

A bibliometric analytics study of hybrid choice modeling evolution and decision-making trends from 2010 to 2026

Péter Czine*, Sándor Kovács

Coordination and Research Centre for Social Sciences, Faculty of Economics and Business, University of Debrecen, Böszörményi Street 138, 4032-Debrecen, Hungary

ARTICLE INFO

Keywords:

Hybrid choice modeling
Latent variable models
Bibliometric analysis
Travel behavior
Research evolution
Decision modeling

ABSTRACT

Over the past 15 years, Hybrid choice modeling (HCM), also known as the integrated choice and latent variable (ICLV) approach, has become an increasingly important extension of discrete choice modeling. Despite the rapid growth in HCM publications, the structural and temporal evolution of this research field has not yet been systematically examined. This study presents a comprehensive bibliometric analysis of the HCM literature from 2010 to 2026. The primary novelty of this work lies in its analysis of a comprehensive multi-database corpus to trace the shift in focus from methodological development to application-oriented paradigms. These patterns are synthesized into an original four-tier conceptual framework. Using a PRISMA-guided and deliberately inclusive sample, the analysis examines publication trends, citation structures, leading journals and authors, and the evolution of thematic patterns and research priorities. The findings indicate sustained and accelerating growth, a concentration of scientific influence among a relatively small group of contributors, and a clear shift from methodological development to application-oriented research. While HCM is now widely institutionalized, model specification practices remain highly heterogeneous, and methodological reflection is often overshadowed by empirical application.

1. Introduction

The study and modeling of consumer preferences are central to multiple fields, including transportation economics, health economics, and marketing research. This is because it enables a systematic analysis of decision-making processes. Discrete choice modeling (DCM) provides a formal framework for analyzing individual decisions among a finite set of alternatives [1–3]. DCM is based on random utility theory (RUT), which posits that individuals assign latent utilities to available alternatives and select the option with the highest utility through random utility maximization [3–5]. Consequently, discrete choice models have been widely used to forecast and evaluate the influence of observed attributes on choice. The evolution of these models has been documented in extensive historical analyses. For example, Haghani et al. [6] track the scholarly landscape of discrete choice modeling over the past five decades and note the significant shift toward diverse applications in health, environmental, and transportation research. However, classical DCM has limitations when it comes to incorporating unobservable latent constructs, such as attitudes, norms, perceptions, and motivations, which can substantially affect decisions. The traditional framework emphasizes directly observed attributes and thus offers limited scope for the methodologically consistent integration of latent factors [7,8].

This limitation has motivated the development of hybrid choice modeling (HCM), also known as integrated choice latent variable (ICLV) modeling. HCM is an important extension of discrete choice modeling. Despite the growing use of HCM, however, this research field has not yet been systematically examined from a bibliometric perspective. No comprehensive review has mapped empirical HCM applications within a unified analytical framework over an extended period. No study has jointly examined the field's publication dynamics, citation structure, key contributors, geographical patterns, and thematic development.

From a methodological standpoint, HCM builds upon the discrete choice modeling framework using structural equation modeling (SEM) logic. This allows for the systematic integration of latent constructs into the decision-making process. HCM adds value at three interconnected levels:

- Incorporating latent variables directly into the utility specification
- Linking latent variables to observed indicators through measurement equations
- Explaining latent variables using observed covariates in structural equations [7,9,10]

* Corresponding author.

E-mail addresses: czine.peter@econ.unideb.hu (P. Czine), kovacs.sandor@econ.unideb.hu (S. Kovács).

Previous research on HCM has advanced in both methodological and empirical terms. Progress has been made in model formulation, estimation methods, and practical applications in various decision-making contexts [11–16]. More specific methodological contributions have also improved the estimation and calibration of ICLV models. For example, Fu and Juan [17] compared three estimation approaches — one sequential and two simultaneous — to evaluate their ability to recover the parameters of the multinomial probit-kernel ICLV model. Their results show that the two simultaneous approaches outperform the sequential approach in terms of estimate accuracy and efficiency, irrespective of sample size. The maximum approximate composite marginal likelihood (MACML) approach is preferable due to its superior performance in recovering true parameter values with relatively small standard errors, particularly when the sample size is large. Sohn [18] adapted the expectation–maximization algorithm to address empirical identification issues and simplify the calibration procedure for ICLV models. However, existing review studies have primarily focused on specific application areas (e.g., environmental economics, [19]) or broader trends in discrete choice modeling (e.g., data-driven and machine learning approaches, [20]). While these studies provide valuable insights, they do not offer a comprehensive bibliometric evaluation of HCM/ICLV research as a distinct modeling paradigm.

Additionally, recent studies in the *Decision Analytics Journal* highlight the importance of modeling heterogeneity and applying advanced analytical approaches in decision-making contexts. For example, Mahpour et al. [21] examined how taste variations influence travelers' destination choices. They demonstrated that Mixed Logit (ML) and Latent Class Variable (LCV) models can better capture psychological heterogeneity than traditional Multinomial Logit benchmarks. Similarly, Cao et al. [22] evaluated the ability of parametric, semiparametric, and nonparametric machine learning approaches, such as twin neural networks and Gaussian processes, to learn and predict complex behavioral choice rules under varying conditions of data availability and choice determinism. Related contributions in the journal further demonstrate the relevance of advanced decision-analytic methods across forecasting, travel behavior, and sustainable transportation contexts. Chhajet et al. [23] evaluated the performance of advanced predictive algorithms in financial and investment forecasting. Aydin et al. [24] examined how disruptive events such as the pandemic reshape travel patterns and mode-choice perceptions, and Palit et al. [25] proposed an integrated Principal Component Analysis and Interpretive Structural Modeling approach to identify hierarchical factors influencing electric vehicle adoption and sustainable transportation systems. Collectively, these studies position the present work within the broader scope of *Decision Analytics Journal* by linking predictive analytics, behavioral modeling, and strategic decision support.

The present study complements these works by shifting the focus from individual modeling applications to field-level bibliometric analysis. It demonstrates how HCM/ICLV research has evolved in terms of publication dynamics, citation structure, collaboration networks, and thematic priorities.

To address this gap, the present study provides a comprehensive bibliometric analysis of empirical HCM research published between 2010 and 2026. This analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26]. This analysis pursues four objectives:

- Examine temporal trends in publication output and scientific impact
- Identify leading journals, authors, and citation hubs
- Analyze patterns of international collaboration and knowledge transfer
- Map thematic directions through keyword structures and their evolution over time

The year 2026 was chosen as the cutoff year because the bibliographic data were extracted on January 12, 2026. This enables the study to incorporate the latest literature and use these preliminary publications to predict future research trajectories and short-term trends within the HCM domain.

The present study has two contributions. First, it provides a structured overview of the empirical HCM literature, clarifying its publication dynamics, intellectual core, and thematic development. Second, by interpreting bibliometric patterns, it identifies the dominant application domains and methodological orientations that have shaped HCM research over the past decade and a half. This establishes a foundation for more targeted empirical and methodological inquiry.

2. Methodology

This study uses bibliometric analysis to examine the scientific landscape of HCMs systematically. As a quantitative approach, bibliometrics enables the identification of research trends, structural configurations, and dominant patterns within a defined field. To ensure transparency and methodological rigor, the procedure followed guidelines adapted from PRISMA 2020 guidelines for systematic reviews [26]. Fig. 1 summarizes the complete workflow, including data collection, cleaning, and analysis, as a PRISMA-style flow diagram.

2.1. Data collection and source selection

The analysis drew exclusively from the Web of Science (WoS) and Scopus databases because they are recognized for their rigorous peer review standards and high-quality, English-language coverage, which is consistent with the scope of the study. To maintain a strict focus on peer-reviewed sources, Google Scholar was excluded. To comprehensively map the research landscape related to hybrid choice models, we employed a relatively inclusive search strategy. Rather than limiting the sample to studies that strictly implemented fully specified HCM frameworks, the search captured methodologically related approaches that contributed to the development and diffusion of the HCM/ICLV research ecosystem. This broader scope increases the dataset's heterogeneity. However, this trade-off was necessary to identify methodological linkages, citation structures, collaboration patterns, and emerging research directions that would likely remain obscured under a narrower definition. Accordingly, the relatively large number of retrieved publications reflects the study's broad bibliometric scope and should not be interpreted as representing the core HCM literature alone. To ensure comparability between the two databases, both Scopus and Web of Science were queried on January 12, 2026. The search strategy targeted the terms “Hybrid choice model*”, “Integrated choice and latent variable*”, and “ICLV” in titles, abstracts, and keywords. The queries were restricted to the document type “Article” and the language “English”:

— Scopus

“Hybrid choice model*” OR “Integrated choice and latent variable*” OR “ICLV”) AND (LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (LANGUAGE, “English”))

— WoS:

“Hybrid choice model*” OR “Integrated choice and latent variable*” OR “ICLV”) (Topic) and Article (Document Type) and English (Language)

The search returned 2657 records from Scopus and 527 from WoS, yielding 3184 initial records. Data were exported in their native formats: WoS as a plain text file (*savedrecs.txt*) and Scopus as a .csv file. The datasets were integrated using the `mergeDbSources()` function in the *bibliometrix* package. De-duplication was primarily anchored by DOI-based matching and supplemented by the package's automated fuzzy-matching algorithm to reconcile metadata inconsistencies across the two databases.

