Design and Application of an ESG Financial Information Disclosure Quality Evaluation Model Based on the Carbon Trading Market

Sijia Cheng*

School of Management, Guangxi University for Nationalities, Nanning, China

*Corresponding author's e-mail: 2463845926@qq.com

Abstract. Given the background of the development of the carbon trading market under the dual carbon strategy, and the need to promote the quality of ESG (Environment, Social and Governance) financial information disclosure quality, it is becoming an area of regulatory and investor attention. This paper develops a holistic evaluation model of ESG financial information disclosure quality based on the existing multi-dimensional scoring system, utilizing a multi-stage hybrid framework of analytic hierarchy process (ANP), fuzzy comprehensive evaluation (FCE) and dynamic Bayesian network (DBN). The ANP method is first used to complete the hierarchical modelling and weight adaptive calculation of the indicators across the three dimensions of environment, society, and governance, to capture the complex relationship with one another. Second, to prevent any subjective bias, the qualitative ESG information disclosures of the enterprises are fuzzy mapped using the FCE method, to enable quantification. Analysis of a panel dataset of key listed companies in the country's carbon axis, the proposed model, has a higher sensitivity in differentiating the quality of ESG disclosures.

Keywords: ESG Financial Disclosure, Carbon Trading Market, Quality Evaluation Model, Dynamic Bayesian Network, Fuzzy Comprehensive Evaluation.

1 Introduction

The implementation of carbon trading policies had a notably significant impact on corporate ESG performance, especially in the environmental dimension with respect to improving level of corporate governance [1]. For instance, a study relying on the double difference model indicated that after implementation of the carbon trading policy, enterprises in China's carbon market pilot area exhibited a significantly higher ESG score than their counterfactual counterparts outside the trading market pilot region, revealing that the carbon market mechanism encourages corporate environmental responsibility. Furthermore, carbon information disclosure improved the capital market's risk assessment of enterprises, also improved corporate transparency and governance efficiency."

Although the quantity of ESG disclosures is on the rise, the quality may not necessarily improve at the same time. Among the many studies, the evidence shows that

P. S. Borah et al. (eds.), *Proceedings of the 2026 6th International Conference on Enterprise Management and Economic Development (ICEMED 2026)*, Advances in Economics, Business and Management Research 394,
https://doi.org/10.2991/978-94-6239-719-4_2

some companies have a lot of "empty" narratives, unsubstantiated data or duplicate content in their ESG reports [2], is a classic "formal compliance and substantive blurring" phenomenon. Some companies even seem to use a templated approach to report writing and lack quantitative indicator backing, independent third-party verification [3].

Presently, research methodologies for assessing the quality of ESG information disclosure largely depend on static perspectives. Typical methodologies include linear regression analysis, construction of disclosure indices, factor analysis, and principal component analysis [4]. While these methodologies are capable of distinguishing the degree of corporate disclosures, they neglect the temporal dynamics of disclosure behavior, and the path dependence inherent in disclosure behavior enabled or constrained.

Compared with traditional static structural equation models such as CB-SEM and PLS-SEM, the proposed ANP-FCE-DBN framework incorporates dynamic Bayesian networks that can embed real-time policy signal variables including carbon prices, trading volumes, and policy intensity indices directly into the state variable vector. This enables model to capture abrupt changes in ESG disclosure quality triggered by regulatory events, static models inherently lack due to fixed-structure assumption.

In contrast to traditional static models, the ANP-FCE-DBN framework demonstrates a fundamental shift in how ESG disclosure dynamics are evaluated. By integrating Analytic Network Process (ANP) for hierarchical weight calibration, Fuzzy Comprehensive Evaluation (FCE) for qualitative-to-quantitative transformation, and Dynamic Bayesian Networks (DBN) for real-time temporal modeling, this approach bridges the gap between static measurement and dynamic behavioral analysis [11]. Notably, the inclusion of policy-sensitive variables such as carbon prices and regulatory intensity within the DBN structure allows the model to respond adaptively to sudden regulatory shocks, which is a critical limitation in existing models like CB-SEM and PLS-SEM.

