Journal of Classification (2022) 39:218–239
https://doi.org/10.1007/s00357-021-09398-1

The Spatial Representation of Consumer Dispersion Patterns via a New Multi-level Latent Class Methodology

Sunghoon Kim[1] · Ashley Stadler Blank[2] · Wayne S. DeSarbo[3] · Jeroen K. Vermunt[4]

Accepted: 13 August 2021 / Published online: 8 September 2021
© The Classification Society 2021

Abstract
Consumer dispersion analysis divides aggregate markets into smaller geographic units that marketers can target with their promotional mix. However, dispersion patterns are not always contiguous. Using survey data from National Football League (NFL) fans, we introduce a new hierarchical expectation-maximization (EM) bi-level clustering model that iteratively classifies both teams and fans (nested within teams) based on the spatial heterogeneity of fans in terms of both distance and direction. The proposed multi-level latent class model with a variable number of classes at the lower level outperforms benchmark models in a Monte Carlo simulation study and points to three non-contiguous team segments with a varying number of fan group vectors in the NFL application. We present these results in two-dimensional consumer dispersion maps and report corresponding differences in consumer behavior.

Keywords Multi-level clustering, Spatial heterogeneity, Latent class analysis, Geographic segmentation, Consumer dispersion

1 Introduction

We define consumer dispersion as the geographic distribution pattern of consumers of a given product or service. One of the earliest manifestations of this concept stems from past research that employs geographic segmentation to divide an aggregate consumer market into smaller

✉ Wayne S. DeSarbo
desarbows@aol.com

Sunghoon Kim
sunghoonkim.rutgers@gmail.com

Ashley Stadler Blank
stadlera@xavier.edu

Jeroen K. Vermunt
j.k.vermunt@tilburguniversity.edu

Extended author information available on the last page of the article

distinct geographic units—such as nations, states, regions, counties, cities, zip codes, or neighborhoods—that marketers can target with various elements of the promotional mix (e.g., advertising). More recent approaches combine geographic data with demographic and behavioral data to yield richer descriptions of consumers in each segment (e.g., PRIZM by Claritas). Importantly, past research demonstrates that market segments do not need to be geographically contiguous and may display geographic heterogeneity within each segment (e.g., a mixture of disparate zip codes; see Chang et al. 2020).

Given the size and importance of the sport industry, we examine consumer dispersion in a sport marketing context. The sport industry (which includes gate revenues, media rights, sponsorship, and merchandising) (PwC 2018) is growing at a faster rate than the global GDP (Collignon n.d.), and the estimate for the global sport industry is about $1.3 trillion ($540 billion in the US sport industry alone; Plunkett Research Ltd 2018). Interestingly, although fans often move away from their hometown, many continue to root for their hometown teams. In fact, more than half of all fans of major professional leagues in the USA do not live in the state where their favorite team plays (Fain 2013). Here, we examine consumer dispersion in the National Football League (NFL), which has the most displaced fans across major professional leagues in the USA (Fain 2013). Headlines such as "Why 'America's Team' home games are dominated by visiting team fans" (Tuttle 2014) and "Visiting fans are taking over Chargers' stadium" (Joseph 2018) provide preliminary model-free evidence of sizeable consumer dispersion in the NFL.

However, we are not aware of any existing models that can represent consumer dispersion patterns both within and outside of a team's local market. In order to do so, such models must be able to account for the relationship between teams and fans, as fans are nested within their favorite teams. This presents a challenge for existing classification methods which typically only segment/group at a single level. Segmenting at the team level (and aggregating over fans) ignores the potential heterogeneity across fans, and segmenting at the fan level (and aggregating over teams) ignores the nested structure of fans within teams. In either case, failure to account for both levels can provide poor results and potentially misleading managerial insights. Therefore, we introduce a new multi-level latent class model (i.e., the MLLC-VC) that accommodates a variable number (and nature) of classes at the lower level that accounts for the nested relationship between teams and fans in a flexible manner. Although the lower-level model in our empirical application takes the form of a bivariate normal mixture model, researchers can utilize the proposed MLLC-VC framework with any distributional form of latent class model.[1]

In our NFL data, the MLLC-VC segments both teams (higher-level segments) and fans nested within teams (lower-level group vectors) based on the spatial heterogeneity of fans and extends existing multi-level latent class models in two ways. First, although existing multi-level latent class models (e.g., Bijmolt et al. 2004; Vermunt 2003, 2008) can accommodate

[1] We focus on latent class models and do not consider clustering techniques. Comparing latent class models to non-latent class clustering techniques presents a methodological challenge for the following reasons: 1) we are not aware of any clustering techniques designed to accommodate bi-level data; 2) we are not aware of any past research to inform which clustering techniques would provide similar versus different results for bi-level data; 3) past research documents the fact that different clustering techniques can provide different results, which makes it difficult to select an appropriate clustering technique; and 4) while we can specify distributional forms for latent class models, clustering techniques do not allow us to do so—so we lack comparable model selection criteria, which makes it difficult to select the optimal number of clusters or directly compare results to a latent class model.

multi-level data, they restrict the fan group vectors (lower-level classes) to be the identical across all team segments (higher-level classes); that is, team segments are restricted to differ only in the fan group vector proportions. In contrast, our proposed MLLC-VC is much more flexible and can accommodate a variable number and nature of fan group vectors for each team segment. For example, existing multi-level latent class models may specify three team segments that all have the same number of fan group vectors, whereas the MLLC-VC may (optimally) derive three team segments with two, three, and five fan group vectors, respectively. Second, while existing multi-level latent class models provide a single model selection criterion value (e.g., one Bayesian information criterion (*BIC*) value; Schwarz 1978) for each team segment and fan group vector combination, the proposed model iteratively calculates model selection criteria at both the team segment and fan group vector level to determine the optimal number of team segments and fan group vectors within each team segment. The ability to apply a classification expectation-maximization (CEM) algorithm (Celeux and Govaert 1992) at a higher-level (team segments) as well as a standard expectation-maximization (EM) algorithm (Dempster et al. 1977) at a lower-level (fan group vectors) within each higher-level (team) segment (Fraley and Raftery 1998) is methodologically new and yields an algorithmic approach that allows researchers to estimate and select models with a different number of fan group vectors within each team segment.

Applying the MLLC-VC to our NFL data, we find that the proposed model outperforms existing benchmark models (we also replicate this finding in a Monte Carlo simulation study; see Supplementary Information). In our empirical application, results point to three team segments with five fan group vectors in Team Segment 1, seven fan group vectors in Team Segment 2, and six fan group vectors in Team Segment 3. Fan group vectors within each segment differ in terms of both distance and direction. We provide two-dimensional consumer dispersion maps for each team segment to illustrate the nature of consumer dispersion in the NFL. We also identify differences in consumer behavior between local and non-local markets within each team segment. Next, we discuss consumer dispersion, review existing multi-level latent class models, describe the technical details of the new MLLC-VC model, apply the MLLC-VC to survey data from NFL fans across the USA, compare results to existing benchmark models, and discuss limitations and opportunities for future research.

2 Sport Consumer Dispersion

Consumer dispersion is relevant in the sport industry, particularly in the USA where "74 percent of NFL fans, 69 percent of NBA fans, 67 percent of NCAA fans, 63 percent of MLB fans, and 54 percent of NHL fans root for teams that do not play in the state where they reside" (Fain 2013). Related to sport consumer dispersion is the notion of sport fan diaspora—the "primarily voluntary dispersion of sport fans to other geographical locations, due to better economic (e.g., job acceptance) or educational (e.g., attending college) opportunities, lifecycle changes (e.g., marriage, kids, retirement), health-related reasons (e.g., to seek medical care or a change in climate), international migration, etc." (DeSarbo et al. 2017). The size of sport fan diaspora is quite large, with up to 40–80% of fans attending games to watch the away (vs. home) team play (Joseph 2018; Tuttle 2014). However, many teams fail to recognize their value or engage distant fans in meaningful ways (Fain 2013) by making geographic assumptions that limit distant fans' ability to engage with the team (Stanfill and Valdivia 2017). Consequently, industry insiders urge teams to redefine fan value to include overall investment in and interaction with the team (Dwyer et al. 2015; Fain 2013) regardless of geographic location.

Given the prevalence of consumer dispersion in sport, some past research examines this phenomenon. Kraszewski (2008) finds that many distant fans use team bars to connect with home and maintain home identities. Collins et al. (2016) build on past research that argues that geographic location explains why fans identify with and follow teams (Heere and James 2007; Kerr and Emery 2011) and find that social media use, internet streaming use, and hometown identification improve hometown team identification among sport fans. DeSarbo et al. (2017) provide empirical evidence for the spatial heterogeneity of sport fans. Finally, Mazodier et al. (2018) examine how sponsorship affects fans outside of a team's local market and find that separation from the local market enhances sponsorship outcomes (e.g., recall, attitude, word-of-mouth, purchase intentions, and choice) among highly identified fans.