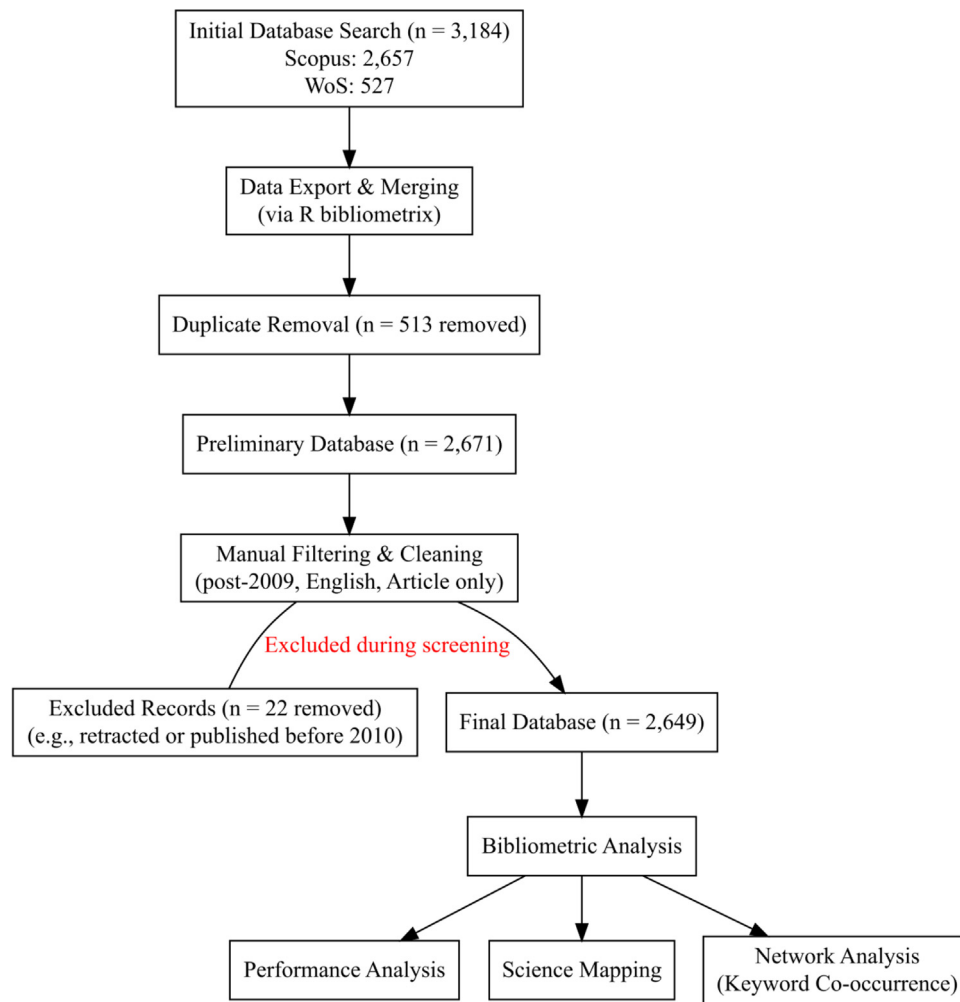


Fig. 1. Research workflow.

2.2. Data processing and refinement

All records were processed using R v. 4.2.3 and its *Bibliometrix* package, primarily through the graphical interface *Biblioshiny*. First, the raw files were imported into *Biblioshiny* and converted into compatible formats (*RData* for Web of Science and *Excel* for Scopus). The datasets were merged, and automated de-duplication removed 513 duplicate records. This resulted in a preliminary database of 2671 articles. A manual post-processing step was conducted within the merged Excel file (*merged.xlsx*). Citation data were standardized by transferring entries from the *CR_Raw* column to the *CR* column to ensure accurate parsing of citations. An additional manual screening phase involved removing 22 records identified as ineligible (e.g., editorial notes, retracted papers, or publications dated prior to 2010). This produced a final dataset of 2649 articles. To increase procedural transparency, this screening was subjected to a verification process by a second checker. A second researcher independently validated the eligibility of all removed records based on the predefined exclusion criteria. Inter-rater agreement was assessed and found to be 100%, confirming that the manual filtering was consistent and replicable.

2.3. Data analysis

The finalized dataset of 2649 articles was reimported into the *Bibliometrix* R package for analysis. In line with established bibliometric research standards, three analytical components were implemented:

- **Performance analysis** assessment of publication output and citation metrics, including total publications, citation counts, leading authors, and country contributions
- **Science mapping**: citation analysis of influential journals and articles, complemented by thematic mapping to identify dominant and emerging topics
- **Network analysis**: keyword co-occurrence analysis to examine the conceptual structure of the field and patterns of scholarly collaboration

Performance analysis utilized *biblioAnalysis*. The keyword co-occurrence network was generated using *biblioNetwork* (*analysis* = “co-occurrences”, *network* = “author_keywords”, *sep* = “;”) with a minimum frequency threshold of 5. Thematic mapping was executed via *thematicMap()* using the Walktrap clustering algorithm to delineate thematic clusters.

3. Results

3.1. Publication dynamics and descriptive statistics

The first stage of the bibliometric analysis provides a descriptive statistical overview of the dataset and establishes the scope and core parameters of the study. [Tables 1a](#) and [1b](#) present the key quantitative indicators, distinguishing between general corpus-level descriptors and document-type, authorship, and collaboration-related metrics.

Table 1a
Descriptive statistics of the bibliometric corpus.

Description	Value
Timespan	2010–2026
Sources (Journals, Books, etc.)	602
Documents	2649
Annual Growth Rate %	13.79
Document Average Age	4.18
Average citations per document	22.06
Average citations per year per document	3.35
References	34,396
Keywords Plus (ID)	8842
Authors' Keywords (DE)	7111

Table 1b
Document type, authorship, and collaboration indicators.

Description	Value
Article	2627
Article; Book Chapter	1
Article; Data Paper	1
Article; Early Access	15
Article; Proceedings Paper	5
Authors	5176
Author Appearances	9570
Authors of single-authored documents	117
Single-authored documents	125
Documents per author	0.512
Co-authors per document	3.61
International co-authorships %	7.74

The corpus comprises 2649 documents published between 2010 and 2026, with an annual growth rate of 13.79% and an average document age of 4.18 years. These indicators suggest a young and rapidly expanding research field. The authorship structure also shows that HCM research is predominantly collaborative, although international co-authorship remains relatively moderate. The annual production pattern further indicates that growth accelerated particularly after 2018 and reached its highest level in the mid-2020s. This expansion was likely driven by the increasing relevance of travel behavior research after the Covid-19 period, the growing policy importance of sustainable and emerging mobility technologies, and the wider availability of estimation tools such as the R Apollo package.

Overall, these patterns indicate that the growth of HCM research reflects more than a simple increase in publication volume. They suggest that HCM has gradually become institutionalized as a practical modeling framework in research areas where attitudes, perceptions, and behavioral heterogeneity play a central role in explaining decision-making.

3.2. Scientific structure in the field of HCM: journals, references, and authors

The citation structure indicates that HCM research has consolidated around a set of transport, methodological, behavioral, and environmental journals. Table 2 reports the top sources based on global citations, local citations, and local citation share. Global citations capture broader scientific visibility, while local citations indicate the extent to which a journal functions as an internal reference point within the analyzed HCM/ICLV corpus. The local citation share therefore helps distinguish journals with general citation influence from those with a more concentrated field-specific role.

The ranking shows that the intellectual core of the field is anchored in transportation and choice-modeling outlets. *Transportation*, *Transportation Research Part A: Policy and Practice*, and the *Journal of Choice Modeling* occupy central positions in the local citation network, while *Transportation Research Part B: Methodological* has the strongest field-specific profile among core transport journals based on its local citation

share. At the same time, the presence of *Travel Behavior and Society*, *Transport Policy*, and *Science of the Total Environment* indicates that HCM research is not confined to methodological development, but is increasingly connected to behavioral, policy-oriented, and environmental debates.

Table 3 supplements the citation-based ranking by comparing journals based on two interrelated criteria: publication productivity within the analyzed corpus and influence within the local citation network. Publication counts indicate a journal's activity within the domain, and local citations capture its role as a foundational reference within the analyzed corpus. For example, *Transportation Research Part A* was cited 2307 times in the reference lists of the sampled articles, underscoring its status as a primary source of knowledge.

Transportation Research Part A and *Transportation* served as foundational knowledge sources within the field. Their prominence reflects both high publication output and sustained citation reliance, confirming their role as core reference outlets for HCM-related research. The most frequently cited studies (Table 4) further evidence the intellectual depth and international relevance of the field. Examining these highly cited contributions clarifies the thematic and methodological foundations that shape the trajectory of HCM research.

One important conclusion from Table 4 is that frequently cited works are not always direct empirical applications of HCM. Instead, several of them served as methodological, review-oriented, or conceptually related contributions that shaped the broader academic context in which HCM/ICLV research evolved. Based on their thematic focus and methodological approach, these studies can be classified into three main categories.

The first group consisted of methodological and review-oriented contributions that, while not directly applying HCM, significantly shaped related modeling frameworks and research directions. Manering and Bhat [27] reviewed advances in traffic safety analysis and identified the HCM/ICLV framework as a promising approach for incorporating psychological latent constructs. Hess and Palma [29] introduced the Apollo software package, a general platform for choice model estimation in which HCM is one example of an application. Their main contribution is making it easier to implement advanced choice models in practice. Similarly, Hair et al. [35] proposed an alternative to the HCM/ICLV framework for discrete choice analysis using partial least squares (PLS) structural equation modeling (SEM).

The second group included empirical and review studies that examined the behavioral and social determinants of emerging mobility technologies outside of formal choice modeling frameworks. For example, Wang et al. [28] analyzed intentions to adopt hybrid electric vehicles using a behavioral model grounded in psychological constructs, without applying choice modeling. Liao et al. [30] reviewed consumer preferences for electric vehicles, identifying HCM as a sophisticated tool for addressing preference heterogeneity, while noting its practical limitations in decision-support contexts. Mersky et al. [31] investigated electric vehicle adoption in Norway using aggregate empirical data without employing HCM. Garrow et al. [36] conducted a systematic review of urban air mobility and electric and autonomous transport technologies. They referenced HCM primarily as a related methodological option rather than as an applied framework.

The third group consisted of empirical studies closely aligned with typical HCM application domains that employed alternative modeling strategies. For example, Alemi et al. [33] examined ride-hailing usage with binary logit models incorporating attitudinal constructs as manifest variables derived from factor analysis. This approach references HCM conceptually but does not implement it. Hackbarth and Madlener [34] analyzed preferences for alternative fuel vehicles using a mixed logit specification that did not implement latent variables. Among the top-cited contributions, Jensen et al. [32] was the only study that directly applied HCM. Using two waves of experimental data, the authors demonstrated how real-world electric vehicle experience reshapes preferences and how attitudinal constructs influence choice

Table 2
Top 10 sources based on local and global citations.

