ARTICLE IN PRESS
IJRM-01259; No of Pages 17
International Journal of Research in Marketing xxx (2018) xxx–xxx
Contents lists available at ScienceDirect
IJRM
International Journal of Research in Marketing
journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

# A random coefficients mixture hidden Markov model for marketing research

Eelco Kappe [a,*], Ashley Stadler Blank [b], Wayne S. DeSarbo [a]

[a] *Marketing at the Smeal College of Business, Pennsylvania State University, University Park, PA 16802, United States of America*
[b] *Marketing at the Opus College of Business, University of St. Thomas, St. Paul, MN 55105, United States of America*

ARTICLE INFO

*Article history:*
First received on May 8, 2017 and was under review for 6 1/2 months
Available online xxxx

Senior Editor: Bruce G.S. Hardie

*Keywords:*
Hidden Markov model
Time-varying effects
Unobserved heterogeneity
Attendance demand model
Major League Baseball

ABSTRACT

The hidden Markov model (HMM) provides a framework to model the time-varying effects of marketing mix variables. When employed in a panel data context, it is important to properly account for unobserved heterogeneity across individuals. We propose a new random coefficients mixture HMM (RCMHMM) that allows for flexible patterns of unobserved heterogeneity in both the state-dependent and transition parameters. The RCMHMM nests all HMMs found in the marketing literature. Results of two simulation studies demonstrate that 1) averaging across a large number of different data generating processes, the RCMHMM outperforms all its nested versions using both in-sample and out-of-sample performance and 2) the RCMHMM is more robust than its nested versions when underlying model assumptions are violated. In addition, we apply the RCMHMM to an empirical application where we examine the effectiveness of in-game promotions in increasing the short-term demand for Major League Baseball (MLB) attendance. We find that the effectiveness of four promotional categories varies over the course of the season and across teams and that the RCMHMM performs best.


## 1. Introduction

The hidden Markov model (HMM) has made significant inroads into marketing over the past three decades (e.g., Montgomery, Li, Srinivasan, & Liechty, 2004; Netzer, Lattin, & Srinivasan, 2008; Poulsen, 1990). The HMM enables researchers to model the time-varying effects of marketing mix variables via the formulation of unobserved (or hidden) states. The HMM allows for different parameters for the group of observations that belong to each hidden state (state-dependent equation) as well as switching between hidden states over time (transition equation). Recent work has increased the value of the HMM in marketing by incorporating unobserved heterogeneity across individuals (e.g., Montoya, Netzer, & Jedidi, 2010; Schweidel & Knox, 2013), which may be present in both the state-dependent and transition parameters. Unobserved heterogeneity exists in many marketing applications (e.g., panel data) and failure to account for it leads to biased parameters and inaccurate managerial insights (see Netzer, Ebbes, & Bijmolt, 2016 for a discussion of the consequences of ignoring unobserved heterogeneity in the HMM).

While past research incorporates some patterns of unobserved heterogeneity into the HMM (e.g., random coefficients or latent classes), this body of work largely ignores the comparative performance of various alternative forms of unobserved heterogeneity in the state-dependent and transition parameters. An improper representation of the heterogeneity structure can lead to biased inferences (Hsiao, 2003)—especially in dynamic models, such as HMMs, where disentangling unobserved heterogeneity and

* Corresponding author.
*E-mail addresses:* erk11@psu.edu, (E. Kappe), stad8452@stthomas.edu, (A. Stadler Blank), wsd6@psu.edu (W.S. DeSarbo).

https://doi.org/10.1016/j.ijresmar.2018.07.002


Please cite this article as: Kappe, E., et al., A random coefficients mixture hidden Markov model for marketing research, *International Journal of Research in Marketing* (2018), https://doi.org/10.1016/j.ijresmar.2018.07.002

state dependence can be challenging (Heckman, 1991; Hyslop, 1999). If an HMM fails to properly account for unobserved heterogeneity, the effects are picked up by the dynamic elements of the model, which leads to incorrect conclusions about the time-varying effects.

We contribute to the marketing and HMM literatures by proposing a new random coefficients mixture HMM (RCMHMM) that allows for flexible patterns of unobserved heterogeneity in *both* the state-dependent and transition parameters. To achieve this flexibility, we draw on the random coefficients mixture (RCM) model (see Allenby, Arora, & Ginter, 1998; Lenk & DeSarbo, 2000). The RCM model combines the flexibility of semi-parametric latent class models and allows for parametric variation in the parameters within latent classes through the formulation of random coefficients. Incorporating an RCM representation of unobserved heterogeneity in both the state-dependent and transition parameters within an HMM framework makes it possible to nest all HMMs in the marketing literature and compare model performance of the RCMHMM to its 16 nested versions (combining homogenous, random coefficients, latent classes, and individual-specific independent state-dependent and transition parameters). Thus, the RCMHMM enables researchers to identify and select the optimal heterogeneity structure using a single model structure, reducing parameter bias.