To summarize, a large contingent of sport fans does not reside in the local market where their favorite team plays but still constitute a sizable and potentially lucrative market. Therefore, it is important for teams to identify and segment their consumers to better understand their fan market structure. Given that fans are nested within teams, we should use a segmentation model that preserves the relationship between these two levels of data. Thus, we introduce our MLLC-VC model that accounts for the relationship between two levels of data (e.g., teams and fans) that iteratively segments both teams and fans nested within teams based on the spatial heterogeneity of fans.

3 Existing Classification Models for Bi-level Data

Existing classification models (in a marketing context) typically segment the market by identifying latent groups of customers with homogeneous characteristics within segments and heterogeneous characteristics between segments (Punj and Stewart 1983; Sarstedt and Mooi 2014). Although marketing data typically include more than one level of data (e.g., household sales nested within brands), traditional segmentation and clustering models only classify such data at a single level (e.g., lower-level households or higher-level brands). Similarly, in our empirical application, we have bi-level data where fans are nested within teams. If we utilize single-level traditional classification models, we can only segment at the team or fan level—not both. There are problems associated with such traditional approaches. First, if you segment at the team level (and aggregate over fans), you ignore the potential heterogeneity across fans. Second, if you segment at the fan level, you ignore the nested structure of fans within teams and the team level dependency structure. In either case, failure to account for both levels in bi-level data will likely provide poor results and potentially misleading managerial insights. Hence, we require a classification model that accounts for the relationship between the two levels of data.

In reviewing the classification literature, there are very few methodological options that incorporate multi-level data. One notable exception extends traditional latent class models by proposing a multi-level latent class model (Vermunt 2003, 2008) that accounts for the dependence between individuals within groups (e.g., consumers nested within countries, employees nested within firms, and students nested within schools; Bijmolt et al. 2004; Lukočienė et al. 2010; Mutz and Daniel 2013). This is important given that past research demonstrates how ignoring group-level dependency can lead to problematic classification results (Asparouhov and Muthen 2008; Bacci et al. 2020; Lee et al. 2018; Park and Yu 2016). Although this multi-level latent class model acknowledges the relationship between the two levels of data and accounts for group-level dependency, it does not capture a different number of lower-level group vectors within each higher-level segment, which limits the model's ability to represent the heterogeneous structure of lower-level data (i.e., mixture of vectors) flexibly. In contrast, the proposed MLLC-VC iteratively provides model selection criteria at both levels and allows for a different number of group vectors in each derived segment. The technical details of the proposed MLLC-VC procedure follow.

4 The MLLC-VC Model

We introduce a new multi-level latent class (MLLC) model that accommodates a variable number of latent classes (VC) at the lower level by developing a hierarchical expectation maximization algorithm that iteratively formulates higher-level segments and lower-level group vectors based on the spatial heterogeneity of consumers. Because we use survey data from NFL fans in our empirical application, we explain the model in terms of this specific context (i.e., higher-level team segments and lower-level fan group vectors). Let:

1. $j = 1, \ldots, J$ represents each team.
2. $i(j) = 1, \ldots, I(j)$ represents each fan nested within their favorite team j, and $I(j)$ represents the number of fans whose favorite team is j.
3. $k = 1, \ldots, K$ represents each latent (unknown) team segment, and K represents the number of team segments.
4. $v(k) = 1, \ldots, V(k)$ represents each latent (unknown) fan group vector in team segment k, and $\mathrm{V}(k)$ represents the number of fan group vectors in team segment k.
5. $\underline{X}_{i(j)} = \left(X_{i(j)}^{(1)}, X_{i(j)}^{(2)}\right)$ follow a bivariate normal distribution and represent the coordinates of fan i's location (in terms of both distance in miles and direction in north-south and east-west) from their favorite team j's location (using 5-digit US zip codes, normalized to the origin in a two-dimensional space).

The MLLC-LC procedure estimates the following model parameters for K latent team segments:

1. λ_k = the prior probability (or mixing proportion) for team segment k $(\sum_{k=1}^{K} \lambda_k = 1, 0 < \lambda_k < 1, \forall k)$.
2. $\pi_{\mathrm{v}(k)}$ = the prior probability (or mixing proportion) for fan group vector $\mathrm{v}(k)$ in team segment k $\left(\sum_{\mathrm{v}(k)=1}^{\mathrm{V}(k)} \pi_{\mathrm{v}(k)} = 1, 0 < \pi_{\mathrm{v}(k)} < 1, \forall \mathrm{v}(k)\right)$.
3. $\underline{\theta}_{\mathrm{v}(k)}$ = the mean coordinates of parameters for fan group vector $\mathrm{v}(k)$ in team segment k (a 1×2 vector); $\underline{\theta}_k$ collects the $\mathrm{V}(k)$ latent fan group vector-specific mean coordinates in team segment k.

4. $\underline{\Sigma}_{v(k)}$= a 2 × 2 covariance matrix for v(k) in team segment k; $\underline{\Sigma}_k$ collects the V(k) latent fan group vector-specific 2 × 2 covariance matrices in team segment k.

Using the model parameters and the observed data, we compute the following posterior probabilities:

1. P_{jk} = the posterior probability (or mixing proportion) of team j membership in team segment k, $(0 \leq P_{jk} \leq 1, \sum_{k=1}^{K} P_{jk} = 1, \forall j, k)$
2. $P_{i(j)v(k)}$ = the posterior probability (or mixing proportion) that fan i from team j belongs to fan group vector v(k) given that team j belongs to team segment k $(0 \leq P_{i(j)v(k)} \leq 1, \sum_{v(k)=1}^{V(k)} P_{i(j)v(k)} = 1, \forall i, v(k))$

Note that although we restrict fans to their favorite team's segment, fans within each team segment can be classified into any fan group vector based on spatial dispersion heterogeneity patterns.

Our primary objective is to estimate *both* team segment membership (P_{jk}, using $\underline{\theta}_k$ and $\underline{\Sigma}_k$) and fan group vector memberships within team segments ($P_{i(j)v(k)}$). To do this, we employ a new hierarchical EM algorithm diagrammed in Fig. 1. Given a fixed value of K, we iteratively apply a CEM approach to classify teams into latent team segments (Step I) and a standard EM algorithm to determine the number of latent fan group vectors and their parameters within each team segment (Step II). We repeat this iterative process for different values of K to select the optimal number of team segments.

In Step I (in Fig. 1), we define the incomplete data likelihood function for segmenting teams as:

$$\mathrm{L} = \prod_{j=1}^{J} \sum_{k=1}^{K} \lambda_k \mathrm{L}_{jk}, \tag{1}$$

where the conditional likelihood function for team j in team segment k is:

$$\mathrm{L}_{jk} = \prod_{i(j)=1}^{I(j)} f\left(\underline{X}_{i(j)}; \underline{\theta}_k, \underline{\Sigma}_k\right), k = 1, \ldots, K. \tag{2}$$

Here, $f\left(\underline{X}_{i(j)}; \underline{\theta}_k, \underline{\Sigma}_k\right)$ represents the density function of $\underline{X}_{i(j)}$, given the segment-level parameters $\underline{\theta}_k$ and $\underline{\Sigma}_k$. For the team segmentation step, first we need to obtain P_{jk}, the posterior probabilities for team segments (i.e., team j membership in team segment k), using Bayes' rule (see Step I-a in Fig. 1):

$$P_{jk} = \frac{\lambda_k \mathrm{L}_{jk}}{\sum_{k=1}^{K} \lambda_k \mathrm{L}_{jk}} = \frac{\lambda_k \prod_{i(j)=1}^{I(j)} f\left(\underline{X}_{i(j)}; \underline{\theta}_k, \underline{\Sigma}_k\right)}{\sum_{k=1}^{K} \lambda_k \prod_{i(j)=1}^{I(j)} f\left(\underline{X}_{i(j)}; \underline{\theta}_k, \underline{\Sigma}_k\right)}. \tag{3}$$

Since the density function of $f\left(\underline{X}_{i(j)}; \underline{\theta}_k, \underline{\Sigma}_k\right)$ has the form of a mixture of fan group vector-specific densities:

