



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing

Sam K. Hui, Tom Meyvis, Henry Assael

To cite this article:

Sam K. Hui, Tom Meyvis, Henry Assael (2014) Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing. Marketing Science 33(2):222-240. <https://doi.org/10.1287/mksc.2013.0835>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing

Sam K. Hui, Tom Meyvis, Henry Assael

Stern School of Business, New York University, New York, New York 10012
{khui@stern.nyu.edu, tmeyvis@stern.nyu.edu, hassael@stern.nyu.edu}

Researchers often collect continuous consumer feedback (moment-to-moment, or MTM, data) to understand how consumers respond to a variety of experiences (e.g., viewing a TV show, undergoing a colonoscopy). Analyzing how MTM judgments are integrated into overall evaluations allows researchers to determine how the structure of an experience influences consumers' post-experience satisfaction. However, this analysis is challenging because of the functional nature of MTM data. As such, previous research has typically been limited to identifying the influence of heuristics, such as relying on the average intensity, peak, and ending.

We develop a Bayesian functional linear model to study how the different "moments" in the MTM data contribute to the overall judgment. Our approach incorporates a (temporally) weighted average of MTM data as well as specific "patterns" such as peak and trough, thus nesting previous approaches such as the "peak-end" rule as special cases. We apply our methodology to analyze data on TV show pilots collected by CBS. Our results reveal several interesting empirical findings. First, the last quintile of a TV show is weighted about four times as much as each of the first four quintiles. Second, patterns such as peak and trough do not play substantial roles in driving overall evaluations for TV shows. Finally, the last quintile is more important for procedural dramas than for serial dramas. We discuss the managerial implications of our results and other potential applications of our general methodology.

Keywords: moment-to-moment data; functional data analysis; Bayesian functional linear model; TV show pilot testing; peak-end rule

History: Received: January 5, 2013; accepted: November 10, 2013; Preyas Desai served as the editor-in-chief and Michel Wedel served as associate editor for this article. Published online in *Articles in Advance* January 10, 2014.

1. Introduction

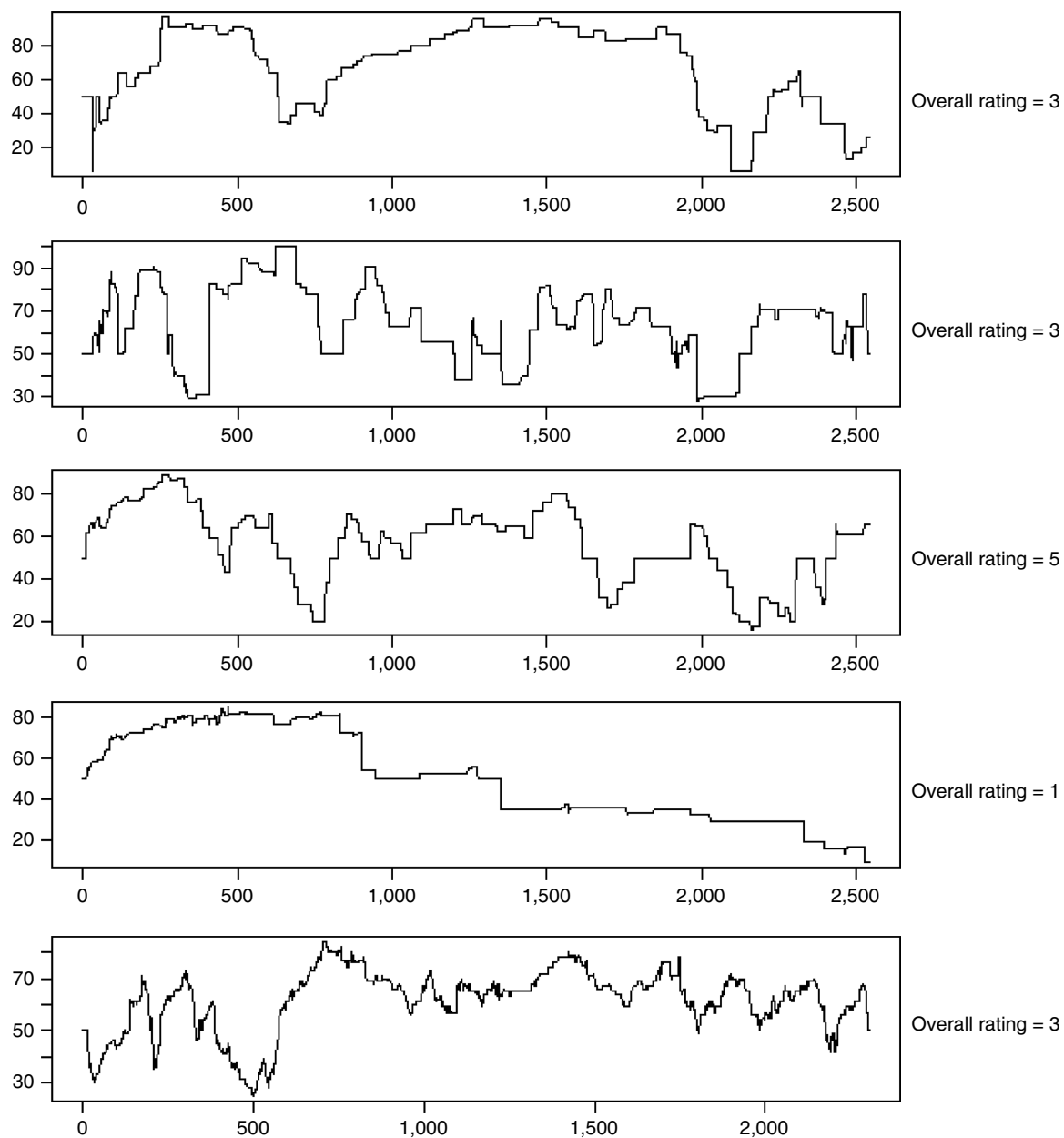
Marketing researchers often use continuous consumer feedback to track the progression of consumers' responses to a variety of experiences. This continuous feedback is then used to craft more effective advertisements or to create better consumer experiences. For instance, advertising researchers examine consumers' moment-to-moment emotional response to commercials to identify ways to design more effective ads (Baumgartner et al. 1997; Polsfuss and Hess 1991; Woltman Elpers et al. 2003, 2004). Similarly, television networks test new shows by asking viewers to provide real-time feedback using an electronic dial (e.g., Perception Analyzer; see <http://www.perceptionanalyzer.com>). Show producers then use these data to identify parts of the show that viewers like or dislike and fine-tune the script accordingly.