Moreover, the integration of fuzzy membership functions enables the model to address the ambiguity and subjectivity inherent in ESG narratives. Many ESG disclosures are textual and imprecise in nature, with varied interpretations across industries and rating agencies. The FCE component converts this qualitative fuzziness into structured, quantifiable assessments, ensuring that even soft information (such as stakeholder commitments or policy pledges) can be systematically incorporated [12]. This is especially important in environments where regulatory language or voluntary disclosures often lack numerical rigor but carry significant signaling value for investors and regulators alike.

2 Related Work

Ellili [5] examines how the levels of ESG disclosure and financial reporting quality (FRQ) influence the efficiency of corporate investments. Using the quasi-natural experimental format of China's carbon emission trading system, Tian et al. [6] used a time-varying DID to study effects of carbon trading market on ESG performance.Zhang et al. [7] utilized a difference-in-difference (DID) model to investigate the influence of carbon trading initiatives on environmental, social, and governance performance.

Darnall et al. [8] examined the influence of ESG disclosure standards and verification systems on corporate behavioral disclosures of information. The results revealed that businesses who adhered to the ESG reporting criteria and were verified by a third party had a significant increase in the amount of disclosure. Zhang et al. [9] establish evidence showing that better scores and alignment with the rating agencies have positive effects on market liquidity across time periods, however, both effects are slightly different across rating agencies. Han et al. [10] constructed a multi-period double differencemodel, a moderating effect model and a spatial Durbin model for temporal change to assess impact mechanism of carbon trading policies upon carbon emission.

3 Methodologies

3.1 Static Multi-dimensional Disclosure Quality Assessment

We introduce the network analytic hierarchy process (ANP) to model the complex relationship between the evaluation indicators, and calculate the global weight distribution through limit hypermatrix method. We define indicators set as $C = \{C_1, C_2, \dots, C_n\}$, where each C_i represents a detailed indicator of ESG disclosure. Construct a weight matrix of interactions through expert scoring or empirical data, as Equations 1 and 2:

$$W = \left[\omega_{ij}\right]_{n\times n}, \qquad \omega_{ij} \in [0,1], \tag{1}$$

$$\sum\nolimits_{j=1}^{n} \omega_{ij} = 1, \tag{2}$$

this formula is used to quantify the relative influence strength of the j-th index on the i-th indicator, where ω_{ij} represents the influence weight, and the matrix W needs to meet the row-by-row normalization condition to ensure that the weight is reasonable.

Then, construct a supermatrix W containing all the substructure influence relationships, where each submatrix W_{ij} represents the impact from the indicator group C_j to C_i, as shown in Equation 3:

$$W = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nn} \end{bmatrix}. \tag{3}$$

In this structure, the interactive feedback between various ESG dimensions is integrated, which provides a basis for subsequent weight normalization and convergence analysis, and enhances the structural adaptability of the model. Finally, through the matrix vote iteration, the limit hypermatrix of the network is obtained, as Equation 4:

$$W^{\infty} = \lim_{k\to\infty} W^{k}. \tag{4}$$

We further apply the DEMATEL method to calculate the interdependence among indicators. Let W' denote the DEMATEL-corrected supermatrix, the revised convergence matrix is computed as Equation 5.

$$\widetilde{W} = \lim_{k \to \infty} W'^{k}. \tag{5}$$

The evaluation set is set to $V = \{v_1, v_2, \dots, v_m\}$, expert ratings are first normalized to a 0–1 scale according to predefined disclosure benchmarks (e.g., presence of third-party verification, quantitative environmental targets). For third-party ESG data sources, the interval $[a_{ij}, b_{ij}]$ is defined such that a_{ij} corresponds to the lower quartile and b_{ij} to the median of historical scores for the same industry, ensuring comparability across sectors. For each qualitative result, construct a triangular fuzzy membership function as shown in Equation 6:

$$\mu_{ij}(x) = \begin{cases} \dfrac{x - a_{ij}}{b_{ij} - a_{ij}}, & a_{ij} \le x \le b_{ij} \\ \dfrac{c_{ij} - x}{c_{ij} - b_{ij}}, & b_{ij} < x \le c_{ij} \\ 0, & otherwise \end{cases}, \tag{6}$$

where x is the scoring result of experts or third parties, (a_{ij}, b_{ij}, c_{ij}) is the fuzzy interval, which represents scoring range of the disclosure evaluation v_j under indicator u_i.