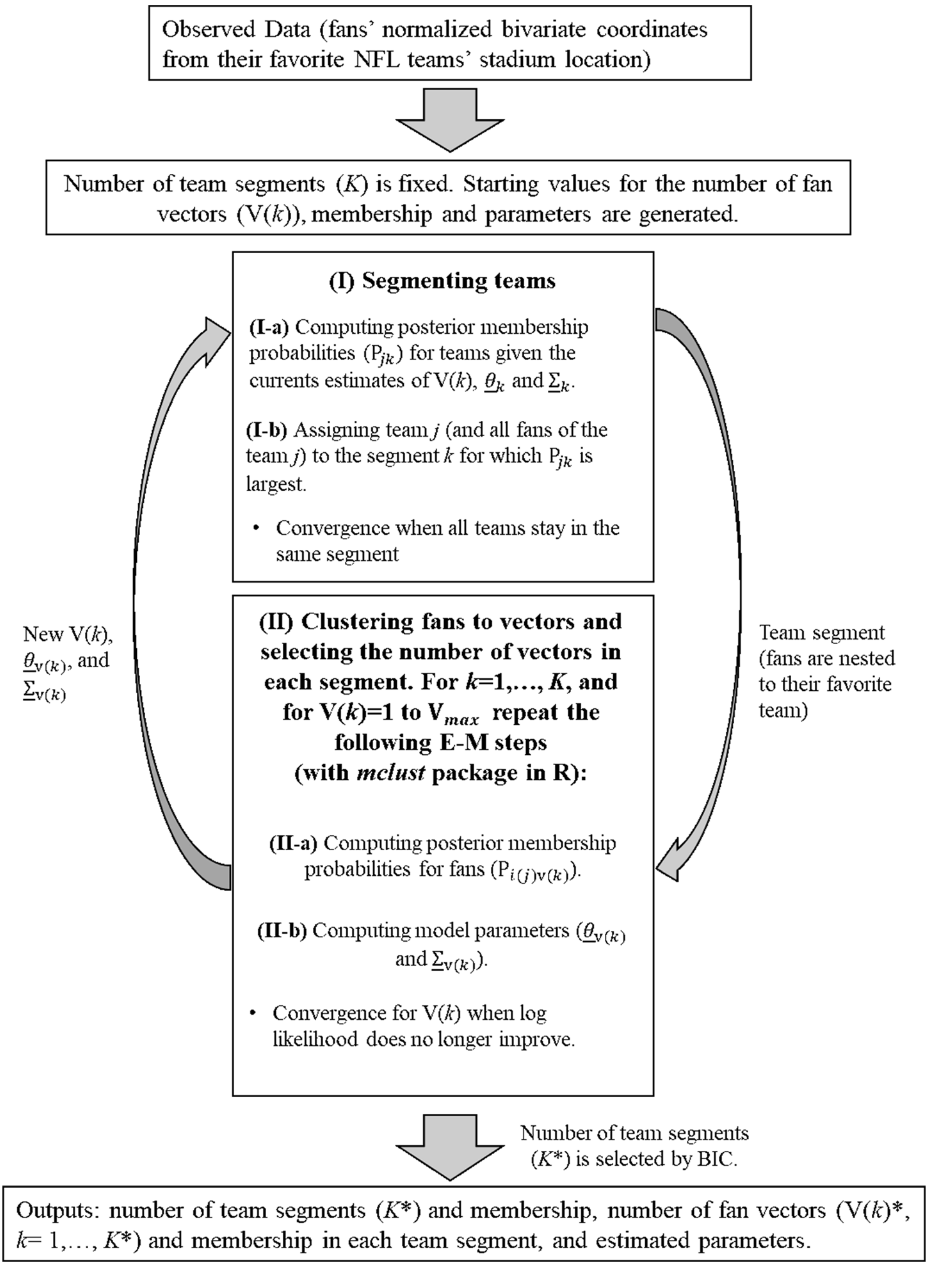


Fig. 1 Hierarchical EM bi-level clustering algorithm. Segmenting teams depends on the EM algorithm for classifying fans to group vectors and corresponding group vector-level parameters $\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$ (see Eqs. (3) and (5))

$$f\left(\underline{X}_{i(j)};\underline{\theta}_k,\underline{\Sigma}_k\right)=\sum_{v(k)=1}^{V(k)}\pi_{v(k)}g\left(\underline{X}_{i(j)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right), \tag{4}$$

We can rewrite Eq. (3) as:

$$P_{jk}=\frac{\lambda_k\prod_{i(j)=1}^{I(j)}\sum_{v(k)=1}^{V(k)}\pi_{v(k)}g\left(\underline{X}_{i(j)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)}{\sum_{k=1}^{K}\lambda_k\prod_{i(j)=1}^{I(j)}\sum_{v(k)=1}^{V(k)}\pi_{v(k)}g\left(\underline{X}_{i(j)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)}. \tag{5}$$

Here, $g\left(\underline{X}_{i(j)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)$ represents the bivariate normal density function of $\underline{X}_{i(j)}$ with parameters $\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$ in fan group vector v(k). So, P_{jk} represents the posterior probability of team j belonging to segment k, and we can interpret segment k as a segment of teams containing fan group vectors of similar dispersion patterns in terms of both distance and direction. Note that in the very first iteration, we need initial values for the number of fan group vectors V(k) in each team segment and for the parameters $\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$. In subsequent iterations, we can use the estimates from Step II, as we describe below. After computing the P_{jk} in Step I-a, the algorithm assigns team j (and all fans of that team) to team segment k for which the current P_{jk} is largest in Step I-b, which implies that we are using a CEM algorithm (Celeux and Govaert 1992) in Step I.

In Step II (in Fig. 1), we determine the number of fan group vectors and corresponding parameters for each team segment by applying a standard EM algorithm for models from 1 to $V_{\max}$ fan group vectors. The likelihood function for segment k is:

$$L_k=\prod_{i(k)=1}^{I(k)}\sum_{v(k)=1}^{V(K)}\pi_{v(k)}L_{iv(k)}, \tag{6}$$

where $i(k)$ indexes fans in team segment k and $I(k)$ is the total number of fans in team segment k. Recall that while fans are restricted to their favorite team's segment, they can be classified into any fan group vector in that segment. The conditional likelihood function for fan i in v(k) is:

$$L_{iv(k)}=g\left(\underline{X}_{i(k)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right). \tag{7}$$

In the E-step of the EM algorithm, we obtain $P_{i(j)v(k)}$, the posterior probability that fan i belongs to fan group vector v(k) given that team j belongs to team segment k, using Bayes' rule (see Step II-a in Fig. 1):

$$P_{i(j)v(k)}=\frac{\pi_{v(k)}g\left(\underline{X}_{i(k)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)}{\sum_{v(k)=1}^{V(k)}\pi_{v(k)}g\left(\underline{X}_{i(k)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)}, v(k)=1,\ldots,V(k). \tag{8}$$

In the M-step, we obtain the fan group vector-level parameters $\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$ by maximizing the expected complete data log-likelihood assuming a bivariate normal distribution for $g\left(\underline{X}_{i(k)};\underline{\theta}_{v(k)},\underline{\Sigma}_{v(k)}\right)$ (see Step II-b in Fig. 1):

$$\underline{\theta}_{v(k)}=\frac{\sum_{i=1}^{N_k}P_{i(j)v(k)}\underline{X}_{i(k)}}{\sum_{i=1}^{N_k}P_{i(j)v(k)}}, \tag{9}$$

where N_k is the number of fans in team segment k and:

$$\underline{\Sigma}_{v(k)} = \frac{\sum_{i=1}^{N_k} P_{i(j)v(k)} \left(\underline{X}_{i(k)} - \underline{\theta}_{v(k)}\right) \left(\underline{X}_{i(k)} - \underline{\theta}_{v(k)}\right)^T}{\sum_{i=1}^{N_k} P_{i(j)v(k)}}. \quad (10)$$

In Step II, we determine the number of latent fan group vectors ($V(k)$) for fans in team segment k in each iteration by running models from 1 to $V_{\max}$ fan group vectors and selecting the best model using the *BIC*. More specifically, we employ the "Mclust" function from the *mclust* package in R, which automatically searches $V(k)$ in team segment k using the *BIC* (see Fraley and Raftery 1998; Scrucca et al. 2016). Note that V(k) can change from one iteration to the next.

To summarize, the proposed hierarchical EM bi-level clustering algorithm segments teams with a CEM approach using P_{jk} (given V(k), $\underline{\theta}_k$, and $\underline{\Sigma}_k$) in Step I of Fig. 1 and groups fans into vectors in each team segment (with $P_{i(j)v(k)}$) with an EM algorithm, updating fan group vector-level parameters ($\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$) in Step II of Fig. 1. Again, the team segment-level parameters ($\underline{\theta}_k$ and $\underline{\Sigma}_k$) automatically update when estimating fan group vector-level parameters ($\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$), as $\underline{\theta}_k$ and $\underline{\Sigma}_k$ contain $\underline{\theta}_{v(k)}$ and $\underline{\Sigma}_{v(k)}$. The algorithm iterates between Step I and Step II until it achieves convergence. Note, we suggest running the algorithm with multiple starts for each K to find the overall best solution for the specified K. We determine the final number of team segments (K*) by running the algorithm for multiple values of K and use the *BIC* as the model selection criterion as in Vermunt (2008) for such multi-level latent class models (see also Fraley and Raftery 1998; Wang and Liu 2006). In the next section, we compare the MLLC-VC performance to several existing benchmark models using our NFL data (see also our Monte Carlo simulation study using synthetic data in the Supplementary Information).