Journal	Global citations	Local citations	Local citation share (%)
Transportation	3813	333	8.7
Transportation Research Part A: Policy and Practice	7430	288	3.9
Journal of Choice Modeling	1722	202	11.7
Transportation Research Part B: Methodological	1091	165	15.1
Transportation Research Record	1018	116	11.4
Travel Behavior and Society	2192	94	4.3
Transportation Research Part F: Traffic Psychology and Behavior	1211	67	5.5
Transport Policy	1943	63	3.2
Transportation Research Part C: Emerging Technologies	2518	62	2.5
Science of the Total Environment	304	60	19.7

Note: Local citation share was calculated as local citations divided by global citations, multiplied by 100.

Table 3
Sources by articles and local citations.

Journal (from local database)	No. Articles published	Journal (from cited references)	Local citations
Transportation Research Part A: Policy and Practice	193	Transport Res A-Pol	2307
Sustainability (Switzerland)	106	Transportation	1187
Transportation	97	Transport Res D-Tr E	916
Travel Behavior and Society	93	Transport Res Rec	899
Transportation Research Part D: Transport and Environment	90	Transport Policy	870
Transport policy	85	Transport Res B-Meth	724
Transportation research record	80	Transport Res F-Traf	642
Transportation Research Part F: Traffic Psychology and Behavior	58	Transport Res C-Emer	638
Journal of Choice Modeling	55	J Transp Geogr	402
Transportation Research Part C: Emerging Technologies	52	Travel Behav Soc	394

Table 4
Top 10 globally cited papers in the corpus.

Paper	Total citations	Total citations per Year	Normalized total citations
Mannering & Bhat, 2014 [27]	1054	81.1	11.25
Wang et al. 2016 [28]	649	59.0	9.59
Hess & Palma, 2019 [29]	635	79.4	14.04
Liao et al. 2017 [30]	601	60.1	11.10
Mersky et al. 2016 [31]	435	39.5	6.43
Jensen et al. 2013 [32]	425	30.4	5.63
Alemi et al. 2018b [33]	424	47.1	9.93
Hackbarth & Madlener, 2013 [34]	404	28.9	5.35
Hair et al. 2019 [35]	390	48.8	8.62
Garrow et al. 2021 [36]	348	58.0	11.71

within a hybrid modeling framework. Overall, this distribution indicates that, although HCM is a central methodological reference in the literature, the most highly cited works do not empirically implement the HCM framework.

The citation structure of the sampled publications provides further insight into the impact dynamics of the field. A closer examination of citation patterns clarifies the dynamics underlying the previously reported average of 22.06 citations per document. While this average indicates substantial scholarly attention, the distribution is highly skewed. Most publications — 74.44% (1972 articles) — had fewer than 50 citations. An additional 14.65% remained uncited, which reinforces the relative recency of much of the corpus. This indicates that the overall average is disproportionately driven by a small group of highly cited works at the apex of the citation hierarchy. Only 30 studies (1.13%) had received more than 200 citations, forming the field's most influential intellectual core. Along with the 3.02% of articles cited between 100 and 200 times, these works served as theoretical anchors that substantially influenced the global visibility of the corpus. This asymmetric structure reflects a common bibliometric pattern in which a limited number of foundational contributions define the main directions of scientific discourse and a broad base of publications consolidates and extends emerging themes.

The concentration of citations was closely linked to author-level performance. Highly cited publications were typically associated with

a relatively small group of influential scholars whose methodological contributions and theoretical frameworks served as reference points for the broader community. Thus, the overall citation average (22.06) and the elite segment exceeding 200 citations were largely sustained by the output of these leading contributors. Since more than 4% of documents accounted for a disproportionate share of total citations, it is essential to identify the authors behind these seminal works to understand the field's intellectual structure. These scholars contributed not only quantitatively, but also by shaping methodological standards and research trajectories. Table 5 lists the most influential authors, ranked by publication volume and citation impact. It also highlights the key contributors to the HCM research network.

The author-level results reinforce the concentration of scientific influence within a limited core of scholars. While traditional indicators such as total citations and the h-index capture individual impact, the network measures show how these authors are positioned within the collaboration structure of the field. Zhang, Y. has the broadest direct co-authorship network, whereas Hess, S. occupies an important bridging position across thematic and institutional clusters. The PageRank results further indicate that influence in HCM research depends not only on publication volume, but also on the structural position of authors within the wider knowledge network.

This structural dominance aligns with the broader research directions observed in the global citation landscape. While highly cited

Table 5
Top 10 authors ranked by PageRank centrality.

Author (h-index)	Total citations	Articles (fractionalized)	Degree	Betweenness	PageRank
Hess S. (20)	2064	46 (15.48)	84	0.04708	0.00255
Zhang Y. (12)	529	39 (8.95)	127	0.02229	0.00255
Wang Y. (14)	980	47 (11.07)	112	0.03806	0.00245
Ortúzar J. (16)	1093	33 (9.84)	77	0.02162	0.00209
Li J. (10)	307	27 (6.27)	93	0.01497	0.00184
Wang S. (10)	1247	28 (6.69)	73	0.01393	0.00165
Arellana J. (14)	519	22 (5.95)	46	0.00957	0.00146
Timmermans H. (18)	1660	30 (9.44)	40	0.00902	0.00138
Irawan M. (11)	411	21 (5.63)	31	0.00252	0.00132
Feng T. (10)	423	28 (8.11)	44	0.00201	0.00131

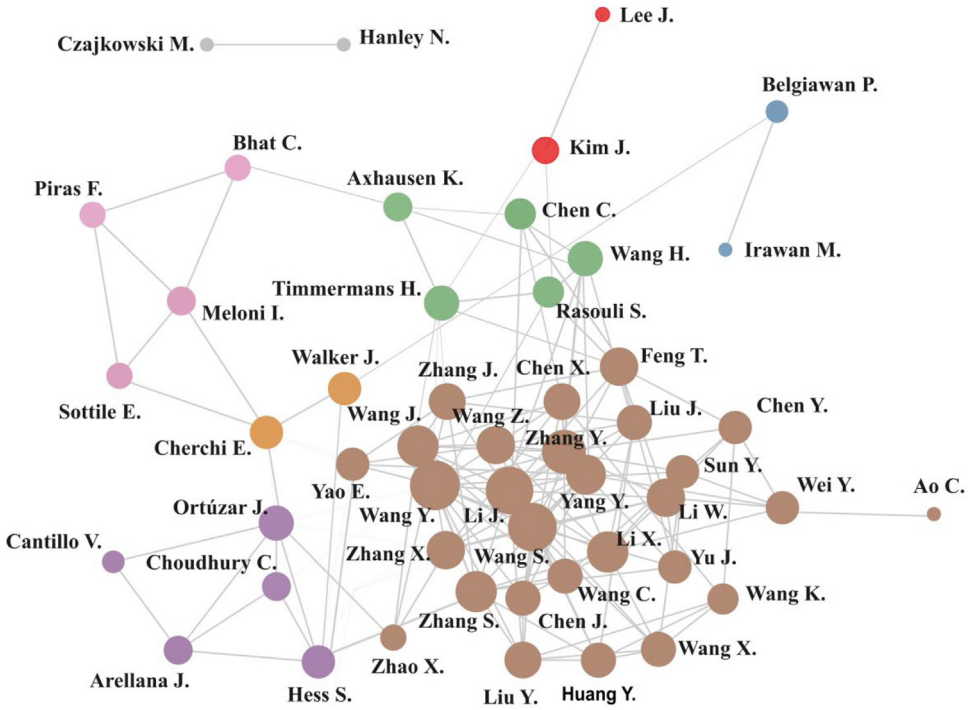


Fig. 2. Author collaboration network in hybrid choice modeling research.
Note: Node sizes are proportional to an author's total publication output. Edge thicknesses indicate the frequency of joint co-authorship. The network reveals three major clusters: (1) a dense brown cluster representing the central Chinese core, which is characterized by high internal collaboration, (2) a purple cluster representing a Western methodological bridge connecting classical discrete choice traditions; and (3) the smaller green and pink peripheral clusters representing specialized research communities surrounding key senior scholars.

classic works [27,34] anchor the theoretical foundations, authors such as Hess, S. and Wang, S., sustained and extended these foundations within the contemporary network. Hess's contributions were primarily methodological, strengthening the formal architecture of choice modeling. Meanwhile Wang, S. occupied a central position in empirical research on transportation mode choice and technology adoption. These patterns suggest that leading scholars within the local collaboration network continued to build systematically on established theoretical and methodological traditions up to 2026, preserving intellectual continuity while advancing applied research.

Following the identification of key authors and centrality metrics, collaboration structures were directly examined. Fig. 2 visualizes the co-authorship network: node size represents publication volume, colors denote collaborative clusters, and connecting lines indicate joint publications.

The co-authorship network reveals three main collaboration patterns. First, a dense Chinese cluster forms the most interconnected segment of the network, reflecting both the country's high publication output and strong internal collaboration. Second, a Western methodological cluster, including authors such as Hess, S., Ortúzar, J., and Choudhury, C., acts as a bridge between classical choice-modeling

traditions and contemporary HCM applications. Third, smaller peripheral clusters are organized around specialized research communities and senior scholars. Overall, the network indicates that HCM research has expanded globally, but knowledge production remains structured around a few dominant regional and methodological communities.

3.3. Geographical distribution and international cooperation

Table 6 reports scientific productivity by country, based on the locations of corresponding authors. This provides insight into the geographical distribution of research activity and patterns of international collaboration.

Table 6 shows a clear geographical concentration of HCM research. China is the dominant contributor, followed by the United States and the United Kingdom, while several European countries form a stable secondary research base. This distribution suggests that the global development of HCM has been shaped by a combination of large national research systems, transport-policy relevance, and strong methodological traditions.