We compare the performance of the RCMHMM to its 16 nested versions in two simulation studies. First, we estimate all 17 models on 17 corresponding datasets. Averaging across datasets, the RCMHMM outperforms its nested model versions and often results in a more parsimonious model (requiring fewer latent classes and hidden states to explain the data). The HMM with random coefficients for both the state-dependent and transition parameters (i.e., the state-of-the-art HMM in marketing) performs second-best. Comparing these two models, the RCMHMM performs better when the unobserved heterogeneity structure is multimodal (i.e., the data includes latent classes or an RCM distribution). When the unobserved heterogeneity structure is unimodal, both models have similar performance (as the RCMHMM perfectly nests the HMM with random coefficients for both the state-dependent and transition parameters), and there is no need to formulate the more complex RCMHMM. In general, the additional flexibility of the RCMHMM can be somewhat harmful when there is little unobserved heterogeneity in the data. Second, we examine what happens with different data dimensions and when the underlying model assumptions are violated. RCMHMM performance improves with more time periods and is more robust to different data generating processes; specifically, when an omitted variable is present, or when the error distribution is misspecified. Importantly, these two simulation studies show that using an incorrect form of unobserved heterogeneity biases parameter estimates. The RCMHMM enables researchers to minimize such bias by nesting 16 HMMs and selecting the model that most accurately reflects the unobserved heterogeneity structure in the data.

We also use the RCMHMM and its 16 nested versions to examine the effectiveness of in-game promotions in increasing the short-term demand for Major League Baseball (MLB) attendance. This empirical application is an ideal context, as past attendance research provides preliminary evidence that promotional effectiveness varies over time (Boyd & Krehbiel, 1999; Lemke, Leonard, & Tlhokwane, 2010) (facilitating hidden states) and across teams (Lemke et al., 2010; Marcum & Greenstein, 1985) (facilitating latent classes). We find that promotional effectiveness varies over the course of the season and across teams and that the RCMHMM performs best.

We organize the remainder of the article as follows. First, we review past research on incorporating unobserved heterogeneity into HMMs. Next, we present our RCMHMM along with the estimation procedure and predictive validation techniques to determine the best model. Then, we test the RCMHMM in two simulation studies and an empirical application. Finally, we highlight practical and methodological implications and discuss directions for future research.

## 2. Unobserved heterogeneity in HMMs

### *2.1. Modeling unobserved heterogeneity*

Properly incorporating unobserved heterogeneity into a marketing model satisfies both practical and methodological concerns. From a practical standpoint, modeling heterogeneity is important to managers, so they can make inferences for individual consumers, stores, demographic market areas—or, as in our empirical application, teams (Allenby & Rossi, 1998). From a methodological standpoint, ignoring unobserved heterogeneity across individuals or incorporating an improper representation of the heterogeneity structure can lead to biased inferences (Dubé, Hitsch, & Rossi, 2010; Hsiao, 2003; Lenk & DeSarbo, 2000). This problem is even more severe in dynamic models, such as HMMs (Heckman, 1991; Hyslop, 1999). Deciding which heterogeneity structure to use can be difficult, as the true nature of the unobserved heterogeneity is usually unknown a priori and the selection of a discrete or continuous distribution is largely an empirical issue (Andrews, Ansari, & Currim, 2002; Michalek, Ebbes, Adigüzel, Feinberg, & Papalambros, 2011; Otter, Tüchler, & Frühwirth-Schnatter, 2004). Typically, a discrete distribution leads to a latent class model and a continuous distribution leads to a random coefficients model (Wedel et al., 1999). A third and more flexible representation of unobserved heterogeneity is the RCM model (Allenby et al., 1998; Lenk & DeSarbo, 2000), which combines the flexibility of the semi-parametric latent class model and allows for parametric variation within the latent classes through the formulation of random coefficients.[1] The RCM model can capture skewed and multimodal distributions of unobserved heterogeneity and can accommodate both discrete and continuous forms of unobserved heterogeneity in the data (Lenk & DeSarbo, 2000).

[1] This model is also known as the Bayesian mixture approach and the mixture-of-normals model.

Please cite this article as: Kappe, E., et al., A random coefficients mixture hidden Markov model for marketing research, *International Journal of Research in Marketing* (2018), https://doi.org/10.1016/j.ijresmar.2018.07.002