5 Empirical Application: Consumer Dispersion in the National Football League

We examine consumer dispersion in the NFL, which consists of 32 professional football teams in the USA in two conferences (the American Football Conference and the National Football Conference) with four divisions each (north, east, south, and west) (see Table 1). The NFL has the highest fan avidity across domestic professional sports leagues in the USA (Scarborough Sports Marketing 2011) and the most displaced fans at around 50 million (Borrison 2018; Fain 2013), making it an ideal context to study consumer dispersion. Also, consumer dispersion is particularly relevant to the NFL due to recent team relocations. While team relocation may present financial opportunities for teams, moving to a new city displaces an entire market of fans (Kulczycki and Hyatt 2005).

5.1 Study

We collected survey data via Amazon Mechanical Turk (MTurk) in two waves (in the 2015 and 2016 NFL seasons). In exchange for a small financial incentive, a total of 1568 participants completed the survey. In the survey, participants selected their favorite team, rated the extent to which they identify with the team (one item: "I identify with the team," measured on a 7-point scale from 1 = strongly disagree to 7 = strongly agree), and indicated how much

Table 1 Summary statistics for fan distance from NFL teams

American Football Conference		M	SD	Median	National Football Conference		M	SD	Median
North	Baltimore Ravens	380.72	706.45	44.45	North	Chicago Bears	451.10	561.35	127.05
	Cincinnati Bengals	327.09	572.75	61.68		Detroit Lions	263.73	458.82	90.92
	Cleveland Browns	413.28	661.27	113.54		Green Bay Packers	686.14	583.71	664.86
	Pittsburgh Steelers	465.97	641.08	185.94		Minnesota Vikings	474.70	506.79	207.53
East	Buffalo Bills	670.50	770.53	384.79	East	Dallas Cowboys	642.73	703.74	453.82
	Miami Dolphins	584.78	532.77	547.98		New York Giants	435.29	596.56	133.26
	New England Patriots	708.25	837.46	331.15		Philadelphia Eagles	456.72	727.01	66.44
	New York Jets	432.10	710.65	63.38		Washington	327.53	535.98	53.39
South	Houston Texans	144.85	329.64	24.77	South	Atlanta Falcons	355.25	536.74	120.70
	Indianapolis Colts	220.21	395.65	99.49		Carolina Panthers	409.99	793.62	126.20
	Jacksonville Jaguars	152.76	238.63	54.11		New Orleans Saints	360.35	406.10	267.36
	Tennessee Titans	272.36	564.11	101.99		Tampa Bay Buccaneers	449.97	646.17	86.22
West	Denver Broncos	580.43	530.15	548.27	West	Arizona Cardinals	651.67	951.28	100.89
	Kansas City Chiefs	302.10	481.63	117.94		Los Angeles Rams	477.65	838.43	22.62
						St. Louis Rams	477.29	585.34	222.55
	Los Angeles Chargers	367.28	684.40	26.14		San Francisco 49ers	629.72	873.97	104.82
	Oakland Raiders	773.36	877.05	395.04		Seattle Seahawks	703.65	897.35	182.63

M = mean, *SD* = standard deviation. The Chargers played in San Diego for the 2015 and 2016 seasons (i.e., our data collection period) and moved to Los Angeles for the 2017 season. The Rams played in St. Louis for the 2015 season and Los Angeles for the 2016 season

money they spent on the team (on tickets as well as merchandise and media subscriptions). They also provided the 5-digit US zip code where they currently reside, which we used to calculate distance and direction from their favorite team (i.e., the 5-digit US zip code for the stadium where the team plays).

5.2 Aggregate Results

First, we calculate the distance between teams and their fans (see Table 1 for means, standard deviations, and medians). Across all teams, the mean distance is 511 mi, the standard deviation is 687 mi, and the median distance is 165 mi. Mean distances are more than three times median distances, and standard deviations are quite large. This is due to the presence of very large distances (i.e., consumer dispersion) in the dataset. We highlight fan distance (in miles) and direction (north-south and east-west) in an aggregate consumer dispersion map in Fig. 2, where the origin represents the teams' normalized locations and each line (vector) represents a fan. The inner circle represents the 25% quartile (32 mi), the middle circle represents the median (165 mi), and the outer circle represents the 75% quartile (777 mi) (see DeSarbo et al. 2017). Note that 25% of fans live more than 777 mi away from the team (as a comparison point, the driving distance between New York City and Chicago is about 800 mi—more than a 12-h drive). This provides preliminary evidence for the presence of considerable consumer dispersion in the NFL.

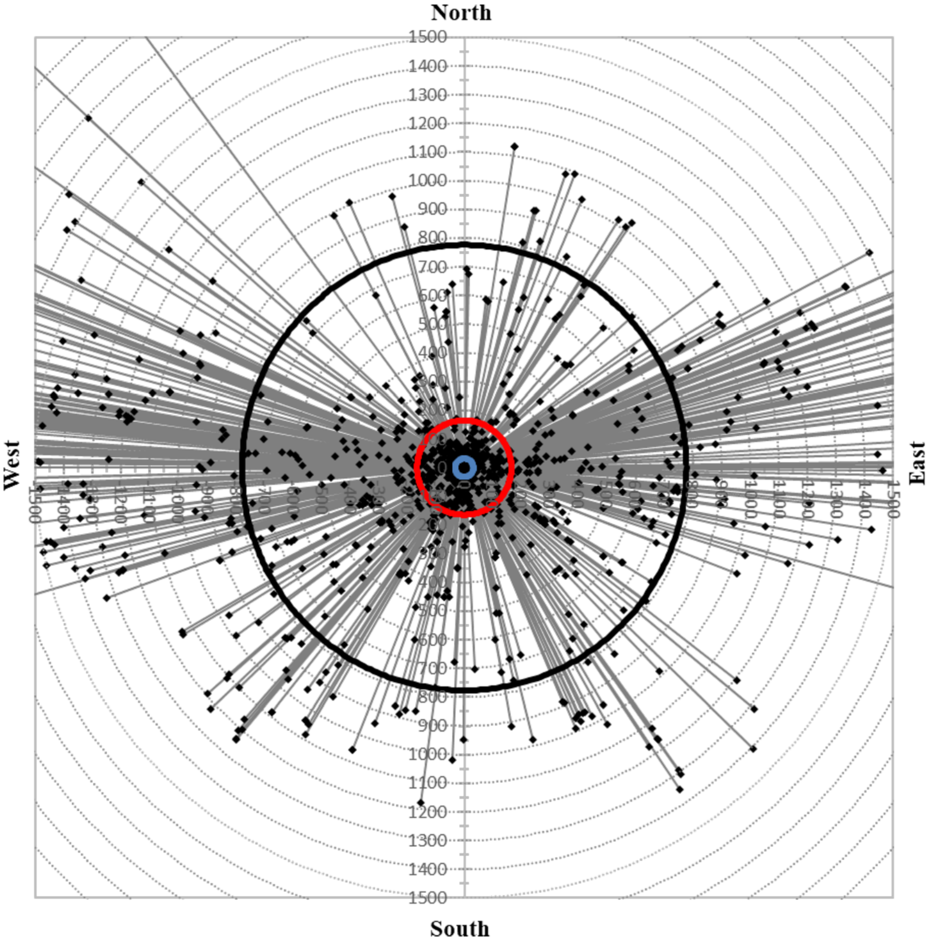


Fig. 2 Aggregate NFL consumer dispersion map. The inner circle represents the 25% quartile distance (32.17 mi), the middle circle represents the median distance (164.93 mi), and the outer circle represents the 75% quartile distance (776.71 mi)

As a simple demonstration of why consumer dispersion is important to marketers, we compare how much money fans who live inside (vs. outside) the median distance (165 mi) from their favorite team spent. First, we find that there is no significant difference in total spend (e.g., tickets, merchandise, media subscription) between the two groups (i.e., \$248 for fans inside the median distance, "inside fans," vs. \$241 for fans outside the median distance, "outside fans," $p > .10$). As one would expect, inside fans spend significantly more (\$123) than outside fans (\$99, $p < .05$) on tickets. But interestingly, outside fans spend significantly more (\$142) than inside fans (\$125, $p < .05$) on merchandise and media subscriptions. This finding indicates that marketers should consider different marketing strategies when targeting fans in local versus non-local markets. In addition, non-local fans are also economically valuable to teams, providing additional support that NFL teams need to rethink the way they understand and target fans. That said, these aggregate results do not provide information on the heterogeneous structure of consumer dispersion across teams. To

Table 2 BIC for the number of team segments for the MLLC-VC

Number of team segments	BIC
$K = 1$	3605.1
$K = 2$	3258.2
$K = 3$	3083.5*
$K = 4$	3191.5

*indicates the best solution (lowest BIC) for the number of team segments

address this limitation, we apply our proposed MLLC-VC procedure which iteratively classifies both teams and fans nested within teams based on the spatial heterogeneity of fans. We initially estimate the number of team segments and fan group vectors and compare MLLC-VC performance to existing benchmark models.

