeeding work which is under publishing procedure. So far, we have got 85 responses but only 65 forms were answered fully, and so, usable for analysis. Of those, 43 belong to university education and experts, and 5 were of doctorate degree or registered in PhD course.

Table 2. The surviving questionnaire after piloting

Taxes' policies	Financial stability
Education development and strategies	Social care and projects
Crime convictions and general order	Wages sectors improvement
Environment care and management	Legislation in general (completeness)
Youths' perspective and employment	Rule of law
Law enforcement	Justice and reforms
Elderly support and policies	Free market guarantees and medium of doing business
Military advance in the framework of recent development.	Accuracy of Legislation for Proprieties and public terrains
Health and medicine	Effectiveness of legislation for assets
Foreign affairs performance	Management of political mediums and democracy

Clearly, we had a very small sample, and initially we observed that several cases did have their indices out of desired range for indices and initially we have attributed this behavior to the clearness of the draft formulation. However, the non-unique IID could be among the causes, despite rigor analysis not possible dye to the ultra-small quantity of data for every level of education (12-32-5), so we have conducted a description of possible occurrences of non IID symptoms. For these purposes we have considered two simplest methods. First, the set of data is portioned in three subsets by identifying a geographical parameter (rural area, small cites) and as said above, by education level (high school or equivalent, university degree, professional diploma including doctor degree). For ethical reasons, we choose numerical qualifiers (I, II, III) in both cases. Then we calculated arithmetic meaning and its variance. The idea of such calculation for the very small number of data is based on the following assumption: each item is answered based on the latent traits (the supportiveness for government performance), so practically we have 20 'random' values for every individual. In this context, the supportiveness calculated for a person (i) represent in fact an averaged over 20 values. Though this is not very correct, it supports the idea that S_i can be regarded as a random value more than an isolated point one, legitimating to some extent the calculation of subset means m_I; m_{II}; m_{III} and the variance $\sigma(\sigma_i, \sigma_j, n_i, n_j)$ where indices (i,j) are se respectively I,II, II. So far, it has resulted that for up to 80% confidence level, the values obtained were different, $m_I \neq m_{II} \neq m_{III}$ in both cases. We do not provide details in numerical values as they have little significance due to small numbers of the subsets, but we believe that the results qualitatively have supported the presence of the non-unique IID effect. By direct testing we adopted the test using - variance in all combinations. In another similar observation we have considered a survey conducted anonymously among 386 physics and mathematics high mathematics teachers, that has been conducted for other purposes. Again, we had selected only the questionnaires fully completed, that were 143 in total. In this case, 8 items out of 25 questions comply with our requirements for instrument that measure a latent variable, which here is identified as the teacher's supportiveness for efficacity of the education practices and conduct in the country. In this case we do not observe significant change between common variables such as 'physics/mathematics teacher, nor for regional/small cities/big cities allocation of their school. Alternatively, we have drawn randomly 10 subsamples of 70 records which are used to construct histograms and for identifying the underlying normal distribution of their z-score parameter. The average supportiveness for the full sample has been evaluated at $33\% \pm 5.5\% \div 42\% \pm 6.2\%$ indicating a large spread of the values. The t-test confirmed again that the supportiveness variable follows a non-unique distribution and that the heterogeneity is significantly high. By those empirical observations we concluded that the syndrome of non-unique IID is present in our systems so, a rigor check for its presence is recommended in case of measurement by using questionnaire, survey. It becomes more important when we measure latent variables. Theoretically, for those cases we must address the variance of the form $Var\big(\{X_i\} + \{X_j\}\big) = var\big(\{X_i\} + var\{X_j\} + 2cov(\{X_i\}, \{X_j\}\big)$, where the curl sign includes the sum of variables or their set. Noticing that if variables are independent, the covariance term is dropped and $\sigma^2 = \sigma_1^2 + \sigma_2^2$ if (1) and (2) are two subpopulations with distinct IID, so in principle it can be calculated, but Firstly, we must consider that the mean or expected value of composite distribution is $\mu = \sum x_i \rho_i + \sum x_j \rho_j$ where ρ represent the underlying distribution, so we must know the subpopulations size. It followed that if the non-uniqueness of the distribution is verified, the measurement procedure must comply with advance statistics, that is ignored in most of the applications in practice, or at least, as by observations, it is not clarified. In this case, sampling procedure must be revised substantially. Whereas those occurrences are of a general nature for similar systems, scholars and practitioners could skip this analysis by increasing the sample size, but again, this cannot be realized by solely guessing or by increasing sample at abnormal level, so, this precaution is advisable for in a broad sense for sociometric measurements.

