

Article

# A Symmetrical Analysis of Decision Making: Introducing the Gaussian Negative Binomial Mixture with a Latent Class Choice Model

Irsa Sajjad <sup>1</sup>, Ibrahim Ali Nafisah <sup>2</sup>, Mohammed M. A. Almazah <sup>3</sup>, Osama Abdulaziz Alamri <sup>4</sup>   
and Javid Gani Dar <sup>5,\*</sup> 

<sup>1</sup> School of Mathematics and Statistics, Central South University, Changsha 410083, China; irsasajjad@csu.edu.cn

<sup>2</sup> Department of Statistics and Operations Research, College of Sciences, King Saud University, P.O. Box 2454, Riyadh 11451, Saudi Arabia; inafisah@ksu.edu.sa

<sup>3</sup> Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia; mmalmazah@kku.edu.sa

<sup>4</sup> Statistics Department, Faculty of Science, University of Tabuk, Tabuk 47512, Saudi Arabia; oalmughamisi@ut.edu.sa

<sup>5</sup> Department of Applied Sciences, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

\* Correspondence: javid.dar@sitpune.edu.in

**Abstract:** This research presents a model called the ‘Gaussian negative binomial mixture with a latent class choice model’, which serves as a robust and efficient tool for analyzing decisions across different areas. Our innovative model combines elements of mixture models, negative binomial distributions, and latent class choice modeling to create an approach that captures the complexities of decision-making processes. We explain how the model is formulated and estimated, showcasing its effectiveness in analyzing and predicting choices in scenarios. Through the use of a dataset, we demonstrate the performance of this method, marking a significant advancement in choice modeling. Our results highlight the applications of this model and point towards promising directions for future research, especially in exploring symmetrical patterns and structures, within decision-making processes.

**Keywords:** unsupervised machine learning; Gaussian mixture model; latent class choice model; negative binomial mixture model; indoor environmental quality; thermal comfort models



**Citation:** Sajjad, I.; Nafisah, I.A.; Almazah, M.M.A.; Alamri, O.A.; Dar, J.G. A Symmetrical Analysis of Decision Making: Introducing the Gaussian Negative Binomial Mixture with a Latent Class Choice Model. *Symmetry* **2024**, *16*, 908. <https://doi.org/10.3390/sym16070908>

Academic Editor: Xuelong Hu

Received: 11 June 2024

Revised: 8 July 2024

Accepted: 10 July 2024

Published: 16 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The discrete choice model (DCM) has rapidly become popular due to its practical applicability and theoretical robustness under individual preferences. The primary interest in the DCM was in the context of transportation. The contribution of [1] played a vital role from the 1970s to 1990. Since that time, the DCM has only been considered a realistic behavior of choice model when used as a closed-form model like nested logit, an open-form model like mixed logit, and a latent class model.

The satisfaction of occupants in office buildings is linked to various factors, including the indoor environmental quality (such as thermal, visual, and air quality and acoustic conditions), as well as the characteristics of the workspace and the building itself, such as the size, esthetics, furniture, and cleanliness. The ten studies listed in Table 1 have acknowledged the parameters that constitute occupants’ satisfaction with buildings (see, e.g., [2–16]).

**Table 1.** Research on workspace satisfaction factors.

Study	Data Analysis	Findings
[2]	Pearson correlation	Satisfaction with the indoor environment was associated with satisfaction regarding acoustics, thermal conditions, visual aspects, and air quality
[3]	Principal component analysis, Pearson correlation, and linear regression	The overall fulfillment was influenced by satisfaction with the thermal, acoustic, and lighting conditions, air quality, control over the indoor environment, level of privacy, as well as the office layout, decor, and sanitation.
[4]	Pearson correlation	Satisfaction with the indoor environment was positively associated with satisfaction regarding air quality, thermal conditions, lighting, acoustics, and spatial conditions.
[5]	Multiple linear regression	Satisfaction with the warmth, air quality, air circulation, noise, humidity, and lighting had an impact on overall workplace comfort.
[6]	Multivariate logistic regression	The suitability of the overall indoor environment was influenced by the acceptability of the thermal conditions, acoustics, lighting, and air quality.
[7]	Pearson correlation	Satisfaction with the workspace showed correlations with several factors, including lighting, noise levels, air quality, heating, drafts, available space, furniture quality, privacy, and the color and layout of walls and partitions.
[8]	Squared multiple correlations (SMCs)	CLS can advance the working environment of hospital staff employed in a neuro-ICU or PACU.
[9]	Exploratory and confirmatory factor analysis and structural equation modeling	Satisfaction with the indoor workstation environment was determined by factors such as noise, air circulation, air quality, temperature, lighting, privacy, the view to the outside, as well as workspace size, esthetics, and the level of enclosure.
[10]	Multivariate logistic regression	The acceptability of the overall indoor environment was influenced by the acceptability of the thermal conditions, air quality, noise level, and illumination level.
[11]	Correspondence analysis and principal component analysis with ideal scaling	Workspace satisfaction was impacted by contentment with temperature, lighting conditions, air quality, acoustics, spatial aspects (including privacy and workspace individualization), office furniture, and office layout.
[12]	Literature review	The review of the existing literature highlights the dual advantages of a favorable indoor environmental quality (IEQ), encompassing both economic and health-related benefits. It underscores the substantial influence of the IEQ on occupant well-being and work efficiency.
[13]	Cross-sectional study design amongst objective measurements and subjective assessments	The study offered a positive suggestion for green buildings with qualitatively and quantitatively measured performance in terms of the IEQ.
[14]	Non-parametric techniques	Satisfaction with the overall environment was influenced by satisfaction with the thermal, acoustic, lighting, and air quality.
[15]	Non-parametric statistical tests	The degree of improvement was more pronounced when moving from a traditional building to one certified under the WELL standard, while it was less significant or even negligible when transitioning from buildings certified under BREEAM to WELL certification.
[16]	Literature review	The recovery of the indoor environment of a healthcare facility for both patients and, more significantly, medical staff.

The studies did not provide a consistent definition of occupant satisfaction. However, all of them approached occupant satisfaction broadly, linking it to either their satisfaction and well-being with the indoor environmental quality or their fulfillment and comfort with their workspace. Specifically, some studies [2,5,6,10] concentrated solely on how the indoor environmental quality influenced the satisfaction of building occupants.

Their findings indicated that the thermal, visual, and acoustic atmosphere and superior air quality influenced the building occupants' sense of fulfillment. While the significance of various indoor environmental factors in building occupants' satisfaction showed slight variations across studies, the thermal environment was consistently ranked as slightly more important than the air quality and the acoustic environment, and significantly more important than the visual environment.

A literature review conducted by [17] highlighted that various non-environmental factors significantly influence building occupants' satisfaction alongside indoor environmental

factors. Factors such as occupants' control over the indoor environment, view satisfaction, privacy levels, and office layout have been identified as crucial determinants ([3,4,8,18]). Moreover, recent advancements propose the application of machine learning techniques to enhance the examination of medical data, potentially surpassing traditional methods [19]. The complexity of human decision making has recently highlighted the need to comprehend risky options. One study, for example, used the choices 13k dataset to train neural networks in a unique way that revealed information about decision noise and dataset bias [20]. For various other applications, one should refer to ([16,21–26]). Further, [27] promote the development of the agent decision model and provide a new way to solve complex decision problems.

This research aims to present a latent class choice model with participant environmental feedback data in an authentic setting. In order to overcome several issues with the conventional approaches, we developed a novel hybrid latent class choice model in this study that combines a Gaussian negative binomial mixture model [28]. To evaluate the performance of the suggested model with more conventional models, we employed micro-ecological momentary assessments (EMA) as secondary feedback data in this investigation.