5.3 MLLC-VC Model Selection

Using the *BIC* (i.e., penalized model fit, where lower values indicate better performance; Schwarz 1978), we identify the optimal number of team segments (see Table 2) and fan group vectors for each team segment (see Table 3). The proposed model iteratively searches for the best *BIC* values at both the team level and fan level and points to three team segments (*BIC* = 3083.5) with five fan group vectors in Team Segment 1 (*BIC* = 1850.3), seven fan group vectors in Team Segment 2 (*BIC* = 917.5), and six fan group vectors in Team Segment 3 (*BIC* = 427.6).

5.4 Comparisons to Existing Competitive Benchmark Models

Next, we use the *BIC* to compare MLLC-VC performance to existing benchmark models. For our model comparisons, we can use the team-level *BIC* as a comparison point, as the *BIC* for the number of team segments reflects the *BIC* performance of fan group vectors within each segment.[2]

For our first comparison, we estimate a traditional two-step sequential finite mixture procedure (Fraley and Raftery 2002) that identifies team segments in the first step and fan group vectors in the second step:

1. Step 1: We obtain P_{jk}, the posterior probabilities for team segments (i.e., team j in team segment k), using centroid values of fans in the team using Eq. (3). This model treats fans as replications for the team, which is the same as assuming a single fan group vector in each team segment (i.e., $V(k) = 1$ for all k).
2. Step 2: We obtain $P_{i(j)v(k)}$, the posterior probabilities for fan group vectors (i.e., fan i in fan group vector $v(k)$ given that team j belongs to team segment k) one at a time for each team segment by estimating a mixture model that assumes bivariate normal distributions within fan group vectors.

[2] We can obtain the log-likelihood of the team-level *BIC* for all competing models by summing the log-likelihood value of each team segment that includes fan group vectors.

Table 3 BIC for the number of fan group vectors in each team segment for the MLLC-VC

Number of fan group vectors	*k = 1*	*k = 2*	*k = 3*
V = 1	3991.8	2373.0	2208.7
V = 2	2107.0	1213.5	782.0
V = 3	1864.9	1033.3	531.3
V = 4	1879.1	961.2	501.4
V = 5	1850.3*	933.3	454.9
V = 6	1854.9	941.5	427.6*
V = 7	1861.2	917.5*	432.4
V = 8	1872.5	925.6	433.8

*indicates the best solution (lowest BIC) for the number of fan group vectors in each team segment

The key difference between the MLLC-VC and this two-step sequential finite mixture procedure is that our MLLC-VC estimates team segments and fan group vectors iteratively in Eq. (5). In contrast, this benchmark model estimates team segments first, and fan group vectors within team segments second in a sequential fashion (i.e., Step 1 then Step 2). The *BIC* for the two-step sequential finite mixture procedure (3605.1) points to a one segment solution (see Table 4). Note that the MLLC-VC and the two-step sequential finite mixture procedure have the same *BIC* value for the one segment solution because the number of fan group vectors in a single team segment is the same across both models. Comparing *BIC* values, the MLLC-VC with three team segments (*BIC* = 3083.5) performs much better. We attribute this to the fact that there is substantial spatial heterogeneity among fans within teams in this data. The MLLC-VC accounts for the relationship between teams and fans—segmenting teams and fans nested within teams more parsimoniously, improving the *BIC* (we also replicate this finding in our Monte Carlo simulation study, see Supplementary Information).

Next, we compare the MLLC-VC to a confirmatory approach that fixes the number of team segments to two NFL conferences (AFC and NFC), four NFL divisions (East, West, North, and South), or eight NFL conference divisions (AFC East, West, North, and South and NFC East, West, North, and South) (see Table 1 for team memberships). Comparing *BIC* values (3547.6, 3308.8, and 4773.8, respectively; see Table 4), the MLLC-VC with three team segments (*BIC* = 3083.5) performs better than these three confirmatory (i.e., conference and division structure) models.

Finally, we compare the MLLC-VC to our approximation of the multi-level latent class model in Vermunt (2003, 2008). This approximation of Vermunt (2003, 2008) represents a constrained version of the MLLC-VC, where the number of fan group vectors is the same across all team segments.[3] Thus, to estimate this model, we restrict the number of fan group vectors to be the same across all estimated team segments. We report *BIC* values in Table 5, where the best solution has four team segments and five fan group vectors for all team segments (*BIC* = 3255.5). The MLLC-VC (*BIC* = 3083.5) performs better than this constrained version of the model.

[3] The multi-level latent class model in Vermunt (2003, 2008) does not force hard partitioning of fans nested within teams when segmenting teams. However, given that fans of a team are usually assigned to the same team segment in most applications, adding the hard-partitioning constraint to the model in Vermunt (2003, 2008) makes sense.

Table 4 BIC for existing benchmark models

Number of team segments	Two-step sequential finite mixture procedure	Conference	Division	Conference divisions
$K = 1$	3605.1*			
$K = 2$	3660.0	3547.6		
$K = 3$	3756.6			
$K = 4$	3792.0		3308.8	
$K = 8$				4773.8

*indicates the best solution (lowest BIC) for the number of team segments

5.5 MLLC-VC Results

The *BIC* values for the MLLC-VC point to three team segments (*BIC* = 3083.5) with five fan group vectors in Team Segment 1 (*BIC* = 1850.3), seven fan group vectors in Team Segment 2 (*BIC* = 917.5), and six fan group vectors in Team Segment 3 (*BIC* = 427.6). As shown in Fig. 3, teams in Team Segment 1 are located in the south and west, teams in Team Segment 2 are located in the midwest, and teams in Team Segment 3 are located in the north and east of the USA.

We illustrate the spatial heterogeneity of fans in each team segment and fan group vector in Fig. 4 (where Panel 1 displays the full consumer dispersion map and Panel 2 zooms in to highlight shorter and overlapping fan group vectors). The origin represents the location of the stadium where the teams play and each line (or fan group vector) represents the distance (in miles) and direction (north-south and east-west) of fans in each team segment (see Table 6 for mixture proportions, distance, and direction for each fan group vector). The teams in each segment are similar in terms of the spatial heterogeneity structure of fans (i.e., the radial mixture pattern of fan group vectors).

Team Segment 1 (south/west) includes 697 fans (44% of the sample) and the following 15 teams: Arizona Cardinals, Atlanta Falcons, Dallas Cowboys, Denver Broncos, Houston Texans, Jacksonville Jaguars, Kansas City Chiefs, Los Angeles Rams, Miami Dolphins, New Orleans Saints, Oakland Raiders, San Diego Chargers, San Francisco 49ers, Seattle Seahawks, and Tampa Bay Buccaneers. For this team segment, more than five out of every 10 fans live 292 mi or more away from their favorite team (with one in 10 about 2047 mi away; as a comparison point, the driving distance between Nashville and Los Angeles is about 2000 mi), which is too far to conveniently commute to a home game in 1 day (assuming a 300-mi, 5-h one-way trip—anything above which the Federal Motor Carrier Safety Administration (n.d.) considers questionable for 1-day roundtrip). In Panel A.1 of Fig. 4, fans in vectors

Table 5 BIC for the constrained MLLC-VC

Number of fan group vectors (V) and team segments (K)	V = 1	V = 2	V = 3	V = 4	V = 5	V = 6
$K = 1$	8904.9	4615.3	3824.5	3726.8	3679.3	3642.1
$K = 2$	8540.5	4379.9	3684.7	3389.0	3452.1	3523.3
$K = 3$	8429.7	3995.5	3385.3	3360.3	3331.5	3457.9
$K = 4$	8380.7	4149.9	3455.7	3345.1	3255.5*	3318.6
$K = 5$	8284.8	4047.7	3671.3	3614.2	3316.7	3713.6

*indicates the best solution (lowest BIC) for the number of team segments and fan group vectors