3.2 Evidencing the stationary

An important issue in data processing and modeling is stationarity**.** This aspect becomes especially significant when the measurement is indirect**,** i.e., obtained through a model. As noted earlier, quantitative and computational political science has now become a standard research and measurement practice; therefore, such a requirement must be considered a priority. For illustration, we refer to the findings of [19] regarding the parameters of the SIDARTHE-V model for the COVID-19 dynamics in our context. According to this study, the parameters were identifiable and measurable only within a very narrow time window. The results showed that the model was applicable only within a very narrow period practically from November 2020 to July 2021. In all other periods, the series reproduced by the model did not match those obtained from the empirical data. Returning to the discussion of the measurements relevant to our study, the system in which these measurements are performed generally does not the system in which these measurements are performed is neither stationary nor linear and therefore requires specialized methods and professional expertise before applying any modeling procedures or simple statistical measurements. In some analysis of our co-authors, related to the electoral mediums in the country, the stationarity of the distribution is measured by using q-Gaussian, and un-stationary has been clearly evidenced, suggesting a rigor analysis of the distribution before sampling standard formula for calculations.

3.4 Preliminary analysis in the design of verification measurements and sequential sampling

In the design and measurements, it is crucial to provide a balanced assessment of error types and tolerances. Suppose we wish to test whether a certain factor has improved following the implementation of a new law. The corresponding values may be taken directly from official statistics or obtained through citizens' perceptions. In such cases, the analyst must first specify the null and alternative hypotheses. For example, the null hypothesis H_0 corresponds to no change, $\mu_1=\mu_0$ whereas the alternative hypothesis H_1 which reflects what we aim to test, denotes a change, $\mu_1\neq\mu_0$. Those measures are tightly dictated by small sample sizes, and the management of errors Type I and II. It turns out that reducing the Type I error inevitably increases the Type II error and vice versa; therefore, a delicate balance must be established [13]. On the other hand, increasing the sample size to reduce the Type II error and correct the imbalance mentioned above not only raises the cost of the survey, but may also raise ethical concerns in several applications [13]. These two elements determine the sample size, according to the theoretical formulation of Wald, Thus, if we aim to verify the non-change of a variable such as the peoples supportiveness, an overall index of performance etc., the sample size is given by $\mathrm{N_{sample}} = \frac{\frac{\left(z_{1-\frac{\alpha}{2}}+z_{1-\beta}\right)^2\sigma^2}{(\mu_1-\mu_0)^2}}{\left[1+\frac{\left(z_{1-\frac{\alpha}{2}}+z_{1-\beta}\right)^2\sigma^2}{(\mu_1-\mu_0)^2 N_{Population}}\right]}$, which becomes $N=\frac{\left(z_{\frac{\alpha}{2}}+z_{1-\beta}\right)^2\sigma^2}{(\mu_1-\mu_0)^2}$ that for infinite or undefined population. Referring to the discussions about the variance and the mean, those formulae reveal the importance of a rigor analysis of distribution for the variable under investigation in the sense of the standard conditions' fulfillment. and foremostly, on verification for the normality of the distribution of the mean. We are not going in detail herein because several mathematical steps must be included, but we're listing in short, some techniques used elsewhere that can be adopted for our purpose. We have used few of them in our research works, but only at qualitative level. In this case we do not consider calculation for very complicated cases where distributions are not known and not unique, but we point out the avoiding uncertified fulfillment of standard assumptions. However, if we obtain by other surveys or piloting measurement that the underlying distribution of the quantity, we are investigating is not normal, or poorly defined, we might implement Chebyshev inequality, [30], [31]