Micro-EMA is a technique that utilizes a smartwatch interface to elicit and gather immediate, in-the-moment subjective feedback from an individual over several weeks [16]. We gained insight into that person's comfort preference patterns by obtaining an extensive volume of feedback from a single individual in various environments and comfort conditions. It is suggested that these behavioral patterns can be employed to categorize individuals into clusters based on their environmental perceptions. Consequently, grouping individuals with comparable comfort preferences could enhance the precision of forecasting where a person will feel comfortable and how the system can respond without extra sensors. Moreover, accumulating substantial quantities of subjective preference data from numerous individuals in a specific area can help define the comfort-related characteristics of that space to complement the data obtained from the installed sensors.

If it is theoretically feasible and minimally disruptive to the occupants, incorporating humans as sensors within buildings can revolutionize post-occupancy evaluations and building and system design, as well as controls and automation procedures. This opens up opportunities for individuals to contribute feedback in various scenarios, whether for short-term episodic purposes (spanning days or weeks), building commissioning extending to long-term assessments (over months or years), or continuous system control and management. This research aligns with the growing interest in other fields that leverage human input as sensors, such as event detection using cybersecurity [18], social media data [24], ecological momentary assessment [29] and emergency detection [30].

In the context of environmental measurements, previous efforts have involved the use of specific sensors mounted on mobile carts [31,32]. However, these sensors were not cost-effective for many building operation scenarios and the affordability issue led to the development of low-cost continuous-sensing sensors, albeit with the requirement for frequent calibration [1,33,34]. Nonetheless, the placement of these sensors within buildings and the interpretation of their readings remained challenging in the literature, primarily due to the heterogeneous nature of indoor spaces [34]. Conversely, surveys introduced their own set of challenges, such as determining the appropriate questions to ask, selecting the right respondents, and interpreting the survey results [1]. Furthermore, [35] explored the concept of 'survey fatigue', wherein survey participants become overwhelmed by the volume of questions, potentially resulting in misrepresentations in responses and decreased response rates.

The literature on choice modeling in the context of environmental perspectives is estimated to have exceeded 14,200 publications. Some of the articles related to occupant preferences and satisfaction with their findings are listed below.

The attention to discrete choice models as a hypothetically sound and practical tool for investigating choice behavior, particularly behavioral outcomes like willingness to pay, has grown rapidly. This development initially took place in the context of transportation, which

was where McFadden made his initial contributions. The authors of [1] offer a historical indication of these contributions from the 1970s to the early 1990s. Since then, there has been a significant expansion of research in the various aspects of choice modeling. This includes the development of more behaviorally realistic discrete choice model forms, such as closed-form models like nested logit, and open-form models like mixed logit, latent class, and generalized mixed logit. New data paradigms have also emerged, including mixed data approaches and expressed preference and choice studies. Additionally, process heuristics have been incorporated by researchers into choice models, such as hybrid logit models, to handle attribute endogeneity and account for attribute non-attendance [36].

## 2. Model Framework

The subject model, namely the Gaussian negative binomial mixture with a latent class choice model, is presented in this section. Then, we give an extensive comparison with benchmark models, i.e., mixed logit, Multinomial Logit, and latent class choice models. We observe that our subject model efficiently performs better than the benchmark models. By including negative binomial distribution, the subject model effectively addresses overdispersion. Additionally, the presence of a latent class choice model makes it more reliable for decision making under heterogeneity. The subject model performed better under the circumstances of heterogeneity in classes and data variability.

### 2.1. Latent Class Choice Model

LCCM contains two models: a class membership model and a class-specific choice model. The class membership model is defined as a function of the features of decision-makers associated with a particular class. The utility  $\omega$  of a decision-maker ' $m$ ' associated with class ' $l$ ' is stated as follows:

$$\omega_{ml} = C'_m \varphi_l + v_{ml} \quad (1)$$

where  $C_m$  is a vector of the features of decision-makers ' $m$ ' and  $\varphi_l$  is the corresponding vector of unknown parameters.  $v_{ml}$  is an error term that follows Extreme Value Type-I distribution over decision-makers and classes, which is assumed to be i.i.d.

The probability of decision-makers ' $m$ ' associated with class ' $l$ ' is specified decision-makers as follows:

$$P(r_{ml} = 1 | C_m, \varphi_l) = \frac{\exp[C'_m \varphi_l]}{\sum_{l'=1}^L \exp[C'_m \varphi_{l'}]} \quad (2)$$

$$r_{ml} = \begin{cases} 1, & \text{if decision-makers associated with class } l \\ 0, & \text{otherwise.} \end{cases}$$

The second model, namely, the class-specific model, is defined as the probability of selecting a particular option as a function of the observed exogenous feature option, conditioned on the person associated with class ' $l$ '. The utility of an individual ' $m$ ' selecting an option ' $k$ ' at a time  $\tau$  is expressed as follows:

$$U_{mk\tau|l} = E'_{mk\tau} \delta_l + v_{mk\tau|l} \quad (3)$$

where  $E_{mk\tau}$  is a vector of the observed features of selecting an option ' $k$ ' at the time  $\tau$ ,  $\delta_l$  is the corresponding unknown parameter's vector, and  $v_{mk\tau|l}$  is an error term that follows Extreme Value Type-I distribution. At the same time, it is assumed to be i.i.d.; the conditional probability of decision-makers ' $m$ ' selecting an option ' $k$ ' at a time  $\tau$  is given as follows:

$$P(t_{mk\tau} = 1 | E_{mk\tau}, r_{ml} = 1, \delta_l) = \frac{\exp[E_{mk\tau|l}]}{\sum_{k'=1}^K \exp[E_{mk'\tau|l}]} \quad (4)$$

where  $K$  is the number of available options.

Let  $t_m$  be a matrix of all the individual options at time  $\tau$ , consisting of  $(k \times \tau_m)$  order and  $E_m$  be  $(k \times \tau_m)$  order matrix, where

$$t_{mk\tau} = \begin{cases} 1, & \text{if decision - makers 'm' selecting an option 'k' at time } \tau \\ 0, & \text{otherwise.} \end{cases}$$

The conditional probability of observing  $t_m$  associated with class ' $l$ ' is expressed as follows:

$$P(t_m | E_m, r_{ml} = 1, \delta_l) = \prod_{\tau_1=1}^{\tau_n} \prod_{k=1}^K [P(t_{mk\tau} = 1 | E_{mk\tau}, r_{ml} = 1, \delta_l)]^{t_{mk\tau}} \quad (5)$$

The likelihood of an individual ' $m$ ' selecting an option ' $k$ ' can be defined by combining the conditional option probability with the probability of an individual associated with class ' $l$ ' as follows:

$$P(t_m) = \sum_{l=1}^L P(r_{ml} = 1 | C_m, \varphi_l) P(t_m | E_m, r_{ml} = 1, \delta_l) \quad (6)$$

The resulting likelihood of all the decision-makers ' $m$ ' can be obtained as follows:

$$P(t) = \prod_{m=1}^M \sum_{l=1}^L P(r_{ml} = 1 | C_m, \varphi_l) P(t_m | E_m, r_{ml} = 1, \delta_l) \quad (7)$$

## 2.2. Proposed Model

The proposed model can be obtained by replacing the class membership probability with the Gaussian negative binomial mixture model. The subject model is a hybrid machine learning approach, as it combines the advantages of two types of models (i.e., discrete and continuous) into one. For this purpose, GMM and NBMM are used for continuous and discrete variables. The vector, consisting of ' $m$ ' decision-makers attributes, is split into two sub-vectors, i.e.,  $C_{cm}$  and  $C_{dm}$ . These two sub-vectors contain the dimensions  $\eta_c$  and  $\eta_d$  which equals the number of elements in  $C_{cm}$  and  $C_{dm}$  (continuous or discrete attributes) respectively.

### Gaussian Negative Binomial Mixture Model

GMM  $\mathbb{N}(C_{cm} | \lambda_{cm}, \Pi_{cm})$  is a collection of ' $L$ ' Gaussian densities, where each density is a segment of the mixture and has its mean  $\mu_{cm}$  and covariance  $\Pi_{cm}$  matrix. The overall likelihood that it represents the mixing coefficient  $\Lambda_l$  comes from the component ' $l$ '.