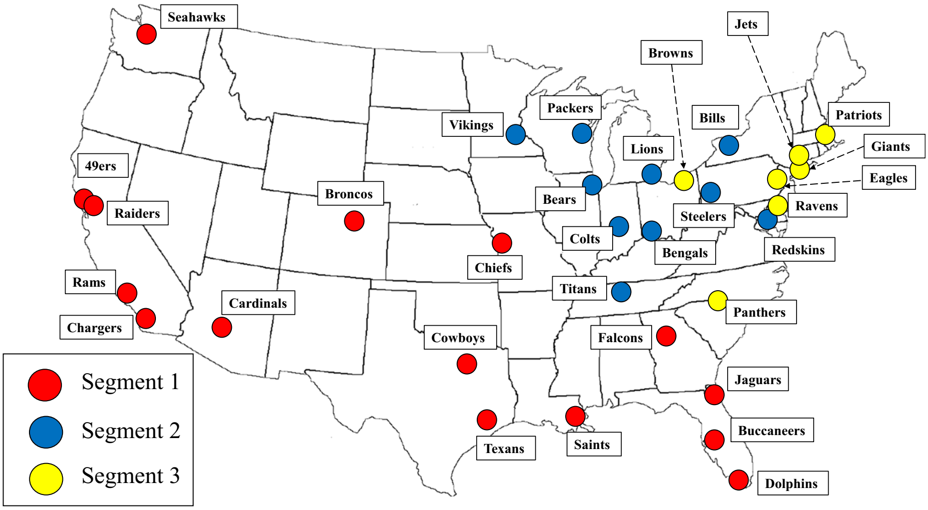


Fig. 3 NFL teams in each team segment

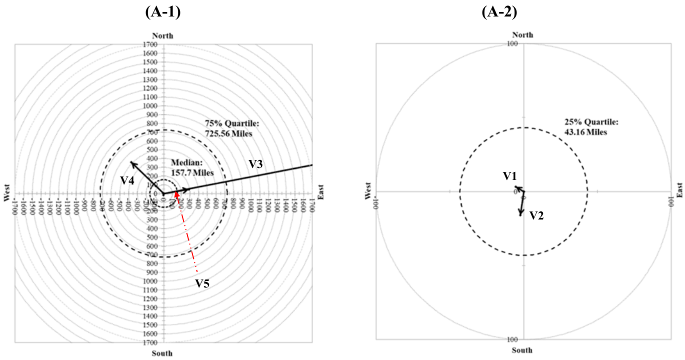
3 (2047 mi) and 5 (292 mi) live to the northeast of their favorite team, while fans in vector 4 (512 mi) live to the northwest. In Panel A.2 of Fig. 4, we zoom in to view shorter and overlapping fan group vectors and observe that fans in vector 1 (6 mi) live to the northwest of their favorite team and fans in vector 2 (16 mi) live to the south of the team. Based on average distances, we label vectors 3, 4, and 5 non-local markets and vectors 1 and 2 local markets.

Team Segment 2 (midwest) includes 469 fans (30% of the sample) and the following 10 teams: Buffalo Bills, Chicago Bears, Cincinnati Bengals, Detroit Lions, Green Bay Packers, Indianapolis Colts, Minnesota Vikings, Pittsburgh Steelers, Tennessee Titans, and Washington. For this segment, four out of every 10 fans live about 423 mi away (with three in 20 about 1224 mi away; as a comparison point, the driving distance between Cleveland and Dallas is about 1200 mi). Overall, 40% of fans live too far away to conveniently commute to a home game in 1 day. In Panel B.1 of Fig. 4, fans in vector 4 (1644 mi) live to the west of their favorite team, fans in vector 3 (1224 mi) live to the southeast of the team, fans in vector 7 (699 mi) live to the east of the team, and fans in vector 5 (423 mi) live to the south of the team. In Panel B.2 of Fig. 4, fans in vector 1 (4 mi) live to the southeast of their favorite team, fans in vector 2 (23 mi) live to the southwest of the team, and fans in vector 6 (99 mi) live to the south of the team. Based on average distances, we label vectors 3, 4, 5, and 7 non-local markets and vectors 1, 2, and 6 local markets.

Team Segment 3 (north/east) includes 402 fans (26% of the sample) and the following seven teams: Baltimore Ravens, Carolina Panthers, Cleveland Browns, New England Patriots, New York Giants, New York Jets, and Philadelphia Eagles. For this segment, four out of every 10 fans live about 557 mi away (with two in 10 at least 1025 mi away; as a comparison point, the driving distance between Chicago and Denver is about 1000 mi). Overall, 40% of fans live too far away to conveniently commute to a home game in 1 day. In Panel C.1 of Fig. 4, fans in vectors 6 (2295 mi) and 1 (557 mi) live to the west of their favorite team, fans in vector 5 (2024 mi) live to the northwest of the team, and fans in vector 3 (1025 mi) live to the southwest of the team. In Panel C.2 of Fig. 4, fans in vector 2 (9 mi) live to the east of their

Diaspora Chart for Latent Vectors across Segments

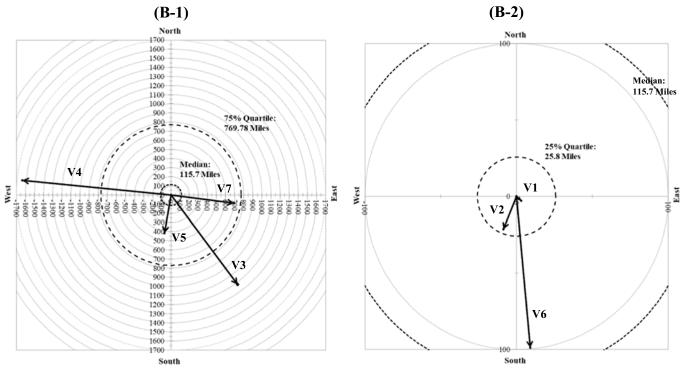
(A) SEGMENT 1



Note, V3 and V4 are not duplicated in this close-up shot but V5 is duplicated for easy understanding.

Diaspora Chart for Latent Vectors across Segments

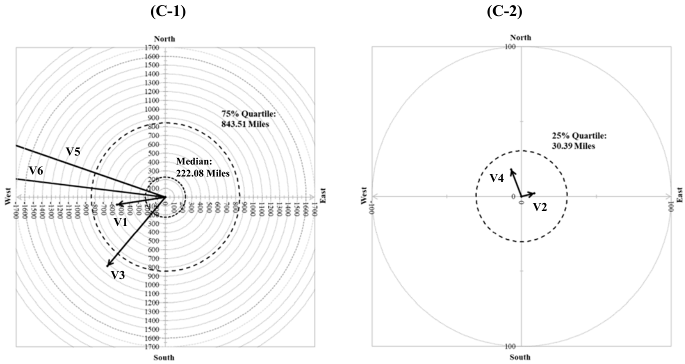
(B) SEGMENT 2



Note, V3, V4, V5 and V7 are not duplicated in this close-up shot

Diaspora Chart for Latent Vectors across Segments

(C) SEGMENT 3



Note, V3, V4 and V6 are not duplicated in this close-up shot but V1 is duplicated for easy understanding.

Fig. 4 NFL consumer dispersion maps for each team segment. We zoom in to view shorter and overlapping fan group vectors in Panels A.2 (V1 and V2), B.2 (V1, V2, and V6), and C.2 (V2 and V4)

favorite team, and fans in vector 4 (19 mi) live to northwest of the team. Based on average distances, we label vectors 1, 3, 5, and 6 non-local markets and vectors 2 and 4 local markets.

These results demonstrate considerable consumer dispersion in the NFL both in terms of team segments and fan group vectors within team segments. Across the three team segments, a large proportion of fans (i.e., 50% of Team Segment 1, 40% of Team Segment 2, and 40% of Team Segment 3) live in locations where it would be difficult to commute to a home game in 1 day. There is also heterogeneity across team segments in that most non-local fans live to the east in Team Segment 1, to the west and south in Team Segment 2, and to the west in Team Segment 3. It makes sense that our model conducts such multi-level segmentation analysis by revealing that geographic location impacts team segments and fan group vectors (as spatial proximity to large bodies of water or international boundaries impact the distance and direction of fan group vectors). Overall, these results provide evidence of sizeable consumer dispersion with significant (non-contiguous) consumer dispersion patterns in terms of both distance and direction across team segments.