The MCP indicator adds an important qualification to this pattern. Although China has the largest absolute output, the United States, the

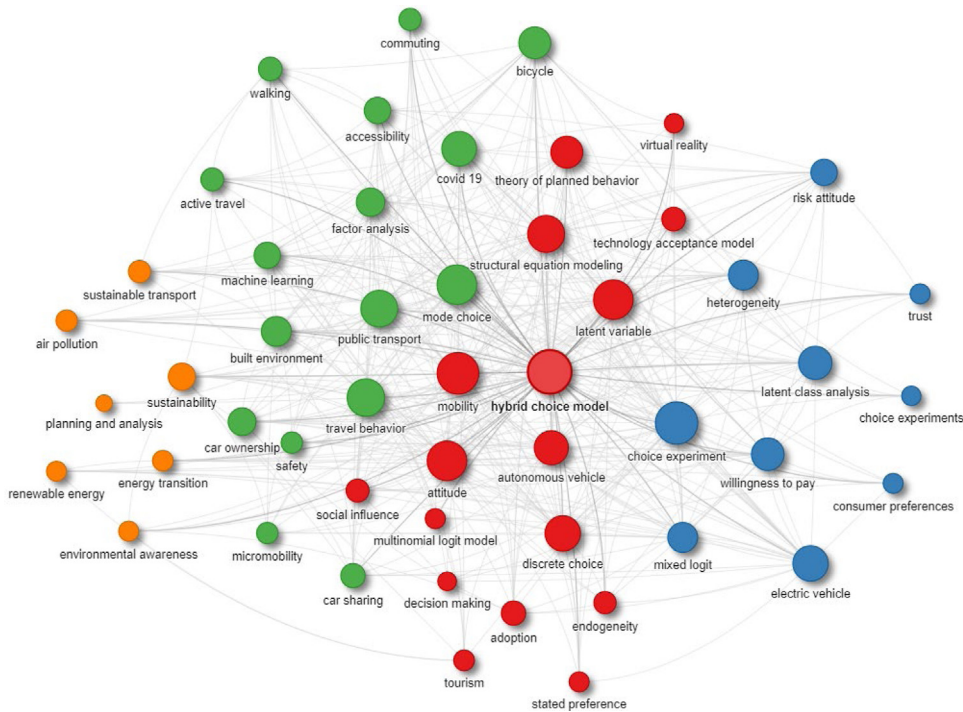


Fig. 3. Keyword co-occurrence network of author keywords.
Note: Node sizes indicate the frequency with which keywords occur in the corpus, and lines represent co-occurrence links. The four colored clusters represent the primary thematic divisions of the discipline: Methodology and Behavioral Modeling (red), Economic Evaluation and Preference Heterogeneity (blue), Travel Behavior and Urban Mobility (green), and Environmental Sustainability and Planning (orange).

Table 6
Top countries by corresponding author publication output.

Country	TP (MCP)	% TP	Country	TP (MCP)	% TP
China	546 (21)	23.7%	Korea	63 (2)	2.7%
USA	240 (23)	10.4%	Iran	54 (3)	2.3%
UK	139 (23)	6%	Indonesia	50 (2)	2.2%
Germany	113 (12)	4.9%	Japan	41 (9)	1.8%
Italy	111 (10)	4.8%	Chile	43 (6)	1.9%
The Netherlands	101 (10)	4.4%	Switzerland	29 (9)	1.3%
India	95 (3)	4.1%	Colombia	32 (3)	1.4%
Australia	84 (10)	3.7%	Norway	30 (4)	1.3%
Spain	58 (11)	2.5%	Sweden	28 (4)	1.2%
Canada	66 (1)	2.9%	Denmark	23 (6)	1%

Note: TP: Total Production; MCP: Multiple Country Publications.

United Kingdom, and several smaller European countries show stronger relative international integration. The collaboration structure therefore suggests a bifurcated global field: one pole centered on China and the United States, and another organized around European research networks. This pattern highlights both the globalization of HCM research and the persistence of regional knowledge communities.

3.4. Thematic structure and temporal evolution

Understanding the intellectual evolution of the field requires examining the temporal dynamics of keywords and thematic clusters beyond descriptive statistics and collaboration networks. Changes in the prominence of keywords reflect shifts in scientific emphasis, from foundational methodological development to the application of HCM in emerging technological and policy domains. Fig. 3 presents the co-occurrence network of author keywords. In this network, node size represents keyword frequency, colors denote thematic clusters, and connecting lines indicate the strength of conceptual associations. This visualization captures the discipline’s internal structure and reveals how its thematic domains are organized and interconnected.

The keyword co-occurrence network shows that HCM research is organized around four closely connected thematic domains: methodological and behavioral modeling, economic evaluation and preference heterogeneity, travel behavior and urban mobility, and environmental sustainability and planning. The central position of “hybrid choice model” indicates that the method functions as a bridge between econometric modeling, psychological constructs, and applied mobility research. At the same time, the links with keywords such as “electric vehicle”, “autonomous vehicle”, “public transport”, “sustainability”, and “environmental awareness” show that HCM has become increasingly embedded in policy-relevant and technology-oriented research contexts.

Fig. 4 presents the trend topics of author keywords based on their frequency and active duration between 2012 and 2025 to capture the chronological emergence and lifespan of these terms. The plot reveals a distinct evolutionary trajectory. Early scholarship (prior to 2016) focused primarily on core econometric foundations, including the MACML estimation approach, structural equation models, and random utility theory. During the intermediate phase (2018–2022), literature shifted toward hybrid choice models, latent variables, and attitudes; electric vehicles emerged as a primary application domain. More recently, from 2023 onward, contemporary, technologically integrated keywords, such as machine learning, risk perception, autonomous vehicles, virtual reality, and mobility as a service (MaaS), have gained prominence. This pattern indicates a clear transition from foundational modeling theory to applied, data-driven, and user-centric mobility frameworks.

Fig. 5 illustrates the evolution of the field’s themes, showing how early foundational concepts were reconfigured and integrated into new research areas by 2026. The evolutionary map illustrates the adaptability of transportation research discourse. While core themes from the 2010–2021 period, such as HCM, attitudes, and electric vehicles, persisted, they progressively merged into more complex, interdisciplinary configurations by 2025–2026.

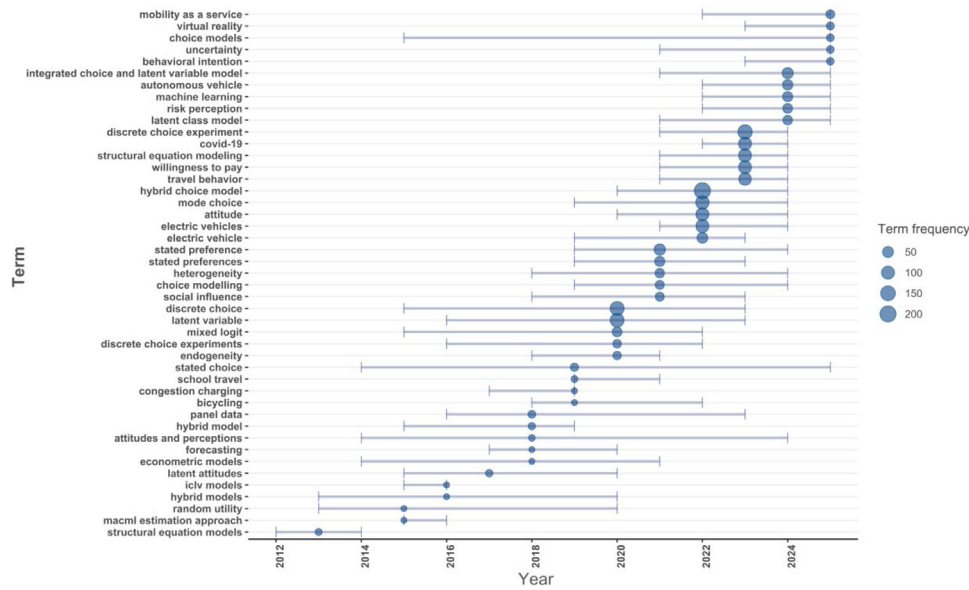


Fig. 4. Trend topics of author keywords.

Note: The vertical axis lists the most prominent author keywords, and the horizontal axis shows the timeline from 2012 to 2025. The horizontal lines show the active lifespan of each term (the first to third quartiles of publication years), and the blue circles show the term's overall frequency. The plot shows a clear shift from the early methodological foundations (before 2016) to the applied, data-driven mobility structures (after 2022).

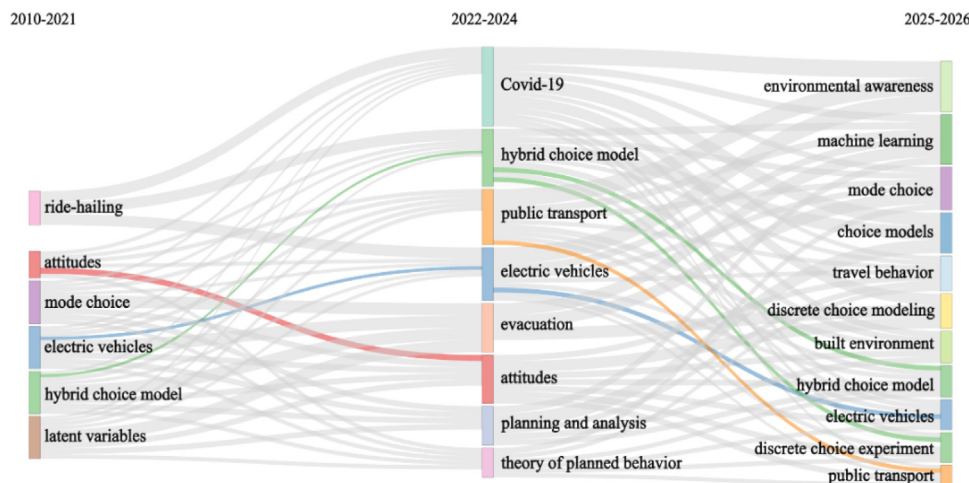


Fig. 5. Thematic evolution of research topics (2010–2026).

Note: This Sankey diagram illustrates the flow and distribution of thematic keywords across three time periods: 2010–2021, 2022–2024, and 2025–2026. The width of the bands represents the volume of publications covering these themes. It shows how traditional concepts, such as ride-hailing and attitudes, have evolved and merged into interdisciplinary fields like machine learning and environmental awareness.

As shown in Fig. 5, machine learning and environmental awareness are the most prominent topics in the 2025–2026 period, as reflected by the width of their bands. Three structural shifts are evident:

- **Methodological Stability and Diversification:** The trajectory from hybrid choice models and latent variables toward DCM combined with machine learning indicates continuity rather than rupture. Traditional decision-theoretic frameworks were progressively integrated with artificial intelligence-based techniques, expanding analytical capacity without displacing foundational principles
- **Growing Centrality of the Human Factor:** Initially, attitudes appeared as a distinct thematic block. By 2026, however, they were embedded within broader domains, such as travel behavior and environmental awareness. This shift suggests that psychological constructs are no longer peripheral, but rather constitute a cross-cutting dimension of the field

- **Technological–Social Convergence:** Themes such as electric vehicles and ride-hailing gained prominence during the 2022–2024 period in connection with crisis management and planning (e.g., the response to the pandemic). By 2026, these topics had converged with concerns related to the built environment and systemic sustainability, reflecting a closer integration of technological innovation and long-term planning frameworks

Following the analysis of temporal dynamics and thematic interconnections, a complementary static assessment of the current intellectual structure is necessary. Fig. 6 (Thematic Map) positions keywords along two dimensions (i.e., network centrality (importance) versus density (internal development)). This distinguishes core drivers, foundational themes, specialized domains, and emerging or declining topics.