A useful and reliable distribution to incorporate count data is the negative binomial distribution. It is a versatile statistical tool that has gained popular significance for dealing with count data with overdispersion. The presence of multiple latent classes within the data allows the model to assume observation counts that are generated from a mixture of negative binomial models. This model can effectively capture the heterogeneity and excess variation within the data through the parameter estimation of the negative binomial mixture model and mixing proportion.

Marginal and posterior probability is estimated using Bayes' theorem after estimating subject model probability assuming Gaussian and negative binomial distributions are independent on continuous and discrete datasets, after considering the conditional independence properties on the graphical structure of the proposed model. The joint probability can be formulated by taking the product of four terms. The first term is class/label probability, and the second and third are the conditional probabilities of  $C_{cm}$  and  $C_{dm}$ . The

fourth term contains the choice probability conditional on the class. We can represent the joint probability as follows:

$$P(C_{cm}, C_{dm}, t_m, r_{ml} = 1 | E_m, \delta_l, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}) = P(r_{ml} = 1 | \Lambda_l) P(C_{cm}, r_{ml} = 1 | \lambda_{cl}, \Pi_{cl}) \cdot P(C_{dm}, r_{ml} = 1 | \lambda_{dl}) P(t_m | E_m, r_{ml} = 1, \delta_l) \quad (8)$$

where

$$P(r_{ml} = 1 | \Lambda_l) = \Lambda \quad (9)$$

$$\sum_{l=1}^L \Lambda_l = 1$$

$$P(C_{cm}, r_{ml} = 1 | \lambda_{cl}, \Pi_{cl}) = \mathbb{N}(C_{cm} | \lambda_{cm}, \Pi_{cm}) = \frac{1}{\sqrt{(2\pi)^{\eta_c} |\Pi_{cl}|}} \exp \left[ -\frac{1}{2} (C_{cm} - \lambda_{cl}) \Pi_{cm}^{-1} (C_{cm} - \lambda_{cl}) \right] \quad (10)$$

$$P \left( C_{dm}^{R_{dm_i}}, r_{ml} = 1 \mid \lambda_{dl} \right) = \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \quad (11)$$

where  $R_{dm_i} = r^{th}$  vectors of favorable features and  $C_{dm_i}$  is the discrete characteristics of decision-makers 'm',  $R_{dm_i}$  is the number of  $\lambda_{cl}$  and  $\lambda_{dl}$  which are the corresponding mean vectors of continuous and discrete distributions.

### 2.3. Joint Probability

The joint probability of  $C_{cm}$ ,  $C_{dm}$ , and  $t_m$  can be accessed by using Equation (8)'s overall component 'k':

$$P(C_{cm}, C_{dm}, t_m | E_m, \delta, \Lambda, \lambda_c, \Pi_c, \lambda_d) = \sum_{l=1}^L P(C_{cm}, C_{dm}, t_m, r_{ml} = 1 | E_m, \delta_l, \Lambda_l, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}) \quad (12)$$

where  $\lambda_c$  and  $\lambda_d$  are matrices containing the 'L' mean vectors of continuous and discrete variables,  $\Pi_c$  is a matrix containing 'L' covariance matrices  $\Pi_{cl}$ , and  $\delta$  is a matrix containing the L vectors of  $\delta_l$ . By omitting the dependencies on the left-hand side of the equation to make the notation more assembled,

$$P(C_c, C_d, t) = \prod_{m=1}^M P(C_{cm}, C_{dm}, t_m | E_m, \delta, \Lambda, \lambda_c, \Pi_c, \lambda_d) \quad (13)$$

$$P(C_c, C_d, t) = \prod_{m=1}^M \sum_{l=1}^L \Lambda_l \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \cdot \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K \left( \frac{e^{E_{mk\tau} \delta_k}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_k}} \right)^{t_{mk\tau}} \quad (14)$$

The overall joint probability can be estimated by using different methods (i.e., maximum likelihood estimation, Hessian Matrix, and Expectation Maximization Algorithm). The traditional maximum likelihood estimation method is inscrutable due to the summation over 'L' that will appear inside the equation on both LCCM and GNBM-LCCM. However, as the number of parameters increases in the model, the MLE becomes more burdensome and lengthened. In addition, the empirical singularity problems might arise during the Hessian Matrix procedures and become numerically challenging, [37]. Therefore, the EM algorithm is an effective way to overcome all these problems. Moreover, it is a powerful technique to estimate the parameters with latent variables.

### 2.4. EM Algorithm

The EM algorithm is divided into two steps, expectation and maximization steps, respectively, as follows:

E-step: This step starts first by taking the joint likelihood function as follows:

$$P(C_c, C_d, t, r) = \prod_{m=1}^M \prod_{l=1}^L \left[ \Lambda_l \mathbb{N}(C_{cm} | M_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \right]^{r_{ml}} \prod_{m=1}^M \prod_{l=1}^L \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K \left( \frac{e^{E'_{mk\tau} \delta_k}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_k}} \right)^{t_{mk\tau}} \quad (15)$$

Then, taking the logarithm of the likelihood, the probability breaks the function into two separate terms, i.e., the class membership model and the class-specific choice model.

$$LL = \sum_{m=1}^M \sum_{l=1}^L r_{ml} \log \left[ \Lambda_l \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \right] + \sum_{m=1}^M \sum_{l=1}^L \sum_{\tau=1}^{\tau_m} \sum_{k=1}^K t_{mk\tau} r_{ml} \log \left( \frac{e^{E'_{mk\tau} \delta_k}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_k}} \right) \quad (16)$$

Now, we find the value of  $r_{ml}$  by taking the expectation using the Bayes theorem.

$$P(r_{ml} = 1 | t_m, C_{cm}, C_{dm}, E_m, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}, \Lambda_l, \delta_l) \propto P(r_{ml} = 1 | \Lambda_l) P(C_{cm} | r_{ml} = 1, \lambda_{cl}, \Pi_{cl}) \cdot P(C_{dm} | r_{ml} = 1, \lambda_{dl}) \cdot P(t_m | E_m, r_{ml} = 1, \delta_l) \quad (17)$$

$$P(r_{ml} = 1 | t_m, C_{cm}, C_{dm}, E_m, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}, \Lambda_l, \delta_l) \propto \Lambda_l \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \cdot \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K \left( \frac{e^{E'_{mk\tau} \delta_k}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_k}} \right)^{t_{mk\tau}} \quad (18)$$

$$E[r_{ml}] = \gamma_{r_{ml}} = \frac{\Lambda_l \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K \left( \frac{e^{E'_{mk\tau} \delta_l}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_l}} \right)^{t_{mk\tau}}}{\sum_{l'=1}^L \Lambda_{l'} \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) \prod_{i=1}^{\eta_d} \lambda_{dl_i}^{R_{dm_i}} (1 - \lambda_{dl_i})^{C_{dm_i}} \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K \left( \frac{e^{E'_{mk\tau} \delta_k}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_k}} \right)^{t_{mk\tau}}} \quad (19)$$

It is important to note that  $\Lambda_l$  in Equation (9) and  $\gamma_{r_{ml}}$  in Equation (16) contemplate prior probability and corresponding posterior probabilities, respectively.