5.6 MLLC-VC Post-hoc Analysis

We conduct a post-hoc analysis to explore whether behavioral patterns vary across local and non-local markets (see Table 6 for local and non-local fan group vectors). Recall that

Table 6 Mixture proportion, mean distance, direction, and behavioral patterns for NFL fan group vectors

Segment	Vector	Mixture proportion	Mean distance	Direction	Attend at least one home game	Attend at least one away game	Spend on tickets	Spend on merchandise and media subscriptions
$k = 1$	v = 5*	31%	292	N-E	19%	15%	$83	$122
	v = 1	30%	6	N-W	46%	7%	$158	$133
	v = 2	19%	16	S-W	43%	8%	$128	$127
	v = 4*	11%	512	N-W	12%	18%	$68	$137
	v = 3*	9%	2047	N-E	13%	23%	$103	$197
	Total				31%	12%	$114	$134
$k = 2$	v = 2	32%	23	S-W	36%	10%	$112	$150
	v = 1	26%	4	S-E	50%	10%	$142	$127
	v = 5*	19%	423	S-W	18%	17%	$77	$136
	v = 4*	14%	1644	N-W	14%	16%	$71	$173
	v = 7*	4%	699	S-E	14%	19%	$78	$209
	v = 6	3%	99	S-E	25%	8%	$51	$51
	v = 3*	2%	1224	S-E	40%	50%	$215	$260
	Total				32%	13%	$107	$147
$k = 3$	v = 4	32%	19	N-W	33%	5%	$100	$102
	v = 2	28%	9	N-E	50%	11%	$157	$139
	v = 1*	22%	557	S-W	20%	12%	$88	$113
	v = 3*	7%	1025	S-W	19%	30%	$133	$137
	v = 6*	7%	2295	N-W	10%	14%	$75	$132
	v = 5*	3%	2024	N-W	7%	0%	$50	$71
	Total				31%	11%	$112	$118

*indicates a non-local (vs. local) market.

N = north, S = south, E = east, W = west. Percentages in the last four columns represent the proportion of fans who engaged in the behavior in the previous season participants rated the extent to which they identify with their favorite team (i.e., team identification) and indicated the ways in which they follow the team. We begin with team identification (one item: "I identify with the team," measured on a 7-point scale from 1 = strongly disagree to 7 = strongly agree). We are interested in team identification, as it affects fans' cognitive, affective, and behavioral responses (Wann and Branscombe 1993). Interestingly, fans in non-local markets in team Segments 1 and 3 report higher team identification than fans in local markets (Team Segment 1: $M = 5.42$ vs. 5.29, $p < .10$; Team Segment 3: $M = 5.56$ vs. 5.32, $p < .05$); there is no difference between fans in non-local and local ($M = 5.45$ vs. 5.46, $p > .10$) markets in Team Segment 2. Although one may expect team identification to be lower for non-local markets, it either met or exceeded that of local markets in this study. Consequently, sport marketers should reach out to and engage highly identified fans in non-local markets to improve business outcomes.

We also assessed how much money fans spent on the team. More specifically, we asked fans how much money they spent on tickets as well as merchandise and media subscriptions. In Team Segment 1, 45% of fans in local markets and 16% of fans in non-local markets attended at least one home game ($\chi^2 = 65.4$, $p < .01$), while 7% of fans in local markets and 17% of fans in non-local markets attended at least one away game ($\chi^2 = 15.2$, $p < .01$). For ticket sales, local markets spent more than non-local markets ($146 vs. $83, $p < .05$). However, for merchandise sales and media subscriptions (although not significant), non-local markets spent more than local markets ($138 vs. $131, $p > .10$). These results highlight the need to develop different business strategies when targeting local vs. non-local markets. As an example, 23% of fans in vector 3 (2047 mi) attended an away game, making it the most profitable non-local market for away game attendance; they also spent the most on merchandise and media subscriptions ($197) among all fan group vectors (including local markets)—almost double what they spent on tickets ($103).

In Team Segment 2, 42% of fans in local markets and 17% of fans in non-local markets attended at least one home game ($\chi^2 = 31.1$, $p < .01$), while 10% of fans in local markets and 18% of fans in non-local markets attended at least one away game ($\chi^2 = 7.1$, $p < .01$). For ticket sales, local markets spent more than non-local markets ($122 vs. $83, $p < .05$). However, for merchandise sales and media subscriptions, non-local markets spent about 20% more than local markets ($164 vs. $136, $p < .10$).

In Team Segment 3, 41% of fans in local markets and 17% of fans in non-local markets attended at least one home game ($\chi^2 = 25.2$, $p < .01$), while 8% of fans in local markets and 15% of fans in non-local markets attended at least one away game ($\chi^2 = 3.9$, $p < .05$). For ticket sales, local markets spent more than non-local markets ($127 vs. $90, $p < .05$). For merchandise sales and media subscriptions (although not significant), non-local markets spent more than local markets ($119 vs. $ 117, $p > .10$). Using these results, teams can cater their marketing strategy to provide the right product to the right fan group vector(s) in each team segment.

To summarize results, team identification among non-local markets exceeded that of local markets in Team Segments 1 and 3 and met that of local markets in Team Segment 2. In terms of attendance, non-local markets attend home games less and away games more. In terms of money spent, while non-local markets spent less on tickets than local markets, they spent more on merchandise and media subscriptions in Team Segment 2 and similarly in Team Segments 1 and 3. These results support the notion that "the family that's sitting at home 1,500 miles away from the stadium watching on their iPad" may be just as valuable as a family with season tickets, so "it's important to connect with both" (Fain 2013).

6 Conclusion

Today, firms compete for consumers in a global economy. Therefore, it is essential for firms to identify and geographically segment their consumers to understand their market structure. That said, marketing data is becoming increasingly complex and often requires a segmentation model that preserves the relationship between multiple levels of data. Therefore, we introduce a new multi-level latent class model that accommodates a variable number of classes at the lower-level (the MLLC-VC). The MLLC-VC preserves the relationship between two levels of data and iteratively segments higher-level segments and lower-level group vectors to measure consumer dispersion in terms of both distance and direction. The proposed model can help marketers understand how consumer behavior varies across segments and group vectors nested within segments, so marketers can develop different strategies to target consumers in each segment and group vector more accurately. Indeed, as we see in our empirical application (and Monte Carlo simulation study in the Supplementary Information), the MLLC-VC performs better than existing benchmark models (including a two-step sequential finite mixture procedure, a model using prespecified segments such as NFL conferences or divisions, and a model that constrains the number of vectors to be equal across all segments) and leads to more accurate conclusions. Note that while we modeled two continuous response variables using a bivariate normal mixture model in our empirical application, researchers can also use the MLLC-VC in combination with other types of latent class models (e.g., latent class models for categorical responses or latent class regression models) in a host of other applications.

Because over half of all fans of major professional sports leagues in the USA do not live in the state where their team plays, we apply the MLLC-VC to measure consumer dispersion in the NFL. The MLLC-VC iteratively estimates team segments and fan group vectors within team segments and outperforms several existent benchmark models. The MLLC-VC points to three team segments: Team Segment 1 (south/west) with five fan group vectors, Team Segment 2 (midwest) with seven fan group vectors, and Team Segment 3 (north/east) with six fan group vectors. Interestingly, about 40–50% of fans in each team segment reside in non-local markets, and we provided two-dimensional maps to illustrate consumer dispersion in each team segment (see Fig. 4). We also examine behavioral differences in team identification, game attendance, and money spent (on tickets as well as merchandise and media subscriptions) between local and non-local markets, and our results support that non-local markets may be just as valuable to teams as local markets.

Overall, we find evidence of sizeable consumer dispersion in the NFL and suggest that teams rethink the way they understand and target fans to improve business outcomes. For example, teams could try to enhance attendance at home or away games among non-local markets. To bolster attendance at home games, teams could implement travel programs similar to Red Sox Destinations (Sox 2020), where distant fans can purchase packages that include tickets, hotel accommodations, tours, meet and greets, gift bags/items, photos, etc. For away games, teams could market to fans located in the away market. For example, in college football, visiting teams typically receive an allotment of tickets for away games (e.g., 7000 tickets in the Southeastern Conference; Dosh 2011). Since attendance at away games has decreased dramatically among fans from a team’s local market (Dooley 2018), teams could target non-local fans that reside in the away markets. Similarly, teams can reach out to non-local markets via the Internet and social media to improve online activity. Teams can also develop specific marketing strategies to promote sales of merchandise and media subscriptions in non-local markets. Identifying what segment a team is in as well as the distance and direction of its fan group vectors can help the NFL and its teams accomplish these goals by reaching out to and engaging more fans.

As with all research, there are some limitations that present opportunities for future research. First, we focus on the relationship between fans and their favorite teams in a professional sport context. We agree with scholars who argue that “while sports markets have some idiosyncrasies… there are also many similarities to experience services such as events, entertainment, or traveling and many other industries that have to manage fixed, perishable inventory” (Wetzel et al. 2018, p. 606). For example, spectator sports share the same NAICS code as other arts, entertainment, and recreation businesses (NAICS Association 2018). Thus, theater and dance companies, music artists and bands, museums, historical sites, zoos and botanical gardens, nature parks, amusement parks, casinos, golf courses, and ski resorts (etc.) can also use the MLLC-VC to segment and market to non-local consumers. Other services, such as retail stores, restaurants and breweries, or universities, may also appeal to consumers outside of a local market, especially if they help consumers connect with home (Torelli et al. 2017). We believe future research could examine multi-level data across these (and a variety of other) domains. Second, while we investigate spatial heterogeneity, future research could also assess other types of heterogeneity—for example, at the brand, product, or attribute level. Third, we did not have access to financial data at the team level; however, future research could investigate the economic value or the impact of differential marketing mix decisions of local versus non-local markets (or higher-level segments and lower-level group vectors, more generally). Finally, from a methodological standpoint, future research could extend the MLLC-VC to more generalized (e.g., non-parametric) models or experiment with other types of distributional forms such as wrapped distributions, Von Mises distributions, and Sine-skewed distributions that may be more appropriate for circular/directional data (see Ley and Verdebout 2017). Somewhat related, future research could work to identify or develop an appropriate non-latent class model (e.g., clustering technique) for bi-level data to compare its performance to the MLLC-VC. Ultimately, we hope that the MLLC-VC helps researchers and practitioners generate more meaningful insights and make better decisions with multi-level data across a variety of different applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00357-021-09398-1.