The four quadrants of the thematic map illustrate the current priorities of the research landscape:

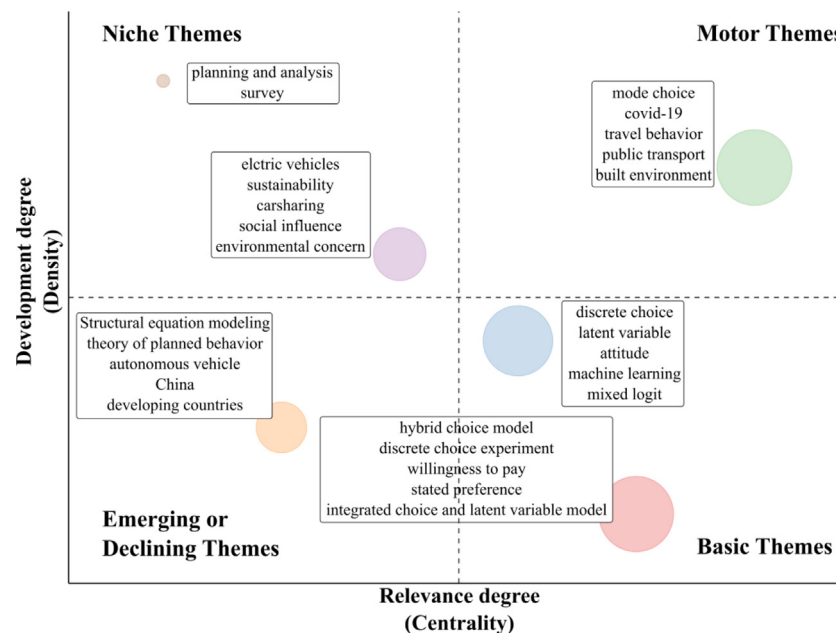


Fig. 6. Thematic map of author keywords.

Note: This diagram divides the thematic domain into four quadrants according to centrality (importance to the network, X-axis) and density (internal development of the subfield, Y-axis). The upper-right quadrant contains mature Motor Themes, the lower-right quadrant contains foundational Basic Themes, the upper-left quadrant represents specialized Niche Themes, and the lower-left quadrant displays Emerging or Declining Themes.

- **Motor Themes (upper-right quadrant):** These themes have both high centrality and high density, indicating well-developed and influential topics. Mode choice, travel behavior, and the built environment occupy this quadrant. The continued presence of pandemic-related topics suggests that this line of inquiry has matured and become structurally embedded.
- **Basic Themes (lower-right quadrant):** These topics have high centrality but lower density, indicating foundational importance with more stable internal development. Core methodological concepts, such as DCM, latent variables, machine learning, and mixed logit, form the analytical infrastructure of contemporary research.
- **Niche Themes (upper-left quadrant):** These themes are internally developed but less central to the broader network. Electric vehicles, car sharing, and sustainability belong to this category, indicating specialized and coherent subfields with more limited influence across the wider transport discourse.
- **Emerging or Repositioned Themes (lower-left quadrant):** Structural equation modeling and autonomous vehicles appear in this quadrant. Their placement suggests thematic consolidation, as research has shifted from initial conceptual interest toward more focused and critical investigation.

To evaluate the depth of methodological engagement within this thematic structure further, a quantitative text-mining analysis was conducted on the titles, abstracts, and author keywords of the entire corpus of 2649 documents. This analysis does not aim to identify every methodological paper in the corpus, but rather to measure the visibility of terms directly associated with methodological reflection in these bibliometrically searchable fields. The search terms were systematically grouped into five core dimensions of methodological rigor: (1) model specification and sensitivity analysis, (2) measurement model validation, (3) model identification and estimation stability, (4) out-of-sample validation and predictive performance, and (5) reproducibility and methodological transparency. The results indicate that only 3.62% (96 unique documents) of the published literature engage with these dimensions explicitly in their metadata. The top five most frequently identified terms in the specific vocabulary were “sensitivity analysis” (29 occurrences), “predictive power” (12 occurrences), “predictive

performance” (10 occurrences), “replicate” (eight occurrences), and “convergence” (seven occurrences). Furthermore, a stark asymmetry was observed in the distribution of these terms: the vast majority of hits — 95 occurrences — were in the abstracts, while only five appeared in the author keywords and two in the document titles. This distribution reveals that methodological reflections are almost exclusively embedded deep within the abstracts and rarely serve as primary identifiers or central selling points of published research. This suggests that recent growth in this modeling domain is largely problem-driven with less emphasis on cumulative methodological refinement, as exemplified by highly applied, policy-oriented simulation frameworks [37].

4. Discussion

The bibliometric results indicate that, between 2010 and 2026, empirical research on HCM expanded rapidly and has consolidated into a thematically and methodologically structured field. The growth in publication volume has been accompanied by an increase in differentiation along thematic, geographical, and institutional dimensions. These patterns reflect the maturation of the field and the diversification of application domains over the past decade and a half.

These results should not be interpreted merely as evidence of quantitative growth. Rather, they suggest that HCM/ICLV has gradually evolved from a specific methodological extension of discrete choice modeling into an increasingly institutionalized analytical framework for examining behaviorally complex decisions. This development has three important implications. First, the rapid growth in publication activity signals the wider adoption of the method, but it also increases the risk that hybrid specifications will be applied routinely without sufficiently explicit theoretical justification. Second, the fact that scientific influence is concentrated around a limited group of journals, authors, and countries suggests that the methodological standards of the field are shaped by a relatively narrow intellectual core. On the one hand, this supports cumulative development; on the other hand, it may limit the diversity of empirical contexts and model-specification strategies. Third, the thematic shift toward sustainability, emerging mobility technologies, and machine learning indicates that HCM is

increasingly being applied to examine policy-relevant and technologically complex problems. Consequently, the future development of this field depends not only on more sophisticated models, but also on clearer reporting standards, stronger validation practices, and a more transparent interpretation of latent psychological constructs.

4.1. Thematic focuses in the field of HCM

Thematic mapping of keywords indicates that the field has expanded from conventional utility-maximizing choice specifications toward behaviorally richer decision models that incorporate latent attitudes, perceptions, and psychological heterogeneity. This development is consistent with the broader macro-trends identified by Haghani et al. [6], who showed that choice modeling has gradually moved from primarily mathematical and econometric abstractions toward greater behavioral realism. Many studies addressed specific policy and behavioral issues, particularly travel mode selection, sustainable mobility transitions, the adoption of new transportation technologies (e.g., electric and autonomous vehicles), and the evaluation of mobility services such as MaaS [38–42]. For example, Bhattarai et al. [43] used the HCM approach to study how people chose modes of transportation during the pandemic, revealing the influence of intangible factors such as safety and discomfort on these choices. Similarly, Tinessa et al. [44] used HCM to analyze how preferences and attitudes of public transit users toward onboard safety measures shifted in response to pandemic alert scenarios. In the realm of sustainable technology, Scorrano et al. [45] used ICLV models to compare the adoption of battery electric vehicles in Norway and Italy. This study revealed that deep-seated perceptions of environmental and economic factors vary by region. Meanwhile, Lindberg et al. [46] developed a hybrid choice model to understand the predictors of public preferences for coastal carbon sequestration. This study demonstrated the versatility of the framework in environmental policy contexts.

Over the past decade, applications related to electric vehicles and active transportation have gained prominence, aligning with broader energy policy and sustainability debates [37,47,48]. These studies often identified environmental attitudes, risk perception, and trust in technology frequently functioned as latent behavioral dimensions [32,49,50]. This trend remains highly relevant in recent assessments of active travel, transit perception, and sustainable transportation preferences [16,51]. A recurring finding across applications is that attitudes do not operate as isolated determinants. Rather, their influence is mediated by the broader contexts of travel behavior and environmental awareness [52]. The importance of mode choice, travel behavior, and the built environment suggests that HCM research is increasingly interpreting mobility decisions within systemic and spatial frameworks. Recent comparative studies contrasting theory-driven ICLV models with data-driven machine learning in ride-sourcing show that machine learning can improve predictive accuracy, while ICLV remains essential for behavioral interpretation [53]. Cao et al. [22] similarly show that model performance depends on training data limitations and the determinism of the choice rule.

Similarly, HCM applications extend beyond conventional attitudinal constructs. Some studies have highlighted subjective perceptions of alternatives (e.g., comfort) as latent dimensions, demonstrating the framework's flexibility in conceptualization and operationalization [54]. While the general discrete choice literature emphasizes the increasing integration of deep and machine learning classifiers [20], our bibliometric analysis reveals that, within the specific field of HCM, researchers are increasingly adopting a hybrid approach. This is consistent with observations in other subfields, such as environmental economics. Recent surveys suggest that traditional random parameter formulations and latent class models remain robust alternatives in this field due to the complexity and risk of misspecification in fully specified HCMs [19]. Mahpour et al. [21] reinforce this perspective. Their empirical comparative analysis of travelers' destination choices

demonstrated that both mixed logit and latent class models successfully capture latent psychological heterogeneity and taste variations. As expected, the mixed logit model offered a highly competitive fit to empirical behavioral data. Works such as that of Tsoleridis et al. [55] support the push toward hybridizing these methods. They introduced a hybrid framework combining probabilistic clustering with utility-based models to capture preference heterogeneity while maintaining policy-relevant interpretability. Similarly, Salazar et al. [56] integrated SHAP (Shapley Additive Explanations) with learning multinomial logit models to provide greater transparency into the predictions of machine learning-augmented choice models.