M-step: In this step, the unknown parameters are estimated, since, in the presence of latent variable  $r_{ml}$ , Equation (16) cannot be estimated directly. Making use of Equations (16) and (19), it gives the following:

$$LL = \sum_{m=1}^M \sum_{l=1}^L r_{ml} \left[ \log \Lambda_l + \log \mathbb{N}(C_{cm} | \lambda_{cl}, \Pi_{cl}) + \sum_{i=1}^{\eta_d} \{ R_{dm_i} \log \lambda_{dl_i} + C_{dm_i} \log (1 - \lambda_{dl_i}) \} \right] + \sum_{m=1}^M \sum_{l=1}^L \sum_{\tau=1}^{\tau_m} \sum_{k=1}^K t_{mk\tau} \gamma_{r_{ml}} \log \left( \frac{e^{E'_{mk\tau} \delta_l}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_l}} \right) \quad (20)$$

By taking the derivatives of unknown parameters and setting them to zero, we obtained the following:

$$\lambda_{cl} = \frac{1}{M_l} \sum_{m=1}^M \gamma_{r_{ml}} C_{cm} \quad (21)$$

$$\Pi_{cl} = \frac{1}{M_l} \sum_{m=1}^M \gamma_{r_{ml}} (C_{cm} - \lambda_{cl})(C_{cm} - \lambda_{cl})' \quad (22)$$

$$\lambda_{dl} = \frac{M_l R_{dml}}{\sum_{l=1}^L C_{dl} + M_l R_{dml}} \quad (23)$$

$$\Lambda_l = \frac{M_l}{M} \quad (24)$$

where  $M_l = \sum_{m=1}^M \gamma_{rml}$

$$\delta_l = \operatorname{argmax} \sum_{m=1}^M \sum_{\tau=1}^{\tau_m} \sum_{k=1}^K t_{mk\tau} \gamma_{rml} \log \left( \frac{e^{E'_{mk\tau} \delta_L}}{\sum_{k'=1}^K e^{E'_{mk'\tau} \delta_L}} \right) \quad (25)$$

Overall, the EM algorithm revolves between E-step and M-step until convergence is attained. First, the unknown parameters are estimated. Second, the latent variable (Equation (19)) is estimated by taking the expectation using the Bayes Theorem. In addition, the closed-form solution of the parameters is derived (from Equation (21) to Equation (24)). Finally, the log-likelihood is examined by utilizing the obtained values of the unknown parameters and then scrutinized for convergence. If the convergence benchmark is not reached, we return to E-step. From Equation (21) to Equation (24), the closed-form solution is available for maximizing coefficient, the Gaussian mean matrix, the negative binomial mean matrix, and Gaussian covariance matrix, respectively. Regarding Equation (25), we cannot obtain any closed-form solution for the parameter  $\delta_l$ . For this purpose, the Gradient-Based Numerical Optimization method is used.

### 2.5. Final Likelihood

After attaining convergence, the marginal probability of observing a vector of ' $t$ ' options of all the decision-makers ' $M$ ' is examined as follows:

$$P(t) = \prod_{m=1}^M \sum_{l=1}^L P(r_{ml} = 1 | C_{cm}, C_{dm}, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}, \Lambda_l) \prod_{\tau=1}^{\tau_m} \prod_{k=1}^K [P(t_{mk\tau} = 1 | E_m, r_{ml} = 1, \delta_l)]^{t_{mk\tau}} \quad (26)$$

where  $P(r_{ml} = 1 | C_{cm}, C_{dm}, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}, \Lambda_l)$  is the posterior probability of the vector  $C_m = \{C_{cm}, C_{dm}\}$  being obtained by the cluster ' $l$ '.

The posterior probability can be expressed using Bayes theorem as follows:

$$P(r_{ml} = 1 | C_{cm}, C_{dm}, \lambda_{cl}, \Pi_{cl}, \lambda_{dl}, \Lambda_l) = \frac{P(r_{ml} = 1 | \Lambda_l) (C_{cm} | r_{ml} = 1, \lambda_{cl}, \Pi_{cl}) P(C_{dm} | r_{ml} = 1, \lambda_{dl})}{\sum_{l'=1}^L P(r_{ml'} = 1 | \Lambda_{l'}) (C_{cm} | r_{ml'} = 1, \lambda_{cl}, \Pi_{cl}) P(C_{dm} | r_{ml'} = 1, \lambda_{dl})} \quad (27)$$

The above posterior probability of Equation (27) can be used to compare the GNBM-LCCM and traditional LCCM from Equation (7). Further, it is used to compute extrapolated sample prediction accuracy.

### 2.6. Real-Life Application with Discussion

**Dataset Overview:** This dataset comprises data collected from the BUDS lab deployments of the Cozie Fitbit smartwatch platform. It involves collecting intensive longitudinal subjective feedback regarding comfort-based preferences through micro-ecological momentary assessments on a smartwatch platform. In an experiment conducted over two weeks with 30 occupants, a total of 4378 field-based surveys were generated to assess thermal, noise, and acoustic preferences.

Throughout the entire study, the environmental variables (such as temperature and relative humidity) in three different buildings were observed. The participants used an open-source application called Cozie on their smartwatches to complete comfort surveys.



Additionally, a custom-designed smartphone application constantly tracked their indoor locations. This location data allowed us to accurately synchronize the timing and spatial aspects of environmental measurements with the thermal preference responses provided by the participants.

In order to extract valuable insights from the dataset, we initiated the exploration by carefully reviewing the features and their corresponding descriptions (refer to Table 2). This preliminary examination served as an informative starting point, granting us a holistic understanding of the dataset's contents. It facilitated the identification of essential variables that play a pivotal role in shaping user choice behavior. These key variables were subsequently selected for more in-depth analysis and model development.

**Table 2.** List of features and their descriptions in the initial dataset.

Feature Name	Type	Description and Values
Index	Numeric	Unique numeric value assigned to occupants
Clothing	Count Data	Clothing value of occupants
Comfort cozie	Count Data	Indoor comfort level of occupants
Heart rate cozie	Count Data	Heart rate of occupants collected from the Fitbit smartwatch device
Lat. Cozie	Continuous	The floor latitude is identified by its grid cell
Light Cozie	Count Data	Environmental comfort by lightning
Lon Cozie	Continuous	The floor longitude is identified by its grid cell
Noise Cozie	Count Data	Noise level in the environment
Response speed Cozie	Continuous	The response speed of occupants collected from the Fitbit smartwatch device
Thermal Cozie	Count Data	The satisfaction or contentment of individuals with the thermal environment
Room	Count Data	Room temperature profile and occupied zone gradient
Co2 sensing	Discrete	Oxygen gradient for assessing comfort
Humidity sensing	Continuous	Relative humidity for assessing comfort
Light sensing	Discrete	Light gradient for assessing comfort
Noise sensing	Discrete	Noise gradient for assessing comfort
Temperature sensing	Continuous	Temperature gradient in the occupied zone for assessing comfort
Voc sensing	Discrete	Velocity in the occupied zone for assessing comfort
Temperature. Mambient	Continuous	Radiant temperature for assessing comfort

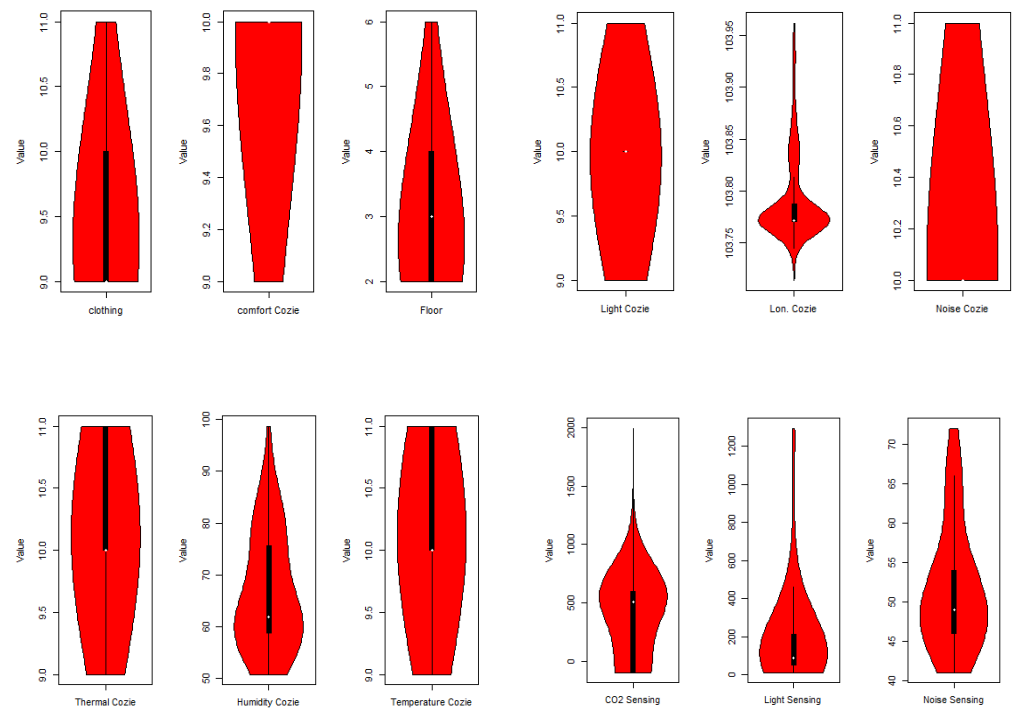
Table 3 presents the mean matrix illustrating the class membership model of the subject data. This matrix offers valuable insights into the distribution of the occupants among the different latent classes within the dataset. Through an examination of this mean matrix, we can gain an understanding of the likelihood of users belonging to each class and identify the underlying class structure within our hybrid model. Figure 1 graphically represents the majority of the data attributes in the case where the distribution of Environmental Light values already exhibits a noticeable overlap for various visual feedback (located at the top-middle distribution in Figure 2).