$$P(-k \leq m-\mu \leq k) \geq 1-\frac{\sigma^2}{k^2 n} \quad (7)$$

where k is the difference that we want to detect or are legitimate to tolerate, and σ is the standard deviation. This is very helpful if we are measuring proportions of the population because we might assign standard deviation at 0.5, whereas for other applications, these values must be known. Notice that it is supposed to be known when applying standard sampling size calculation, too. Hence, by putting the left side of (7) at $1-\alpha$ where α is the risk error that we tolerate, mostly at 0.05, the sample size is evaluated $n=\frac{\sigma^2}{\alpha k^2}$. In a very special case where we assume the tolerance k in standard deviation units, the calculation is simplified further. We used this formula to legitimate our very small sample size in pilot testing. So, when initiating our pilot analysis to check for guessing in the filling of the forms imposed by our questions formulations, we considered that standard deviation is 0.5, and k=1, so n=50 at the 0.05 risk level. By this sampling we observed that our initial questionary (Table 2) was confusing because by the rash analysis we observed a high level of guessing or midpoint selection for multiple scale variables. A general guideline for such application can be found in standard statistics schoolbooks, but we are highlighting some useful steps. Firstly, we might use a preliminary test to verify if the distribution in the sample is the same as the one of whole population. A Kolmogorov-Smirnov test can be used. Finally, if we can formulate alternative hypothesis, an opportunistic solution is to make use of sequential sampling [31]

Conclusions

This paper presented a methodological and descriptive analysis of the measurement process for specific medium and system where heterogeneity and other deviation from the standard assumption on sociometric measurement are present. The work offers a synthesized overview motivated by the increasing interest in quantitative analyses within the fields of auditing, legislative performance evaluation, and regulatory efficiency, supported by illustrative examples from our recent applications. We emphasize that ensuring the quality and rigor of measurements in sociometric and political-science applications is essential for the credibility of conclusions and for the practical applicability of the findings.

References:

[1] Hitt, M. P., Volden, C., & Wiseman, A. E. (2017). Spatial Models of Legislative Effectiveness. American Journal of Political Science, 61(3), 575–590.

[2] Van Aeken, K. (2011). From Vision To Reality: Ex Post Evaluation of Legislation. legisprudence, 5(1), 41–68.

[3] Kollman, Scott E. Page,Computational Methods and Models of Politics, Editor(s): L. Tesfatsion, K.L. Judd, Handbook of Computational Economics, Elsevier, Volume 2, 2006, Pages 1433-1463,

[4] Anglmayer, I., & Scherrer, A. (2020). Ex-post evaluation in the European Parliament: an increasing influence on the policy cycle. The Journal of Legislative Studies, 26(3), 405–426.

[5] Marco Battaglini Valerio Leone Sciabolazza Eleonora Patacchini. Effectiveness Of Connected Legislators Working Paper 24442 Http://Www.Nber.Org/Papers/W24442 National Bureau Of Economic Research 1050 Massachusetts Avenue Cambridge, MA 02138 March 2018

[6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286: 509, 1999.

[7] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. PNAS, 79: 2554, 1982

[8] N. R. Mantegna, E. H. Stanley, An Introduction to Econophysics, Correlations and Complexity in Finance, UK: Cambirdge University Press, (2000)

[9] A. L. Barab´asi, R. Albert (2002). "Statistical mechanics of complex networks". Reviews of Modern Physics. 74 (1): 47–94.

[10] Trenton D. Mize, Bianca Manago (2022), The past, present, and future of experimental methods in the social sciences, Social Science Research, Volume 108, 2022, 102799,

[11] Hrba, M., Maciak, M., Peštová, B., & Pešta, M. (2022). Bootstrapping Not Independent and Not Identically Distributed Data. Mathematics, 10(24), 4671.

[12] Bhattacherjee, Anol. "Social Science Research: Principles, Methods, and Practices" (2012). Textbooks Collection.

[13] Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021 Feb 15;31(1):010502. Epub 2020 Dec 15. PMID: 33380887; PMCID: PMC7745163.

[14] Lamprianou, I. (2019). Applying the Rasch Model in Social Sciences Using R. Routledge.

[15] Taichi Yamashita, Analyzing Likert scale surveys with Rasch models, Research Methods in Applied Linguistics, Volume 1, Issue 3, 2022, 100022, ISSN 2772-7661,

[16] Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of modern physics, 81(2), 591-646.

[17] Lyra, M., Costa, U. S., Costa Filho, R. N., & Andrade, J. (2003). Generalized Zipf's law in proportional voting processes. Europhysics letters (Print), 62(1), 131-137.