However, this thematic expansion is not without risks. Mariel et al. [19] discussed the trade-off between the behavioral richness of HCM and the danger of structural misspecification in environmental choice contexts. Since fully specified simultaneous HCMs introduce substantial parametric constraints, modelers may encounter issues related to convergence, stability, and model interpretation. The text-mining results presented in Section 3.4 quantitatively support this concern, showing that only 3.62% of the literature explicitly addresses model specification and validation challenges in the metadata. The fact that these terms are almost entirely confined to abstracts, rather than titles or author keywords, confirms that methodological reflection often remains secondary to empirical findings. Furthermore, the concentration of hits around practical terms, such as sensitivity analysis and predictive power, as opposed to the near absence of structural identification or transparency-focused concepts, suggests that empirical growth has often been prioritized over cumulative methodological refinement. Recognizing this, authors such as Rodrigues et al. [57] have advocated for a more holistic, iterative approach to model building that bridges the gap between machine learning and econometrics. This approach encourages practitioners to leverage both frameworks based on the specific needs of their research design.

4.2. Geographical, institutional, and methodological concentration

A structural interpretation of geographical and collaborative networks reveals an intellectual division of labor in the HCM ecosystem. Although East Asian research teams, particularly those in China, are a significant source of high-volume publications on electric and autonomous vehicles, the methodological core is firmly rooted in European and South American networks. Classical discrete choice traditions provided the theoretical foundation for later extensions that incorporated latent constructs. This methodological trajectory is evident in the seminal and critical contributions that formalized and evaluated the HCM/ICLV framework [7,58,59]. Based on this, a structured, methodological toolkit emerged. This toolkit includes measurement models, attitudinal indicators, and specification choices, which are recurring analytical issues [58,60]. These issues have been evaluated in recent handbooks [15] in terms of their behavioral realism, statistical efficiency, and policy relevance. Scholars who contributed both methodological and empirical work, and who were frequently cited across studies, played pivotal roles in disseminating methodological standards and linking alternative theoretical perspectives within the HCM community [29,61].

Similarly, bibliometric evidence suggests that although the general logic of HCM is widely accepted, specification strategies vary considerably across empirical applications. Justifications for modeling choices are often not explained in detail. Several authors have questioned the incremental value of HCM/ICLV in specific contexts, emphasizing its sensitivity to specification and its partial equivalence with alternative modeling approaches [59,62,63]. These concerns are particularly salient when models not only extend the role of latent constructs but also the underlying decision mechanism. Incorporating explicit links between latent constructs and decision rules increases sensitivity to theoretical assumptions further and may make empirical conclusions contingent on the chosen framework [64].

4.3. Technological background and interdisciplinarity as driving forces

The rapid increase in publication output is closely associated with advances in software accessibility. Earlier estimation of HCM/ICLV models often required substantial programming expertise, such as with *Python Biogeme*, *Latent Gold*, or custom implementations. However, the introduction of dedicated platforms and libraries, notably the *R Apollo* package in 2019, has substantially lowered these technical barriers. Consequently, hybrid modeling has evolved from a specialized methodological practice into a more widely adopted analytical standard. Although our bibliometric database did not contain formal adoption statistics for the *R Apollo* package, its release and growing recognition coincide with the period of accelerating publication growth observed after 2018. This suggests that the increased accessibility of the software may have contributed to the wider adoption of HCM applications by making complex model estimation more feasible for a broader range of researchers. Recent comprehensive reviews by Benjdiya et al. [65] emphasize this trend. They note that, while traditional discrete choice models are still effective for aggregate analysis, integrating machine learning techniques has become essential for more granular, disaggregate travel mode choice analyses. Similarly, Zhang et al. [66] highlight the broader trend of synergistically integrating metaheuristics and machine learning. This approach has proven highly effective in enhancing the efficiency and precision of complex decision-making systems in fields such as autonomous driving.

Parallel developments occurred in SEM, which constituted the psychological component of the hybrid framework. Software such as *Smart-PLS* and R packages including *lavaan*, *SEMinR*, and *plspm* now routinely support the measurement and validation of latent variables. These tools have facilitated the application of covariance-based SEM and PLS-SEM techniques, allowing researchers to validate measurement structures before integrating them into discrete choice models.

The broader availability of these analytical tools has increased HCM's interdisciplinary reach. The thematic structure identified earlier, particularly the behavioral and mobility-oriented cluster, illustrates how the framework connects transport economics with social psychology. By incorporating constructs derived from TPB and related psychological models, HCM enables systematic integration of subjective attitudes and preferences into decision-theoretic models. This addresses the limitations of approaches that rely solely on observable attributes.

4.4. Limitations of the bibliometric approach and future research directions

The bibliometric design of this study imposed inherent constraints. In particular, the relatively broad search strategy used resulted in a heterogeneous dataset because not all included studies presented strictly defined HCM applications. First, although we used the two standard academic databases, WoS and Scopus, the exclusion of other search tools, such as Google Scholar or Dimensions, may have resulted in the omission of regional or gray literature. Additionally, restricting the search to English-language articles may have introduced language bias by excluding relevant studies published in other languages. Second, our inclusive search terms captured a broader methodological ecosystem, which introduced a degree of thematic heterogeneity and reduced the specificity of the final corpus as a representation of strictly defined HCM applications. However, this approach was intentional because it allows for a more comprehensive characterization of the broader research field and the methodological ecosystem surrounding HCMs. Since the analysis relied on publication output, citation patterns, and keyword structures, it could not provide a detailed assessment of the methodologies used in specific models. Notably, it failed to capture the technical depth of latent variable operationalization and the subtle differences among estimation strategies.

These limitations are particularly relevant for the empirical measurement of latent psychological constructs. Attitudes and perceptions are context-sensitive and susceptible to measurement variation, which

can affect estimated parameters and complicate cross-study comparability [67]. Consequently, the robustness of HCM findings depends not only on data quality but also on theoretical framing and measurement design.

Looking ahead, the growing prominence of keywords related to machine learning techniques in HCM signals a methodological expansion. However, this development also highlights unresolved challenges. The absence of standardized software environments and harmonized validation protocols limits reproducibility and comparability across studies. Further research should systematically address estimation stability, model identifiability, and practical applicability. Previous studies have emphasized that specifying latent constructs and selecting measurement models can significantly impact empirical outcomes [35,62,68]. Specifically, we suggest that future work should focus on the following:

- investigating standardized validation protocols for machine learning–HCM integration
- exploring the utilization of natural language processing (NLP) and large language models (LLMs) to automatically extract and code latent psychological constructs from qualitative data or social media feeds
- systematically comparing estimation performance between classical frequentist and Bayesian approaches across different software implementations

Greater methodological transparency and comparative validation would strengthen the field's cumulative development.

4.5. Synthesis: A conceptual framework for HCM research

The bibliometric findings and identified trends were synthesized into the integrated conceptual framework presented in Fig. 7. The framework conceptualizes the evolution of HCM as a hierarchical and interdependent process. It illustrates the flow from foundational theories (e.g., RUT and TPB) through the methodological engine (latent variable–DCM integration) to contextual mobility applications. Ultimately, it supports the green mobility transition.

The model distinguishes four analytical levels:

- **Foundational Tier** This level captures the theoretical integration of RUT with social-psychological constructs (e.g., TPB). It represents the conceptual basis of hybrid modeling.
- **Methodological Engine**: The second tier encompasses software innovation and methodological refinement, including tools such as *R Apollo* and *Biogeme*. These tools enabled the formal integration of latent variables within DCM.
- **Contextual Applications**: At this level, theoretical and methodological advances are operationalized in empirical contexts. Hybrid models are applied to issues such as public transportation competitiveness, the adoption of shared mobility (e.g., car sharing), and the influence of the built environment on behavioral decisions.
- **Strategic Tier**: The upper tier reflects broader policy implications. Here, hybrid modeling functions as an analytical instrument for sustainability transitions, including energy system transformation and the diffusion of electric and autonomous vehicles.

Rather than serving as a static taxonomy or a simple summary of bibliometric trends, the four-tier conceptual framework provides an integrative interpretation of how HCM research connects microeconomic decision theory with socio-psychological behavioral constructs. Specifically, the framework links foundational theories, such as random utility theory and the theory of planned behavior, to methodological developments in latent variable–DCM integration, empirical applications in mobility and sustainability contexts, and broader strategic policy objectives. In doing so, it advances the existing literature by clarifying the

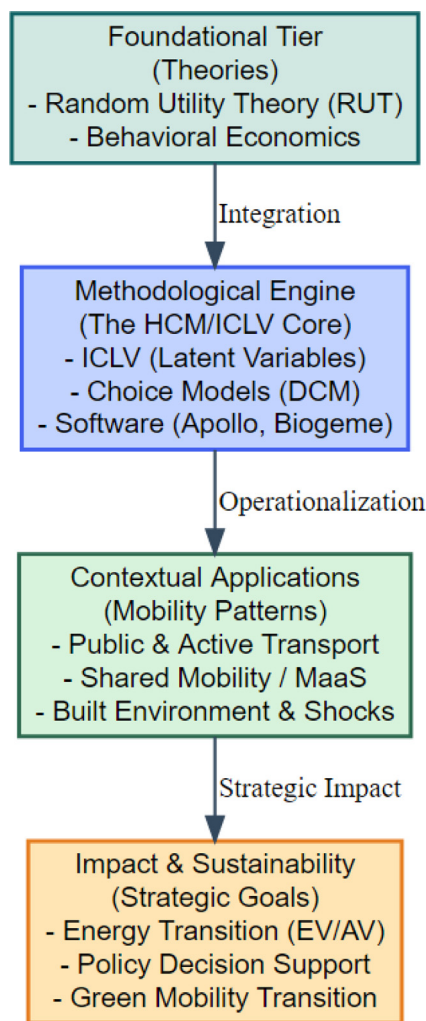


Fig. 7. Conceptual framework of the evolutionary ecosystem of hybrid choice modeling.

Note: This schematic summarizes the evolution of the HCM field into a four-tier, hierarchical system. Directional arrows represent conceptual flows and operational pathways. The Foundational Tier combines utility and behavioral theories, which are mathematically formalized in the Methodological Engine (software and models). The engine is operationalized through empirical studies of contextual applications, such as mobility patterns, which ultimately drive the strategic impact required to support green mobility policy and sustainable transition goals.

operational pathways through which abstract behavioral and utility-based assumptions are translated into empirical model specifications. The framework also highlights that the integration of latent constructs is not merely a technical extension of discrete choice modeling, but a broader conceptual development that expands the behavioral content of utility-based decision analysis. By organizing these components into foundational, methodological, contextual, and strategic tiers, the framework helps researchers more explicitly consider identification issues, model-specification choices, validation needs, and predictive versus explanatory trade-offs. Thus, it provides a conceptual bridge between behavioral science and decision analytics.