The variable 'time' was generated by employing feature engineering techniques on the timestamps corresponding to when the occupants provided feedback. This engineered feature represents the time cyclically, taking into account both the hour of the day and the day of the week. This straightforward feature type was integrated into all the scenarios to identify potential cyclical patterns or factors influencing preference prediction. The attribute 'time' is used to categorize the class membership model into distinct classes, as follows: Class L = 1: This initial latent class corresponds to environmental data recorded from September 28th to October 10th. These applications represent a specific group characterized by an early time frame. Class L = 2: The second class is associated with applications recorded from October 11th to October 22nd. Class L = 3: The third class is related to periods spanning from October 23rd to November 3rd. Class L = 4: The fourth and final class comprises applications with a time frame from November 4th to November 15th, representing the last observation period for the occupants. These class assignments help delineate the temporal structure of the dataset and provide a meaningful segmentation of the occupant observations. This finding strengthens the argument that relying solely on

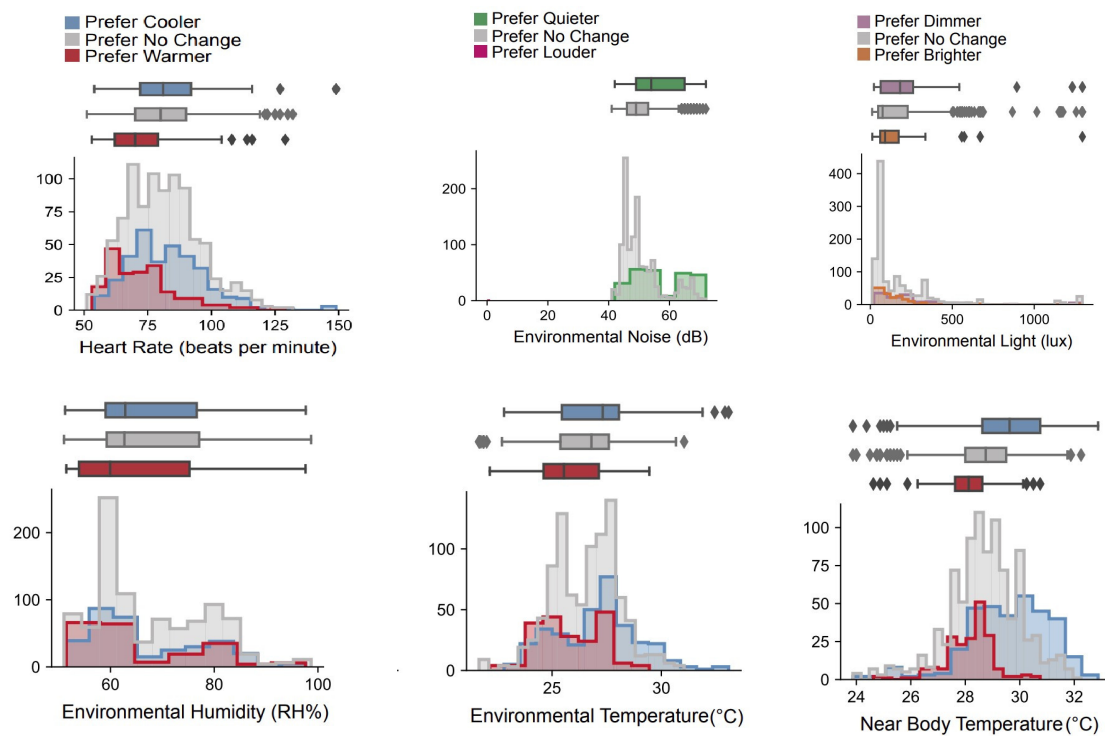
environmental measurements is insufficient for characterizing an individual’s preferences, thus leading to less accurate predictions, as has been observed in earlier research [7].

**Table 3.** Mean matrix of class membership model.

Attribute	Range	Class 1	Class 2	Class 3	Class 4
Clothing	9	0.0126	0.0008	0.0013	0.0026
	10	0.0131	0.0024	0.0040	0.0021
	11	0.0160	0.0019	0.0023	0.0014
Comfort	9	0.0265	0.0050	0.0056	0.0026
	10	0.0606	0.0107	0.0163	0.0111
Latitude Cozie	1.00–1.17	0.3882	0.1025	0.1032	0.0317
	1.18–1.35	0.0194	0.0008	0.0053	0.0036
	1.36–1.50	0.0443	0.0007	0.0095	0.0054
Light Cozie	9	0.5490	0.1274	0.1480	0.0782
	10	0.0475	0.0071	0.0086	0.0046
	11	0.0136	0.0014	0.0024	0.0021
Noise Cozie	9	0.0025	0.0000	0.0002	0.0005
	10	0.0008	0.0000	0.0000	0.0001
Floor	2	0.5211	0.1187	0.1304	0.0581
	3	0.0115	0.0015	0.0036	0.0015
	4	0.0058	0.0005	0.0021	0.0006
	5	0.0025	0.0002	0.0010	0.0003
	6	0.0015	0.0008	0.0006	0.0001
Room	1–4	0.0194	0.0008	0.0053	0.0036
	5–8	0.0443	0.0007	0.0095	0.0054
	9–12	0.1497	0.0410	0.0414	0.0137
	13–16	0.2775	0.0690	0.0706	0.0231



**Figure 1.** Visual illustration of some of the attributes of environmental data.



**Figure 2.** Visual illustrations of dispersion of sensor data based on preference voting.

Table 4 provides an overview of the parameter estimation in the class-specific choice model. This table offers insights into the estimated parameters for each latent class within the hybrid model. These estimated parameters enable us to quantitatively assess the influence of different variables on the choice behavior of users within each class. The proposed model facilitated the division and categorization of these zones, now based on the various comfort preferences exhibited by the occupants in those areas. This outcome primarily offered facility managers an overview of the office spaces they oversee, equipping them with insights to enhance comfort and take necessary actions.