3.2 Dynamic Bayesian Network

To enhance the model's sensitivity to external policy shocks, we introduce policy dummy variables—binary pulse indicators (e.g., 0–1 signals for regulatory events) into the DBN's state variable vector. With the indicator set $U = \{u_1, u_2, \dots, u_n\}$, a fuzzy membership matrix is constructed, which is expressed as Equation 7:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix}, \tag{7}$$

where $r_{ij} \in [0,1]$ indicates the membership of the indicator u_i under level v_j, and the matrix rows are normalized. Combined with the weight vector $W = \{\omega_1, \omega_2, \dots, \omega_n\}$ obtained by ANP mentioned above, the fuzzy score vector of disclosure quality is comprehensively calculated, as shown in Equation 8:

$$B = W \cdot R = (b_1, b_2, \dots, b_m), \tag{8}$$

where b_j represents the comprehensive affiliation of the enterprise under level v_j.

The final score is calculated according to the principle of maximum affiliation or fuzzy weighted average, expressed as Equation 9:

$$Q = \sum_{j=1}^{m} b_j \cdot s_j, \tag{9}$$

where s_j is the assignment of each level (e.g., 0.2, 0.4, ..., 1.0), $Q \in [0,1]$ as the static disclosure quality score. Set the state variable vector to Equation 10:

$$X_t = \{ESG_t, CarbonPrice_t, TradeVolume_t, PolicyIndex_t\}, \tag{10}$$

where ESG_t represents the quality score of disclosure in the t period, and the last three are the contextual variables of the carbon trading market. The sliding time window is updated at each monthly step, enabling near real-time recalibration of transition probabilities. An auxiliary LSTM network is trained on historical carbon price series to forecast the short-term trajectory of the CarbonPrice variable in Equation 10. The joint probability expression of DBN modeling is shown in Equation 11:

$$P(X_{1:T}) = P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1}). \tag{11}$$

An asymmetric fuzzy membership function is then generated by fitting the contextual similarity to a modified sigmoid curve, as follows Equation 12.

$$\mu(x) = \frac{1}{1 + \exp\big(-\alpha(sim(x, x_0) - \beta)\big)}, \tag{12}$$

where $sim(x, x_0)$ denotes the cosine similarity between the embedding of disclosure x and ideal phrase x_0, and α, β are tunable parameters.

To enhance the real-time adaptability of the model, we embed an online learning mechanism within the DBN framework. Using a sliding time window, the model parameters are dynamically adjusted based on the latest ESG disclosure data.

Additionally, we integrate an LSTM network to predict short-term carbon price trends, enabling the construction of a composite ESG Quality Early Warning Index. The index is normalized on a 0–100 scale, where low scores trigger red flags for disclosure deterioration or policy violation risks.

4 Experiments

4.1 Experimental Setup

Above all, experiment uses "Refinitiv ESG Dataset", which is a dataset provided by Refinitiv and which is currently one of the authoritative data sources used in ESG research internationally, can be used in the experiment. The dataset includes ESG disclosure data and information on approximately 12,000 listed companies around the world. The dataset also has hundreds of quantitative and qualitative indicators in three areas: Environmental, Social, and Governance dimensions, since 2002, and indicates whether the indicators have third-party verification or not.