5. Conclusion

This study offers the first comprehensive bibliometric overview of the scientific landscape of HCM from 2010 to 2026. Based on an analysis of more than 2600 publications compiled according to

PRISMA guidelines, it shows that HCM has become an increasingly important analytical approach in transport-related decision-making research. In addition to documenting growth, the synthesis captures the internal maturation of the field, the impact of technological democratization, and the evolving strategic role of methodology in sustainability transitions.

The proposed conceptual framework is a key theoretical contribution of this work. It structures the intellectual development of HCM into four interactive tiers. By mapping the linkages between theoretical foundations, methodological tools, empirical applications, and strategic goals, the framework provides a cohesive interpretive structure that goes beyond descriptive bibliometric synthesis. This structure can help researchers and practitioners more explicitly connect latent-variable validation, utility specification, empirical context, and policy interpretation. In this way, the framework supports the development of empirical models that are both behaviorally meaningful and relevant for decision-making practice.

From a methodological perspective, the study reveals a growing gap between the rapid adoption of estimation software and the consistency of methodological reporting. While this technical accessibility has supported empirical expansion, it may also lead to less systematic attention to identification, specification, validation, and sensitivity checks. Addressing these issues requires clearer reporting standards and more rigorous validation protocols, especially as the field increasingly interfaces with data-driven machine learning algorithms.

In summary, HCM has evolved from a methodological innovation to a core instrument in transportation research for modeling complex human decision-making processes. The primary challenge for the coming decade is to build on the proposed conceptual framework and not only refine HCM incrementally, but also deploy it strategically to support sustainable, human-centered, and behaviorally informed decision-making systems.

Generative AI usage

Generative artificial intelligence (AI) tool was used to support parts of this research. Google's Gemini Pro 3.0 large language model was used to aid in code generation. All outputs produced by this tool were critically reviewed, verified, and edited by the authors to ensure accuracy, originality, and adherence to academic standards. The final responsibility for the content of this article rests entirely with the authors.

Funding

This research received no external funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors of the study would like to thank the University of Debrecen Program for Scientific Publication.

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Data availability

Data will be made available on request.

References

- [1] M.E. Ben-Akiva, S.R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA, 1985, <http://dx.doi.org/10.2307/2582065>.

- [2] J.J. Louviere, D.A. Hensher, J.D. Swait, *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, 2000, <http://dx.doi.org/10.1017/cbo9780511753831>.
- [3] K.E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, 2009, <http://dx.doi.org/10.1177/0049124110371323>.
- [4] D. McFadden, Conditional logit analysis of qualitative choice behavior, in: P. Zarembka (Ed.), *Frontiers in Econometrics*, Academic Press, New York, 1974, pp. 105–142.
- [5] D. McFadden, McFadden economic choices, *Am. Econ. Rev.* 91 (2001) 351–378, <http://dx.doi.org/10.1257/aer.91.3.351>.
- [6] M. Haghani, M.C.J. Bliemer, D.A. Hensher, The landscape of econometric discrete choice modelling research, *J. Choice Model.* 40 (2021) 100303, <http://dx.doi.org/10.1016/j.jocm.2021.100303>.
- [7] M. Ben-Akiva, J. Walker, A.T. Bernardino, D.A. Gopinath, T. Morikawa, A. Polydoropoulou, Integration of choice and latent variable models, in: H.S. Mahmassani (Ed.), *Travel Behaviour Research Opportunities and Application Challenges*, Pergamon, Oxford, 2002, pp. 431–470, <http://dx.doi.org/10.1016/b978-008044044-6/50022-x>.
- [8] J. Walker, M. Ben-Akiva, Generalized random utility model, *Math. Social Sci.* 43 (2002) 303–343, [http://dx.doi.org/10.1016/s0165-4896\(02\)00023-9](http://dx.doi.org/10.1016/s0165-4896(02)00023-9).
- [9] S. Raveau, R. Álvarez-Daziano, M.F. Yáñez, D. Bolduc, J. de Dios Ortúzar, Sequential and simultaneous estimation of hybrid discrete choice models: Some new findings, *Transp. Res. Rec.* 2156 (2010) 131–139, <http://dx.doi.org/10.3141/2156-15>.
- [10] J.L. Walker, *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables* (Ph.D. dissertation), Massachusetts Institute of Technology, Cambridge, MA, 2001, <http://dx.doi.org/10.2139/ssrn.4584802>.
- [11] D. Bolduc, M. Ben-Akiva, J. Walker, A. Michaud, Hybrid choice models with logit kernel: Applicability to large scale models, in: M.E.H. Lee-Gosselin, S.T. Doherty (Eds.), *Integrated Land-Use and Transportation Models*, Oxford, 2005, pp. 275–302, <http://dx.doi.org/10.1108/9781786359520-012>.
- [12] D. Bolduc, N. Boucher, R. Alvarez-Daziano, Hybrid choice modeling of new technologies for car choice in Canada, *Transp. Res. Rec.* 2082 (2008) 63–71, <http://dx.doi.org/10.3141/2082-08>.
- [13] P. Mariel, J. Meyerhoff, S. Hess, Heterogeneous preferences toward landscape externalities of wind turbines—combining choices and attitudes in a hybrid model, *Renew. Sustain. Energy Rev.* 41 (2015) 647–657, <http://dx.doi.org/10.1016/j.rser.2014.08.074>.
- [14] V. O'Neill, S. Hess, D. Campbell, A question of taste: Recognising the role of latent preferences and attitudes in analysing food choices, *Food Qual. Prefer.* 32 (2014) 299–310, <http://dx.doi.org/10.1016/j.foodqual.2013.10.003>.
- [15] M. Abou-Zeid, M. Ben-Akiva, Hybrid choice models, in: S. Hess, A. Daly (Eds.), *Handbook of Choice Modelling*, second ed., Edward Elgar Publishing, Cheltenham, 2024, pp. 489–521, <http://dx.doi.org/10.4337/9781800375635.00026>.
- [16] W. Kriswardhana, D. Esztergár-Kiss, The role of intermodality and environmental consciousness in the preferences for maas bundles: A hybrid choice modeling approach, *Transp. Res. Part A* 191 (2025) 104332, <http://dx.doi.org/10.1016/j.tra.2024.104332>.
- [17] X. Fu, Z. Juan, Estimation of multinomial probit-kernel integrated choice and latent variable model: Comparison on one sequential and two simultaneous approaches, *Transportation* 44 (2017) 91–116, <http://dx.doi.org/10.1007/s11116-015-9626-x>.
- [18] K. Sohn, An expectation–maximization algorithm to estimate the integrated choice and latent variable model, *Transp. Sci.* 51 (2017) 946–967, <http://dx.doi.org/10.1287/trsc.2016.0696>.
- [19] P. Mariel, A. Artabe, U. Liebe, J. Meyerhoff, An assessment of the current use of hybrid choice models in environmental economics, and considerations for future applications, *J. Choice Model.* 53 (2024) 100520, <http://dx.doi.org/10.1016/j.jocm.2024.100520>.
- [20] T. Hillel, M. Bierlaire, M.Z. Elshafie, Y. Jin, A systematic review of machine learning classification methodologies for modeling passenger mode choice, *J. Choice Model.* 38 (2021) 100221, <http://dx.doi.org/10.1016/j.jocm.2020.100221>.
- [21] A. Mahpour, A. Baghestani, A. Mamdoohi, An exploration of heterogeneity in latent psychological variables on travelers' destination choice, *Decis. Anal. J.* 6 (2023) 100179, <http://dx.doi.org/10.1016/j.dajour.2023.100179>.
- [22] S.L.C. Cao, D. Nock, A. Davis, An analysis of machine learning approaches for enhancing decision-making in complex discrete choice tasks, *Decis. Anal. J.* 14 (2026) 100668, <http://dx.doi.org/10.1016/j.dajour.2025.100668>.
- [23] P. Chahajer, M. Shah, A. Kshirsagar, The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction, *Decis. Anal. J.* 2 (2022) 100015, <http://dx.doi.org/10.1016/j.dajour.2021.100015>.
- [24] N. Aydin, A.O. Kuşakcı, M. Deveci, The impacts of COVID-19 on travel behavior and initial perception of public transport measures in Istanbul, *Decis. Anal. J.* 2 (2022) 100029, <http://dx.doi.org/10.1016/j.dajour.2022.100029>.
- [25] T. Palit, A.B.M.M. Bari, C.L. Karmaker, An integrated principal component analysis and interpretive structural modeling approach for electric vehicle adoption decisions in sustainable transportation systems, *Decis. Anal. J.* 4 (2022) 100119, <http://dx.doi.org/10.1016/j.dajour.2022.100119>.
- [26] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *BMJ* 372 (n71) (2021) <http://dx.doi.org/10.1136/bmj.n71>.
- [27] F.L. Mannering, C.R. Bhat, Analytic methods in accident research: Methodological frontier and future directions, *Anal. Methods Accid. Res.* 1 (2014) 1–22, <http://dx.doi.org/10.1016/j.amar.2013.09.001>.
- [28] S. Wang, J. Fan, D. Zhao, S. Yang, Y. Fu, Predicting consumers' intention to adopt hybrid electric vehicles: Using an extended version of the theory of planned behavior model, *Transportation* 43 (2016) 123–143, <http://dx.doi.org/10.1007/s11116-014-9567-9>.
- [29] S. Hess, D. Palma, Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application, *J. Choice Model.* 32 (2019) 100170, <http://dx.doi.org/10.1016/j.jocm.2019.100170>.
- [30] F. Liao, E. Molin, B. Van Wee, Consumer preferences for electric vehicles: A literature review, *Transp. Res. Part D* 37 (2017) 252–275, <http://dx.doi.org/10.1080/01441647.2016.1230794>.
- [31] A.C. Mersky, F. Sprei, C. Samaras, Z.S. Qian, Effectiveness of incentives on electric vehicle adoption in Norway, *Transp. Res. Part D* 46 (2016) 56–68, <http://dx.doi.org/10.1016/j.trd.2016.03.011>.
- [32] A.F. Jensen, E. Cherchi, S.L. Mabit, On the stability of preferences and attitudes before and after experiencing an electric vehicle, *Transp. Res. Part D* 25 (2013) 24–32, <http://dx.doi.org/10.1016/j.trd.2013.07.006>.
- [33] F. Alemi, G. Circella, S. Handy, P. Mokhtarian, What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California, *Travel. Behav. Soc.* 13 (2018) 88–104, <http://dx.doi.org/10.1016/j.tbs.2018.06.002>.
- [34] A. Hackbarth, R. Madlener, Consumer preferences for alternative fuel vehicles: A discrete choice analysis, *Transp. Res. Part D* 25 (2013) 5–17, <http://dx.doi.org/10.1016/j.trd.2013.07.002>.
- [35] J.F. Hair, C.M. Ringle, S.P. Gudergan, A. Fischer, C. Nitel, C. Menicats, Partial least squares structural equation modeling-based discrete choice modeling: An illustration in modeling retailer choice, *Bus. Res.* 12 (2019) 115–142, <http://dx.doi.org/10.1007/s40685-018-0072-4>.
- [36] L.A. Garrow, B.J. German, C.E. Leonard, Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research, *Transp. Res. Part C* 132 (2021) 103377, <http://dx.doi.org/10.1016/j.trc.2021.103377>.
- [37] M.J. Eppstein, D.K. Grover, J.S. Marshall, D.M. Rizzo, An agent-based model to study market penetration of plug-in hybrid electric vehicles, *Energy Policy* 39 (2011) 3789–3802, <http://dx.doi.org/10.1016/j.enpol.2011.04.007>.
- [38] F. Alemi, G. Circella, P. Mokhtarian, S. Handy, Exploring the latent constructs behind the use of ridehailing in California, *J. Choice Model.* 29 (2018) 47–62, <http://dx.doi.org/10.1016/j.jocm.2018.08.003>.
- [39] W. Li, M. Kamargianni, An integrated choice and latent variable model to explore the influence of attitudinal and perceptual factors on shared mobility choices and their value of time estimation, *Transp. Sci.* 54 (2020) 62–83, <http://dx.doi.org/10.1287/trsc.2019.0933>.
- [40] G. Oeschger, P. Carroll, B. Caulfield, Micromobility and public transport integration: The current state of knowledge, *Transp. Res. Part D* 89 (2020) 102628, <http://dx.doi.org/10.1016/j.trd.2020.102628>.
- [41] M. Paulssen, D. Temme, A. Vij, J.L. Walker, Values, attitudes and travel behavior: A hierarchical latent variable mixed logit model of travel mode choice, *Transportation* 41 (2014) 873–888, <http://dx.doi.org/10.1007/s11116-013-9504-3>.
- [42] K. Shalender, N. Sharma, Using extended theory of planned behaviour (TPB) to predict adoption intention of electric vehicles in India, *Env. Dev. Sustain.* 23 (2021) 665–681, <http://dx.doi.org/10.1007/s10668-020-00602-7>.
- [43] K. Bhattarai, S. Zhao, E. Bachir Diop, D. Zhang, M. Joshi, Investigating work commuters mode choice behavior in Dalian city during pandemic: A hybrid choice model approach, *Transp. Plann. Technol.* 48 (2024) 1047–1069, <http://dx.doi.org/10.1080/03081060.2024.2410317>.
- [44] F. Tinessa, C. Román García, F. Simonelli, A. Papola, F. Pagliara, How public transport users would react to different pandemic alert scenarios in the post-vaccine era? An analysis of preferences and attitudes of the users in the metropolitan area of Naples (Italy), *Transp. Res. Part A* 190 (2024) 104301, <http://dx.doi.org/10.1016/j.tra.2024.104301>.
- [45] M. Scorrano, T.A. Mathisen, R. Danielis, O. Simsekoglu, G. Marinelli, Stated car choices in Norway and Italy: a comparison based on the integrated choice and latent variable model, *Res. Transp. Econ* 115 (2026) 101695, <http://dx.doi.org/10.1016/j.retrec.2025.101695>.
- [46] K. Lindberg, S. Klain, M. Pignatari, A.J. Caplan, Diverse coastal carbon sequestration preferences and their hierarchical predictors: Three-phase development of a hybrid choice model, *J. Env. Econ. Policy* (2026) 1–19, <http://dx.doi.org/10.1080/21606544.2026.2667147>.
- [47] Y.N. Sang, H.A. Bekhet, Modeling electric vehicle usage intentions: An empirical study in Malaysia, *J. Clean. Prod.* 92 (2015) 75–83, <http://dx.doi.org/10.1016/j.jclepro.2014.12.045>.