**Table 4.** Parameter estimation of the class-specific choice model (GNBM).

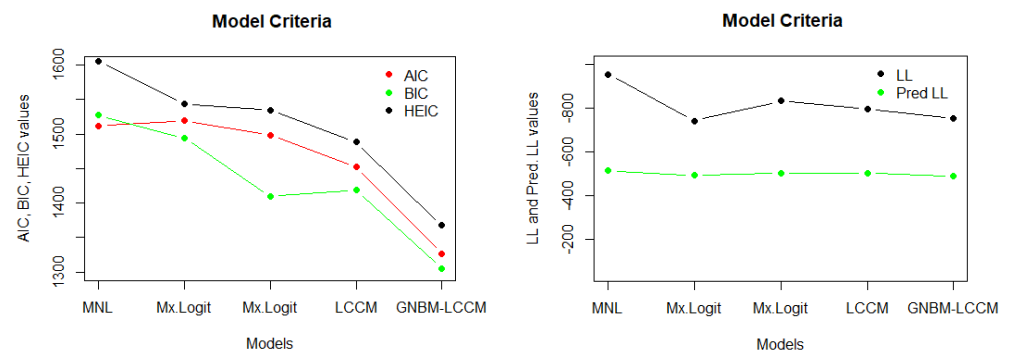
Parameters	Class 1	Class 2	Class 3	Class 4
Heart Rate	−1.3240 (0.018)	0.0061 (0.020)	−0.0134 (0.073)	−0.0999 (0.009)
Response	−1.8896 (0.010)	0.1021 (0.005)	0.0004 (0.000)	−0.2637 (0.006)
Thermal	0.1989 (0.004)	0.0342 (0.016)	0.0261 (0.039)	0.01763 (0.018)
CO <sub>2</sub> Sensing	0.0175 (0.030)	0.0014 (0.036)	0.0163 (0.000)	0.0721 (0.036)
Humidity	0.0013 (0.028)	0.0735 (0.061)	0.0536 (0.047)	0.0165 (0.080)
Light Sensing	0.0092 (0.037)	0.0938 (0.083)	0.0728 (0.023)	0.0828 (0.009)
Noise Sensing	0.0037 (0.008)	0.0015 (0.019)	0.0183 (0.025)	0.0194 (0.028)
Voc Sensing	0.0083 (0.019)	0.0635 (0.051)	0.0387 (0.019)	0.0295 (0.019)
Temperature	0.0163 (0.060)	0.0927 (0.019)	0.0624 (0.014)	0.0576 (0.016)

In Table 5, we compare the proposed Gaussian negative binomial mixture with a latent class choice model (GNBM-LCCM) and traditional models. Our evaluation involves the use of various metrics, including AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), HEIC (Hannan–Quinn Information Criterion), LL (log-likelihood), Joint LL (Joint Log-Likelihood), and Pred LL (Predictive Log-Likelihood). This comparative assessment allows us to determine the superiority of our GNBM-LCCM over traditional models in terms of model fit, complexity, and predictive accuracy. The visual representations of these benchmark comparisons are also provided in Figures 3–5.

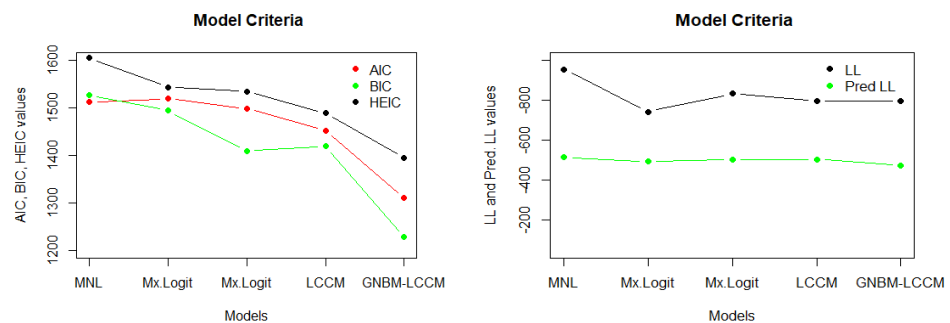
**Table 5.** Model comparison of proposed model with benchmark models using different criteria.

Models	Specifications	LL <sup>1</sup>	Joint LL <sup>2</sup>	Pred. LL	Residual Deviance	AIC	BIC	HEIC
Multinomial Logistic	Normal	−954.81		−514.62	1594.47	1512.17	1527.45	1605.33
Mixed Logit		−743.23		−494.76	1598.38	1519.23	1494.08	1543.65
LCCM		−798.06	−8615.23	−504.88	1499.56	1452.00	1418.89	1489.00
GNBM-LCCM	Diagonal covariance	−755.77	−7856.14	−489.63	1416.00	1326.41	1304.93	1367.41
	Spherical covariance	−797.54	−7904.37	−474.19	1401.98	1310.01	1227.82	1394.56
	Full covariance	−734.21	−7866.56	−415.35	1429.67	1276.20	1224.54	1384.03

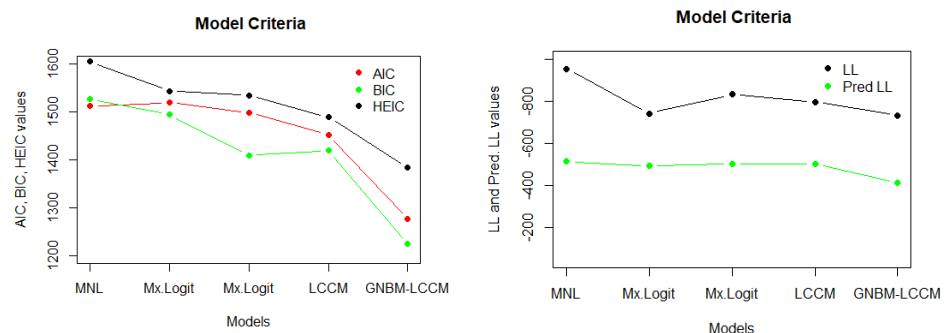
<sup>1</sup> Marginal log-likelihood of GNBM-LCCM (Equation (26)) and LCCM (Equation (7)); <sup>2</sup> Joint log-likelihood of GNBM-LCCM (Equation (14)).



**Figure 3.** Visual illustration of traditional models with proposed model (diagonal covariance) using different criteria.



**Figure 4.** Visual illustration of traditional models with proposed model (spherical covariance) using different criteria.

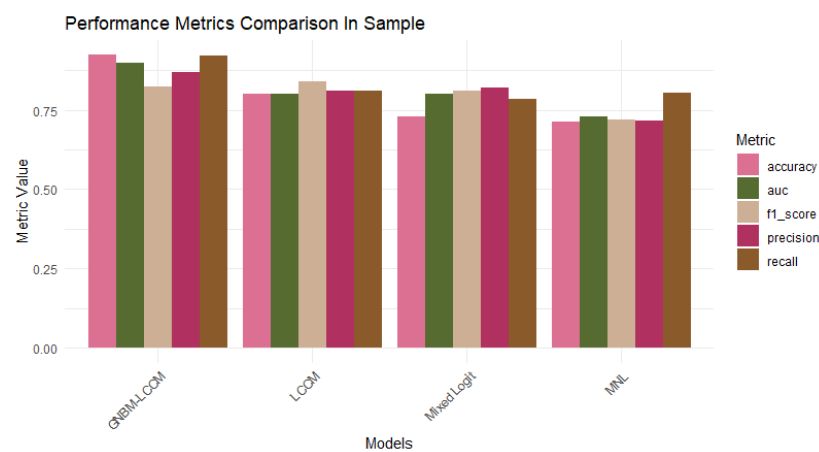


**Figure 5.** Visual illustration of traditional models with proposed model (full covariance) using different criteria.

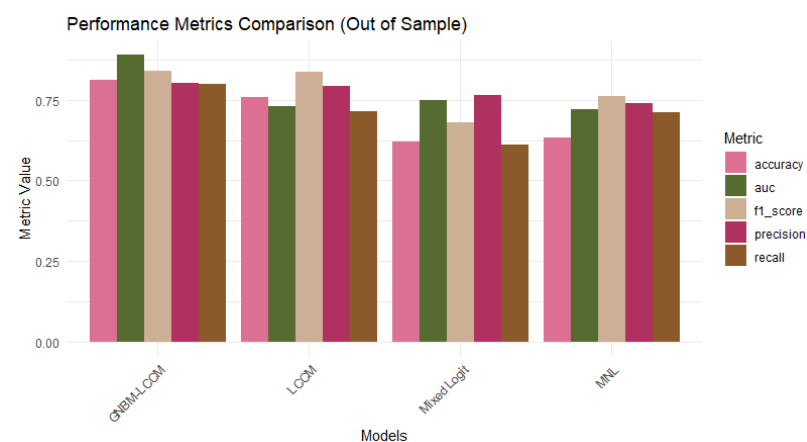
Considering the preference feedback in this methodology occurred at a notably higher frequency compared to typical surveys or occupants’ interactions with thermostats, this study possessed preference data characterized by a relatively diverse temporal and spatial

nature. In the initial observation, it is apparent that the office space generally provided a comfortable environment, whereas outdoor seating areas exhibited an overall higher preference for cooling. The study also captured time-dependent fluctuations, revealing the model's capability to predict comfort preferences that varied across different times of the day or days of the week. Notably, within the office environment, there was a peak in warmer preference around mid-day. However, it is worth noting that the model sometimes attempted to predict comfort preferences inaccurately during periods when no data were available. The square peaks observed in the office area for aural and visual prediction, particularly between the hours of 22:00 and 7:00, were a result of the absence of data to make accurate predictions during those times.