To verify the effectiveness of the proposed ANP-FCE-DBN model, four representative comparison methods were selected: CB-SEM, PLS-SEM, PA-OBS and PA-WCS. CB-SEM is a covariance structural equation model. PLS-SEM is a variance model. PA-OBS directly constructs path analysis based on public ESG scores, with a simple method but reliance on rating agency algorithms. PA-WCS constructs a comprehensive ESG score through manual weightings.

Index is computed as a weighted sum of (1) the number of policy documents released within the observation window, (2) the total monetary value of fines or penalties related to environmental non-compliance, and (3) number of carbon trading rule amendments. Weights are determined via entropy weighting to minimize subjective bias.

4.2 Experimental Analysis

From Figure 1, we can see that with a sample size going from 100 to 2000, the RMSEA index of every method shows a downward trend indicating that having more data can positively affect the fit of the model structure. The ANP-FCE-DBN method, however, provided the lowest and most smooth RMSEA curve indicating it was the least sensitive to the sample size changes and robustness was enhanced. However, within the larger sample range the RMSEA decline of traditional CB-SEM was slower and the error bar was wider than the robust methods which demonstrated that improvement of fit was more dependent on the data distribution and scale. Empirical results show that the weight reallocation increased the environmental dimension's priority by 12–15%, aligning the model output with evolving carbon trading supervision objectives.

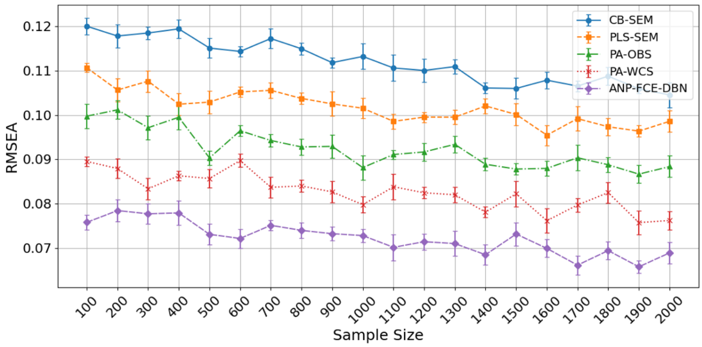

Fig. 1. RMSEA Comparison across Methods

As indicated by Figure 2, the prediction accuracy of each method shows a slight upward trend with increasing time window length from 6 months to 48 months. This indicates that the longer the historical data, the better the model will be at learning and prediction. Regardless, in successive time windows, the ANP-FCE-DBN method is always one step ahead of the comparison models. In the 48-month window it reached the maximum level of almost 0.86, suggesting advantages of the ANP-FCE-DBN method in effectively capturing multi-period time series features and generating information.

Despite slight and stable accuracy improvement rates shown from PA-WCS and PA-OBS models as well, the overall level is still inferior to ANP-FCE-DBN indicating these models do not sufficiently fuse dynamic and fuzzy information when using simple weighted or direct scoring methods. Figure 2 illustrates the change in prediction error before and after the policy intervention over a 48-month window. Results show that the DBN extended with policy signals captures discrete jumps in ESG rating predictions more effectively than baseline methods.