- [48] V. Singh, V. Singh, S. Vaibhav, A review and simple meta-analysis of factors influencing adoption of electric vehicles, *Transp. Res. Part D* 86 (2020) 102436, <http://dx.doi.org/10.1016/j.trd.2020.102436>.
- [49] R.A. Daziano, D. Bolduc, Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model, *Transp. A* 9 (2013) 74–106, <http://dx.doi.org/10.1080/18128602.2010.524173>.
- [50] Y. Wang, S. Wang, J. Wang, J. Wei, C. Wang, An empirical study of consumers' intention to use ride-sharing services: Using an extended technology acceptance model, *Transportation* 47 (2020) 397–415, <http://dx.doi.org/10.1007/s11116-018-9893-4>.
- [51] A. Jameel, A.A. Abdulhussein, Incorporating latent variables into active transportation mode choice: A review, *Al-Rafidain J. Eng. Sci.* 3 (2024) 1–18, <http://dx.doi.org/10.61268/vhyhs009>.
- [52] M. Kroesen, S. Handy, C. Chorus, Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling, *Transp. Res. Part A* 101 (2017) 190–202, <http://dx.doi.org/10.1016/j.tra.2017.05.013>.
- [53] Y. Liu, P. Loa, K. Wang, K.N. Habib, Theory-driven or data-driven? Modelling ride-sourcing mode choices using integrated choice and latent variable model and multi-task learning deep neural networks, *J. Choice Model.* 48 (2023) 100431, <http://dx.doi.org/10.1016/j.jocm.2023.100431>.
- [54] A. Glerum, B. Atasoy, M. Bierlaire, Using semi-open questions to integrate perceptions in choice models, *J. Choice Model.* 10 (2014) 11–33, <http://dx.doi.org/10.1016/j.jocm.2013.12.001>.
- [55] P. Tsoleridis, C.F. Choudhury, S. Hess, Using probabilistic clustering techniques as a specification tool for capturing heterogeneity in choice models, *Transp. Res. Part C* 179 (2025) 105289, <http://dx.doi.org/10.1016/j.trc.2025.105289>.
- [56] E. Salazar, P. Salas, P. Sáez, Deep SHAP explanations for predicted choice probabilities from learning multinomial logit model in modal choice application, *Qual. Quant.* 60 (2026) 887–899, <http://dx.doi.org/10.1007/s11135-025-02275-8>.
- [57] F. Rodrigues, R. Krueger, F.C. Pereira, Model building, inference and interpretation: Developing discrete choice models in the age of machine learning, in: S. Hess, A. Daly (Eds.), *HandBook of Choice Modelling*, second ed., Edward Elgar Publishing, Cheltenham, 2024, pp. 74–115, <http://dx.doi.org/10.4337/9781800375635.00009>.
- [58] A. Daly, S. Hess, B. Patruni, D. Potoglou, C. Rohr, Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behaviour, *Transportation* 39 (2012) 267–297, <http://dx.doi.org/10.1007/s11116-011-9351-z>.
- [59] A. Vij, J.L. Walker, How, when and why integrated choice and latent variable models are latently useful, *Transp. Res. Part B* 90 (2016) 192–217, <http://dx.doi.org/10.1016/j.trb.2016.04.021>.
- [60] S. Hess, N. Beharry-Borg, Accounting for latent attitudes in willingness-to-pay studies: The case of coastal water quality improvements in tobago, *Env. Resour. Econ.* 52 (2012) 109–131, <http://dx.doi.org/10.1007/s10640-011-9522-6>.
- [61] M.E. Kløjgaard, S. Hess, Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: Testing the benefits of a hybrid choice model approach, *Soc. Sci. Med.* 114 (2014) 138–150, <http://dx.doi.org/10.1016/j.socscimed.2014.05.058>.
- [62] C.G. Chorus, M. Kroesen, On the (im-) possibility of deriving transport policy implications from hybrid choice models, *Transp. Policy* 36 (2014) 217–222, <http://dx.doi.org/10.1016/j.tranpol.2014.09.001>.
- [63] T. Dekker, S. Hess, T. Arentze, C. Chorus, Incorporating needs-satisfaction in a discrete choice model of leisure activities, *J. Transp. Geogr.* 38 (2014) 66–74, <http://dx.doi.org/10.1016/j.jtrangeo.2014.05.015>.
- [64] S. Hess, A. Stathopoulos, A mixed random utility—random regret model linking the choice of decision rule to latent character traits, *J. Choice Model.* 9 (2013) 27–38, <http://dx.doi.org/10.1016/j.jocm.2013.12.005>.
- [65] O. Benjdiya, N. Rouky, O. Benmoussa, M. Fri, On the use of machine learning techniques and discrete choice models in mode choice analysis, *Logforum* 19 (2023) 331–345, <http://dx.doi.org/10.17270/j.log.2023.845>.
- [66] R. Zhang, J. Wang, C. Liu, K. Su, H. Ishibuchi, Y. Jin, Synergistic integration of metaheuristics and machine learning: Latest advances and emerging trends, *Artif. Intell. Rev.* 58 (2025) <http://dx.doi.org/10.1007/s10462-025-11266-y>.
- [67] M. Abou-Zeid, M. Ben-Akiva, The effect of social comparisons on commute well-being, *Transp. Res. Part A* 45 (2011) 345–361, <http://dx.doi.org/10.1016/j.tra.2011.01.011>.
- [68] F.J. Bahamonde-Birke, U. Kunert, H. Link, J.D. de Dios Ortúzar, About attitudes and perceptions: Finding the proper way to consider latent variables in discrete choice models, *Transportation* 44 (2017) 475–493, <http://dx.doi.org/10.1007/s11116-015-9663-5>.