Figures 6 and 7 provide a comprehensive comparison of the performance metrics for four models: MNL, mixed logit, LCCM, and GNBM-LCCM, evaluated both in-sample and out-of-sample. Table 6 illustrates the in-sample evaluation criteria, where the GNBM-LCCM consistently outperforms the benchmark models across most metrics. Specifically, the GNBM-LCCM demonstrates superior accuracy (0.9246), Recall (0.9204), precision (0.8693), and AUC (0.8971), with only a slight dip in the F1 Score (0.8249) compared to LCCM (0.8396). These results indicate that the GNBM-LCCM not only enhances the predictive accuracy but also effectively captures the intricacies of decision-making processes within the sample data.



**Figure 6.** Visual illustration of traditional models with proposed model using different evaluation criteria (in-sample).



**Figure 7.** Visual illustration of traditional models with proposed model using different evaluation criteria.

**Table 6.** Evaluation criterion of benchmark model and proposed model (in-sample).

Models	MNL	Mixed Logit	LCCM	GNBM-LCCM
Accuracy	0.7118	0.7298	0.8016	<b>0.9246</b>
Recall	0.8034	0.7839	0.8119	<b>0.9204</b>
F1 Score	0.7185	0.8102	0.8396	<b>0.8249</b>
Precision	0.7155	0.8193	0.8110	<b>0.8693</b>
AUC	0.7287	0.8004	0.7998	<b>0.8971</b>

Table 7 extends this evaluation to out-of-sample performance, where the GNBM-LCCM maintains its dominance over the traditional models. The GNBM-LCCM achieves the highest accuracy (0.8103), AUC (0.8901), competitive Recall (0.7983), F1 Score (0.8398), and precision (0.8024). These metrics suggest that the GNBM-LCCM is robust and generalizes well to unseen data, thus providing reliable predictions beyond the training dataset. In contrast, the mixed logit model shows the weakest performance out-of-sample, with lower accuracy (0.6197) and Recall (0.6115), highlighting its limitations in generalizability. The reliable performance of both the in-sample and out-sample evaluation criteria of the subject model makes it a more robust and accurate choice model.

**Table 7.** Evaluation criterion of benchmark model and proposed model (out-of-sample).

Models	MNL	Mixed Logit	LCCM	GNBM-LCCM
Accuracy	0.6315	0.6197	0.7593	<b>0.8103</b>
Recall	0.7109	0.6115	0.7158	<b>0.7983</b>
F1 Score	0.7610	0.6789	0.8374	<b>0.8398</b>
Precision	0.7398	0.7630	0.7932	<b>0.8024</b>
AUC	0.7198	0.7481	0.7294	<b>0.8901</b>

### 3. Conclusions

In this study, an innovative hybrid choice model has been introduced, namely the Gaussian negative binomial mixture with latent class choice model (GNBM-LCCM). Further, we have checked its practical application by implementing this to environmental preference data. Our primary objective was to make it more reliable as compared to benchmark studies, i.e., Multinomial Logit model, mixed logit model, and latent class choice model (LCCM). By this comparison, we have proved the superior performance of our subject model. The results demonstrate that the proposed model not only outperforms in-sample evaluation, but it also shows superior performance for out-of-sample criteria.

All of the previous studies effectively represent the heterogeneity in individual preferences, but they fail to deal with the overdispersion situation. The benchmark models such as the Multinomial Logit and latent class choice models could easily depict the decision-making process, but it is not accurate in complex criteria like the subject model. We fill the gap of the restricted comprehension of individual preferences and latent classes with overdispersion data by incorporating the hybrid model.

The decision-making process is being scrutinized by implementing GNBM-LCCM in the analysis of environmental preference data. These data are more extensively examined through GNBM with the presence of latent classes in it. These results highlight how important it is to take latent classes into account and correct for overdispersion in a choice model that not only increases its accuracy but also plays a vital role in capturing heterogeneity in individual preferences.

The subject model is a robust framework for analyzing the decision-making process; however, it faces some limitations. For example, it may give over-generalized results as the latent class follows an independent assumption. Secondly, the model performance

might oversimplify the input data, especially with sparse or noisy datasets. Additionally, in large-scale applications, parameter estimation may lead to false predictions, which require more consideration of computational complexity.

Therefore, future research could focus on the GNBM-LCCM for large dataset scalability and model performance. The applicability and robustness of the subject model could investigate different areas using different datasets. Finally, the advancement of the subject model could be enhanced by adding an alternative regularization technique with sparse or noisy data.

**Author Contributions:** Conceptualization, I.S.; methodology, I.S. and J.G.D. software, I.S.; validation, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D. formal analysis, I.S. and J.G.D. investigation, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D. resources, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D. data curation, I.S.; writing—original draft preparation, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D. writing—review and editing, I.S., I.A.N.; visualization, I.S.; supervision, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D. project administration, I.S., I.A.N., M.M.A.A., O.A.A., J.G.D.; funding acquisition, I.A.N., M.M.A.A., O.A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by the Deanship of Research and Graduate Studies at King Khalid University through a Large Research Project under grant number RGP2/41/45.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through a Large Research Project under grant number RGP2/41/45.

**Conflicts of Interest:** The authors declare no competing interests.

## Nomenclature

$\omega_{ml}$	The utility of decision-makers associated with class $l$ .
$E_{mk\tau}$	The vector of observed characteristics of individual $m$ .
$C_{cm}$	The continuous attributes of decision-maker $m$ .
$C_{dm}$	The discrete attributes of decision-maker $m$ .
$\eta_c$	The number of elements in $C_{cm}$ .
$\eta_d$	The number of elements in $C_{dm}$ .
$\delta_l$	The unknown parameter of utility belonging to class $l$ at the time $\tau$ .
$\varphi_l$	The unknown parameter of utility belonging to class $l$ .
$\lambda_{cl}$ and $\lambda_{dl}$	The mean of the Gaussian and negative binomial mixture model.
$\Pi_{cl}$	The covariance of the Gaussian mixture model.
$\delta_l$	Mixing probability associated with class $c$ .