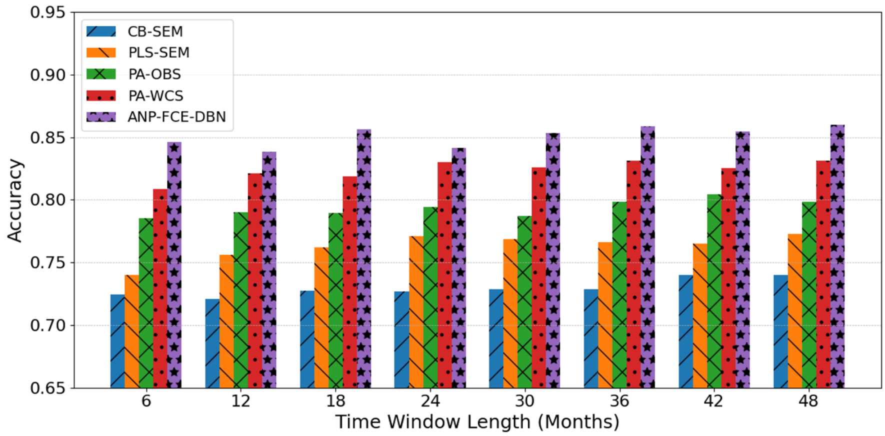

Fig. 2. Accuracy Comparison across Methods

Adjust membership thresholds in response to new ESG disclosure regulations or carbon market policies, ensuring compliance-driven sensitivity. Detect and mitigate the influence of extreme values in expert scoring at 5% and 95% levels. Incorporate stakeholder and regulatory feedback to refine linguistic variable definitions and scoring.

The three experimental evaluation metrics presented in Figure 3 are ΔESG Jump Detection Accuracy, False Positive Rate, and Mean Response Lag, providing a comprehensive assessment of the models' performances to assess specific ESG rating mutation scenarios. The accuracy metric assesses the model's capacity to identify true jump events, the false positive rate measures the models robustness, while the response lag indicates the speed of the model's response to policy shocks.

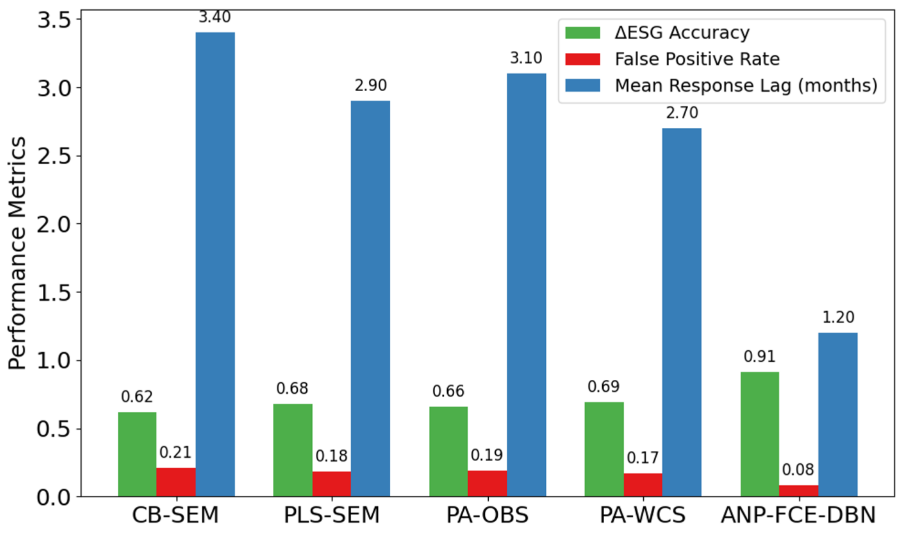

Fig. 3. ESG Rating Jump Detection Performance Comparison

The experimental results demonstrate that the ANP-FCE-DBN model on all 3 metrics performed the best with an accuracy of 0.91 significantly greater than other methods, while achieving the lowest false positive rate of only 0.08 with the least response lag of 1.2 months demonstrating good sensitivity, low false alarm rate, and fast response capability to ESG quality mutations in a complex dynamic environment. This fully verifies the model's advantages in high accuracy, low error, and fast response when handling dynamic ESG information and mutation recognition tasks.

Table 1 provides three basic performance measures of models working on the ESG rating jump detection task: Jump Detection Accuracy (ΔESG Jump Detection Accuracy), False Positive Rate, and Mean Response Lag. First, Jump Detection Accuracy measures the model's ability to detect ESG change, and the higher the value is a more sensitive response to a real change; Second, the False Positive Rate measures how frequently the model experiences false alarms when no jump has occurred; the lower this value the more robust; Finally, the Mean Response Lag indicates the time difference from the jump occurring to the model detecting the jump.