Acknowledgements The authors gratefully acknowledge the financial support from the Smeal College of Business at the Pennsylvania State University. In addition, we wish to thank the Editor, Associate Editor, and anonymous reviewers for their insightful and constructive comments which have improved this manuscript for the Journal of Classification.

Funding This research was supported by the Smeal College of Business at the Pennsylvania State University.

References

Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27–51). Charlotte, NC.

Bacci, S., Bartolucci, F., & Pennoni, F. (2020). Multilevel model based clustering: A new proposal of maximum a-posteriori assignment. In T. Imaizumi, A. Okada, S. Miyamoto, F. Sakaori, Y. Yamamoto, & M. Vichi (Eds.), *Advanced studies in classification and data science* (1st ed., pp. 3–11). New York.

Bijmolt, T. H. A., Paas, L. J., & Vermunt, J. K. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing, 21*(4), 323–340.

Borrison, R. (2018). NFL connects displaced football fans to their teams, Available at https://www.mobilemarketer.com/ex/mobilemarketer/cms/news/strategy/16234.html.

Celeux, G., & Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis, 14*(3), 315–332.

Chang, W., Kim, S., & Chae, H. (2020). A regularized spatial market segmentation method with Dirichlet process—Gaussian mixture prior. *Spatial Statistics, 35*, 1–19.

Collignon, H. (n.d.). Winning in the business of sports, Available at https://www.es.kearney.com/communications-media-technology/article?/a/winning-in-the-business-of-sports.

Collins, D. R., Heere, B., Shapiro, S., Ridinger, L., & Wear, H. (2016). The displaced fan: The importance of new media and community identification for maintaining team identity with your hometown team. *European Sport Management Quarterly, 16*(5), 655–674.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. Series B (Methodological)*, 39*(1), 1–38.

DeSarbo, W.S., Blank, A.S., Kim, S. (2017). Sports diaspora: A national survey of NFL fan dispersion, Available at https://www.sportsbusinessdaily.com/Journal/Issues/2017/10/16/Opinion/Diaspora.aspx.

Dooley, P. (2018). Across the SEC, football fans opting not to travel with their team, Available at https://www.gainesville.com/news/20180714/across-sec-football-fans-opting-not-to-travel-with-their-team.

Dosh, K. (2011). To Texas A&M fans: What to expect in SEC, Available at http://businessofcollegesports.com/2011/08/31/to-texas-am-fans-what-to-expect-in-sec/.

Dwyer, B., Greenhalgh, G. P., & LeCrom, C. W. (2015). Exploring fan behavior: Developing a scale to measure sport efangelism. *Journal of Sport Management, 29*(6), 642–656.

Fain, I. (2013). Displaced fans remain connected – and valuable – to teams, Available at https://www.sportsbusinessdaily.com/Journal/Issues/2013/02/04/Opinion/From-the-Field-of-Fan-Engagement.aspx.

Federal Motor Carrier Safety Administration (n.d.). Section § 392.6: Schedules to conform with speed limits, Available at https://www.fmcsa.dot.gov/regulations/title49/section/392.6.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal, 41*(8), 578–588.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611–631.

Heere, B., & James, J. D. (2007). Sports teams and their communities: Examining the influence of external group identities on team identity. *Journal of Sport Management, 21*(3), 319–337.

Joseph, A. (2018). Visiting fans are taking over Chargers' stadium, Available at https://www.usatoday.com/story/sports/ftw/2018/09/09/visiting-fans-are-still-taking-over-the-los-angeles-chargers-stadium/111296530/.

Kerr, A. K., & Emery, P. R. (2011). Foreign fandom and the Liverpool FC: A cyber-mediated romance. *Soccer and Society, 12*(6), 880–896.

Kraszewski, J. (2008). Pittsburgh in Fort Worth: Football bars, sports television, sports fandom, and the management of home. *Journal of Sport and Social Issues, 32*(2), 139–157.

Kulczycki, C., & Hyatt, C. (2005). Expanding the conceptualization of nostalgia sport tourism: Lessons learned from fans left behind after sport franchise relocation. *Journal of Sport Tourism, 10*(4), 273–293.

Lee, W.-Y., Cho, S.-J., & Sterba, S. K. (2018). Ignoring a multilevel structure in mixture item response models: Impact on parameter recovery and model selection. *Applied Psychological Measurement, 42*(2), 136–154.

Ley, C., & Verdebout, T. (2017). *Modern directional statistics*. CRC Press.

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology, 40*(1), 247–283.

Mazodier, M., Henderson, C. M., & Beck, J. T. (2018). The long reach of sponsorship: How fan isolation and identification jointly shape sponsorship performance. *Journal of Marketing, 82*(6), 28–48.

Mutz, R., & Daniel, H.-D. (2013). University and student segmentation: Multilevel latent-class analysis of students' attitudes towards research methods and statistics. *British Journal of Educational Psychology, 83*(2), 280–304.

NAICS Association (2018). Six digit NAICS codes and titles, Available at https://www.naics.com/six-digit-naics/?code=71.

Park, J., & Yu, H.-T. (2016). The impact of ignoring the level of nesting structure in nonparametric multilevel latent class models. *Educational and Psychological Measurement, 76*(5), 824–847.

Plunkett Research Ltd. (2018). Sports industry statistic and market size overview, business and industry statistics, Available at https://www.plunkettresearch.com/statistics/Industry-Statistics-Sports-Industry-Statistic-and-Market-Size-Overview/.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20*(2), 134–148.

PwC (2018). At the gate and beyond: outlook for the sports market in North America through 2022, Available at https://www.pwc.com/us/en/industries/tmt/library/sports-outlook-north-america.html.
Sarstedt, M., & Mooi, E. (2014). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics* (2nd ed.). Berlin, Germany.
Scarborough Sports Marketing (2011). American consumers root for the underdogs: Gymnastics, Fig. skating, men's golf, high school sports and pro boxing rank high in fan interest, Available at https://www.prnewswire.com/news-releases/american-consumers-root-for-the-underdogs%2D%2Dgymnastics-figure-skating-mens-golf-high-school-sports-and-pro-boxing-rank-high-in-fan-interest-131233464.html.
Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.
Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R Journal, 8*(1), 289–317.
Boston Red Sox (2020). Red Sox destinations, Available at https://www.mlb.com/redsox/tickets/travel.
Stanfill, M., & Valdivia, A. N. (2017). (Dis) locating nations in the World Cup: Football fandom and the global geopolitics of affect. *Social Identities, 23*(1), 104–119.
Torelli, C. J., Ahluwalia, R., Cheng, S. Y. Y., Olson, N. J., & Stoner, J. L. (2017). Redefining home: How cultural distinctiveness affects the malleability of in-group boundaries and brand preferences. *Journal of Consumer Research, 44*(1), 44–61.
Tuttle, B. (2014) Why "America's team" home games are dominated by visiting team fans, Available at https://money.com/dallas-cowboys-jerry-jones-nfl-fans/.
Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*(1), 213–239.
Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research, 17*(1), 33–51.
Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research, 77*(2), 220–225.
Wann, D. L., & Branscombe, N. R. (1993). Sports fans: Measuring degree of identification with their team. *International Journal of Sport Psychology, 24*(1), 1–17.
Wetzel, H. A., Hattula, S., Hammerschmidt, M., & van Heerde, H. J. (2018). Building and leveraging sports brands: Evidence from 50 years of German professional soccer. *Journal of the Academy of Marketing Science, 46*, 591–611.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sunghoon Kim [1] • **Ashley Stadler Blank** [2] • **Wayne S. DeSarbo** [3] • **Jeroen K. Vermunt** [4]

[1] Rutgers Business School, Rutgers University, New Brunswick, NJ 08854, USA

[2] Williams College of Business, Xavier University, Cincinnati, OH 45207, USA

[3] Smeal College of Business, University Park, PA 16802, USA

[4] Tilburg School of Social and Behavioral Sciences, Tilburg University, 5037 DB Tilburg, The Netherlands