## References

1. Heinzerling, D.; Schiavon, S.; Webster, T.; Arens, E. Indoor environmental quality assessment models: A literature review and a proposed weighting and classification scheme. *Build. Environ.* **2013**, *70*, 210–222. [[CrossRef](#)]
2. Astolfi, A.; Pellerey, F. Subjective and objective assessment of acoustical and overall environmental quality in secondary school classrooms. *J. Acoust. Soc. Am.* **2008**, *123*, 163–173. [[CrossRef](#)] [[PubMed](#)]
3. Bluysen, P.M.; Aries, M.; van Dommelen, P. Comfort of workers in office buildings: The European HOPE project. *Build. Environ.* **2011**, *46*, 280–288. [[CrossRef](#)]
4. Choi, J.H.; Aziz, A.; Loftness, V. Decision support for improving occupant environmental satisfaction in office buildings: The relationship between subset of IEQ satisfaction and overall environmental satisfaction. In Proceedings of the Healthy Buildings, Syracuse, NY, USA, 13–17 September 2009.
5. Humphreys, M.A. Quantifying occupant comfort: Are combined indices of the indoor environment practicable? *Build. Res. Inf.* **2005**, *33*, 317–325. [[CrossRef](#)]
6. Lai, A.C.K.; Mui, K.W.; Wong, L.T.; Law, L.Y. An evaluation model for indoor environmental quality (IEQ) acceptance in residential buildings. *Energy Build.* **2009**, *41*, 930–936. [[CrossRef](#)]
7. Marans, R.W.; Yan, X. Lighting quality and environmental satisfaction in open and enclosed offices. *J. Archit. Plan. Res.* **1989**, *6*, 118–131.
8. Schledermann, K.M.; Bjørner, T.; West, A.S.; Hansen, T.S. Evaluation of staff's perception of a circadian lighting system implemented in a hospital. *Build. Environ.* **2023**, *242*, 110488. [[CrossRef](#)]



9. Veitch, J.A.; Charles, K.E.; Farley, K.M.J.; Newsham, G.R. A model of satisfaction with open-plan office conditions: COPE field findings. *J. Environ. Psychol.* **2007**, *27*, 177–189. [[CrossRef](#)]
10. Wong, L.T.; Mui, K.W.; Hui, P.S. A multivariate-logistic model for acceptance of indoor environmental quality (IEQ) in offices. *Build. Environ.* **2008**, *43*, 1–6. [[CrossRef](#)]
11. Schakib-Ekbatan, K.; Wagner, A.; Lussac, C. Occupant satisfaction as an indicator for the socio-cultural dimension of sustainable office buildings—Development of an overall building index. In Proceedings of the Conference: Adapting to Change: New Thinking on Comfort, Windsor, UK, 9–11 April 2010.
12. Al Horr, Y.; Arif, M.; Kaushik, A.; Mazroei, A.; Kafatygiotou, M.; Elsarrag, E. Occupant productivity and office indoor environment quality: A review of the literature. *Build. Environ.* **2016**, *105*, 369–389. [[CrossRef](#)]
13. Lee, J.Y.; Wargocki, P.; Chan, Y.H.; Chen, L.; Tham, K.W. Indoor environmental quality, occupant satisfaction, and acute building-related health symptoms in Green Mark-certified compared with non-certified office buildings. *Indoor Air* **2018**, *29*, 112–129. [[CrossRef](#)] [[PubMed](#)]
14. Altomonte, S.; Schiavon, S.; Kent, M.G.; Brager, G. Indoor environmental quality and occupant satisfaction in green-certified buildings. *Build. Res. Inf.* **2019**, *47*, 255–274. [[CrossRef](#)]
15. Licina, D.; Yildirim, S. Occupant satisfaction with indoor environmental quality, sick building syndrome (SBS) symptoms and self-reported productivity before and after relocation into WELL-certified office buildings. *Build. Environ.* **2021**, *204*, 108183. [[CrossRef](#)]
16. Shen, X.; Zhang, H.; Li, Y.; Qu, K.; Zhao, L.; Kong, G.; Jia, W. Building a satisfactory indoor environment for healthcare facility occupants: A literature review. *Build. Environ.* **2023**, *228*, 109861. [[CrossRef](#)]
17. Frontczak, M.; Wargocki, P. Literature survey on how different factors influence human comfort in indoor environments. *Build. Environ.* **2011**, *46*, 922–937. [[CrossRef](#)]
18. Vielberth, M.; Menges, F.; Pernul, G. Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity* **2019**, *2*, 23. [[CrossRef](#)]
19. Al-Quraishi, T.; NG, C.K.; Mahdi, O.A.; Gyasi, A.; Al-Quraishi, N. Advanced Ensemble Classifier Techniques for Predicting Tumor Viability in Osteosarcoma Histological Slide Images. *Appl. Data Sci. Anal.* **2024**, *2024*, 52–68. [[CrossRef](#)]
20. Thomas, T.; Straub, D.; Tatai, F.; Shene, M.; Tosik, T.; Kersting, K.; Rothkopf, C.A. Modelling dataset bias in machine-learned theories of economic decision-making. *Nat. Hum. Behav.* **2024**, *8*, 679–691. [[CrossRef](#)] [[PubMed](#)]
21. Alghazzawi, D.; Noor, A.; Alolaiyan, H.; Khalifa, H.A.E.W.; Alburaihan, A.; Xin, Q.; Razaq, A. A novel perspective on the selection of an effective approach to reduce road traffic accidents under Fermatean fuzzy settings. *PLoS ONE* **2024**, *19*, e0303139. [[CrossRef](#)]
22. Cheng, Y.; Deng, X.; Qi, Q.; Yan, X. Truthfulness of a Network Resource-Sharing Protocol. *Math. Oper. Res.* **2023**, *48*, 1522–1552. [[CrossRef](#)]
23. Shi, M.; Hu, W.; Li, M.; Zhang, J.; Song, X.; Sun, W. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. *Mech. Syst. Signal Process.* **2023**, *188*, 110022. [[CrossRef](#)]
24. Wang, D.; Amin, M.T.; Li, S.; Abdelzaher, T.; Kaplan, L.; Gu, S.; Pan, C.; Liu, H.; Aggarwal, C.C.; Ganti, R.; et al. Using humans as sensors: An estimation-theoretic perspective. In Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, Germany, 15–17 April 2014; pp. 35–46.
25. Wang, G.; Yang, J. SKICA: A Feature Extraction Algorithm Based on Supervised ICA with Kernel for Anomaly Detection. *J. Intell. Fuzzy Syst.* **2019**, *36*, 761–773. [[CrossRef](#)]
26. Wu, Z.; Liu, G.; Wu, J.; Tan, Y. Are neighbors alike? A semisupervised probabilistic collaborative learning model for online review spammers detection. *Inf. Syst. Res.* **2023**. [[CrossRef](#)]
27. Zhu, C. An Adaptive Agent Decision Model Based on Deep Reinforcement Learning and Autonomous Learning. *J. Logist. Inform. Serv. Sci.* **2023**, *10*, 107–118.
28. Stone, A.A.; Shiffman, S.; Atienza, A.A.; Nebeling, L.; Stone, A.; Shiffman, S.; Atienza, A.; Nebeling, L. Historical roots and rationale of ecological momentary assessment (EMA). In *The Science of Real-Time Data Capture: Self-Reports in Health Research*; Oxford University Press: Oxford, UK, 2007; pp. 3–10.
29. Intille, S.; Haynes, C.; Maniar, D.; Ponnada, A.; Manjourides, J. Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 1124–1128.
30. Avvenuti, M.; Cimino, M.G.C.A.; Cresci, S.; Marchetti, A.; Tesconi, M. A framework for detecting unfolding emergencies using humans as sensors. *Springerplus* **2016**, *5*, 43. [[CrossRef](#)] [[PubMed](#)]
31. Kim, Y.K.; Abdou, Y.; Abdou, A.; Altan, H. Indoor environmental quality assessment and occupant satisfaction: A post-occupancy evaluation of a UAE University Office Building. *Buildings* **2022**, *12*, 986. [[CrossRef](#)]
32. Webster, T.; Arens, E.; Anwar, G.; Bonnell, J.; Bauman, F.; Brown, C. *UFAD Commissioning Cart: Design Specifications and Operating Manual*; Internal Report; Center for the Built Environment, UC Berkeley: Berkeley, CA, USA, 2007.
33. Parkinson, T.; Parkinson, A.; de Dear, R. Continuous IEQ monitoring system: Context and development. *Build. Environ.* **2019**, *149*, 15–25. [[CrossRef](#)]
34. Jin, M.; Liu, S.; Schiavon, S.; Spanos, C. Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring. *Build. Environ.* **2018**, *127*, 268–276. [[CrossRef](#)]



35. Wang, G.; Yang, J.; Li, R. UFKLDA: An unsupervised feature extraction algorithm for anomaly detection under cloud environment. *Etri J.* **2019**, *41*, 684–695. [[CrossRef](#)]
36. Hess, S.; Daly, A. (Eds.) *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference*; Emerald Group Publishing Limited: Leeds, UK, 2010.
37. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.