[18] Prenga, D., & Ifti, M. (2012). Distribution of votes and a model of political opinion formation for majority elections. In International Journal of Modern Physics: Conference Series (Vol. 16, pp. 1-12). World Scientific Publishing Company.

[19] Denaj, A., Prenga, D., & Tahiri, V. (2023, September). General features of the time data series of Covid-19 in Albania. In AIP Conference Proceedings (Vol. 2872, No. 1, p. 030003). AIP Publishing LLC.

[20] Prenga, D., Peqini, K., & Osmani, R. (2021, November). The analysis of the dynamics of the electorate system by using q-distribution-a case study. In Journal of Physics: Conference Series (Vol. 2090, No. 1, p. 012073). IOP Publishing.

[21] Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561–73.

[22] Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV (pp. 321–334). Berkeley, California: University of California Press.

[23] Mobilia, M. (2023). Polarization and Consensus in a Voter Model under Time-Fluctuating Influences. Physics, 5(2), 517-536.

[24] Mobilia, M. Does a single zealot affect an infinite group of voters? Phys. Rev. Lett. 2003, 91, 028701.

[25] Clara Monteiro Vieira, Elisabeth Costa Monteiro, Rasch model and anchoring vignettes for metrological characterization of a democracy measuring system: preliminary studies, Measurement: Sensors, Volume 38, Supplement, 2025, 101507.

[26] John Cornish (2002). RESPONSE PROBLEMS IN SURVEYS Improving response & minimizing the load.

[27] Reed JE, Johnson JK, Zanni R, et al. Quality of locally designed surveys in a quality improvement collaborative: review of survey validity and identification of common errors. BMJ Open Quality 2024;13:e002387.

[28] Giuliano, F., Soriano, M. A., Pino-Mulero, V., & González-García, J. (2025). Testing Consistency: Analyzing the Reliability of Two Lower Limb Isometric Force Measurements in Strength-Trained Athletes. Applied Sciences, 15(1), 303.

[29] Lea JWD, O'Driscoll JM, Coleman DA, Wiles JD. Validity and reliability of the 'Isometric Exercise Scale' (IES) for measuring ratings of perceived exertion during continuous isometric exercise. Sci Rep. 2021 Mar 5;11(1):5334. PMID: 33674722; PMCID: PMC7970867.

[30] Todinov, M.T. (2020). Risk and Uncertainty Reduction by Using Algebraic Inequalities (1st ed.). CRC Press.

[31] R.H. Riffenburgh, Statistics in Medicine (Third Edition), Academic Press, 2012,

[32] David Hestenes, Malcolm Wells, and Gregg Swackhamer. Phys. Teach. 30, 141 (1992).

[33] Noelle M. Crooks, Martha W. Alibali. Defining and measuring conceptual knowledge in mathematics. Developmental Review 34 (2014) 344–377

[34] Antti Savinainen and Philip Scott. The Force Concept Inventory: a tool for monitoring student learning. 2002 Phys. Educ. 37 45

[35] Llakaj, A., Prenga, D. (2025). Remarks on extension of common models for studying legislation, coherence and effectiveness. 3rd International Conference on Pioneer and Innovative Studies. October 06-07, 2025: Konya, Turkey

[36] Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. Physical Review Special Topics—Physics Education Research, 2(1), 010105.

[37] Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of Psychology, 140, 1–27.

[38] Prenga, D., Kushta, E., & Musli, F. (2023). Enhancing Concept Inventory Analysis by Using Indexes, Optimal Histogram Idea, and the Likert Analysis. Journal of Human, Earth, and Future, 4(1), 103-120.

[39] M. Planinic. Assessment of difficulties of some conceptual areas from electricity and magnetism using the conceptual survey of electricity and magnetism. American Journal of Physics, 74(12), 1143–1148.

[40] Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. American Journal of Physics, 73(12), 1172-1182.

[41] Nisbet, R., Elder, J., & Miner, G. (2009). Numerical Prediction. Handbook of Statistical Analysis and Data Mining Applications, 259–284.

[42] McNabb, David E. (2010) Research methods for political science: quantitative and qualitative approaches. Political science—Methodology. JA71.5.M34 2009. 320.072—dc22 2008049000

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.