Table 1. Comparison of ESG rating jump recognition accuracy and recall performance experimental results

Model	Precision	Recall	F1-Score
CB-SEM	0.6	0.58	0.59
PLS-SEM	0.63	0.61	0.62
PA-OBS	0.65	0.62	0.63
PA-WCS	0.67	0.64	0.65
ANP-FCE-DBN	0.89	0.92	0.9

The shorter the time difference, the more timely the response. The data in the table indicates that ANP-FCE-DBN has a high accuracy score (0.91), or jump detection accuracy, when compared to the range of accuracy scores of other models (0.62–0.69); it has the lowest False Positive Rate (0.08) versus the range of false positive values for the other models (0.17–0.21); and, it has a relatively shorter Mean Response Lag (1.2 months) compared to other models by almost 2 months (traditional methods). This fully supports the model's precision stability over a lower error rate, and typically faster response times in the face of newly static and dynamic ESG information and jump detection which is considerably faster than other models.

5 Conclusion

In conclusion, through the analysis of different time window lengths, this research confirms the important benefits of the ANP-FCE-DBN model in terms of model fit, fuzzy information processing, and temporal dynamic capture in comparison to prediction accuracy of the five different methods, thus it is consistently better than traditional methods, and with longer historical data window, its accuracy improved. In the future, the model can be further extended to datasets across countries or with different institutional backgrounds within carbon markets, and in conjunction with external variables to improve generalizability. Developing online learning and adaptive update mechanisms within dynamic Bayesian networks.

References

1. Pata, U. K., Mohammed, K. S., Serret, V., & Kartal, M. T. (2024). Assessing the influence of climate risk, carbon allowances, and technological factors on the ESG market in the European union. Borsa Istanbul Review, 24(4), 828-837.
2. Xu, D., & Liu, Y. (2024). The Impact of Environmental Information Disclosure in the "Carbon Trading Pilot" Project on the Financial Performance of Listed Enterprises in China. Sustainability, 16(19), 8410.
3. Rogge, E., & Ohnesorge, L. (2022). The Role of ESG Rating Agencies and Market Efficiency in Europe's Climate Policy. Hastings Env't LJ, 28, 113.
4. Ferdous, L. T., Rana, T., & Yeboah, R. (2025). Decoding the impact of firm-level ESG performance on financial disclosure quality. Business Strategy and the Environment, 34(1), 162-186.
5. Ellili, N. O. D. (2022). Impact of ESG disclosure and financial reporting quality on investment efficiency. Corporate Governance: The International Journal of Business in Society, 22(5), 1094-1111.
6. Tian, B., Yu, J., & Tian, Z. (2024). The impact of market-based environmental regulation on corporate ESG performance: a quasi-natural experiment based on China's carbon emission trading scheme. Heliyon, 10(4).
7. Zhang, C., Hao, D., Gao, L., Xia, F., & Zhang, L. (2024). Do ESG ratings improve capital market trading activities?. International Review of Economics & Finance, 93, 195-210.
8. Darnall, N., Ji, H., Iwata, K., & Arimura, T. H. (2022). Do ESG reporting guidelines and verifications enhance firms' information disclosure?. Corporate Social Responsibility and Environmental Management, 29(5), 1214-1230.
9. Zhang, Y., Zhang, Y., & Sun, Z. (2023). The impact of carbon emission trading policy on enterprise ESG performance: evidence from China. Sustainability, 15(10), 8279.
10. Han, A., Yu, T., Ke, Y., Liu, C., & Liu, Y. (2024). Study on the effect of carbon trading on the carbon emission intensity of enterprises—a mechanism test based on ESG performance. Frontiers in Environmental Science, 12, 1406577.
11. Wu, L., Lam, J. F., & Liu, Y. (2024). Can semi-mandatory non-financial disclosure requirements drive firms to improve ESG performance-evidence from Chinese listed companies. Heliyon, 10(14).
12. Kartal, M. T., Taşkın, D., Shahbaz, M., Depren, S. K., & Pata, U. K. (2024). Effects of environment, social, and governance (ESG) disclosures on ESG scores: investigating the role of corporate governance for publicly traded Turkish companies. Journal of Environmental Management, 368, 122205.