Aside from its use by marketing researchers to create more effective ads or more enjoyable consumer experiences, continuous consumer feedback is also used in other domains, such as politics and medicine. In politics, undecided voters who watch political

debates are asked to provide continuous ratings of their reactions, which are then used to assess voters' sentiments about various political issues (Correll et al. 2004, Kirk and Shrill 2011); in medicine, researchers continuously record the intensity of pain experienced by patients undergoing medical procedures to construct a "pain profile" and to find ways to minimize the remembered unpleasantness of the medical procedure (Lewis et al. 1995, Redelmeier et al. 2003).

Each of above examples involves the collection and analysis of continuous moment-to-moment (MTM) feedback, denoted as $(x(t); t \in [0, T])$, along with an overall evaluation y . As a concrete example, Figure 1 shows a few data samples collected during a TV show pilot test conducted by CBS that will be analyzed in our empirical application. Depending on the specific setting, the MTM measure may correspond to subjective liking, persuasiveness of an argument, or physical pain. By measuring consumers' reactions during the experience, researchers can identify parts of the experience that stand out (e.g., scenes from a show that are particularly enjoyable, parts of the medical procedure

Figure 1 Moment-to-Moment Dial Data for Five Individuals



Note. Data were collected from a TV show pilot study conducted by CBS.

that are particularly painful). Moreover, by relating these MTM reactions to overall judgments, researchers can also determine how specific patterns and features of the experience influence subsequent judgments and behaviors (e.g., how does the enjoyment of a show over time influence viewers' attitude toward, and subsequent exposure to, that show?).

Understanding how MTM measurements $x(t)$ are aggregated into an overall judgment y has important managerial implications. For instance, in the context of TV shows, if viewers systematically overweight later parts of a show, show producers may want to

devote more resources to improving the end of the show, including possibly shifting the funniest jokes or most exciting scenes toward the end. Similarly, in medical research, understanding how a pain profile is integrated into an overall pain assessment may help medical professionals adjust medical procedures to reduce patients' recalled averseness of the procedure. For instance, Redelmeier et al. (2003) observed that extending a colonoscopy by adding a less painful procedure at the end reduces the patient's overall pain assessment and increases his or her intention to return for another colonoscopy.

Estimating the relationship between $x(t)$ and y is challenging because the independent variable $x(t)$ is a function rather than a scalar, and thus it involves functional data analysis issues (Ramsay and Silverman 2005). Fundamentally, the biggest challenge is that $x(t)$ is often of very high dimensionality, or even of infinite dimensionality in the case of truly continuous measurement. In the TV pilot show data that we analyze in §5, MTM dial ratings are recorded second by second, resulting in approximately 2,500 measurements in each profile. The high dimensionality of $x(t)$ makes it difficult to accurately estimate the contribution of each moment to the overall evaluation. Furthermore, previous research suggests that y may be related to $x(t)$ not only through the temporal average of $x(t)$ but also through certain specific patterns of $x(t)$, such as the peak ($\max_t x(t)$), the trough ($\min_t x(t)$), and the end ($x(T)$). Because of these methodological challenges, previous research has typically bypassed the difficulty of specifying a formal statistical model of how $x(t)$ is integrated into y ; instead, such research has examined how y relates to summary descriptors of the $x(t)$ function, such as the unweighted average of $x(t)$ (Aaker et al. 1986, Polsfuss and Hess 1991, Thorson and Friestad 1989) or the “peak-end rule” (Baumgartner et al. 1997, Fredrickson and Kahneman 1993, Redelmeier and Kahneman 1996).

In this paper, we propose a formal Bayesian functional linear model to study how MTM data are aggregated into overall evaluations. We begin by assuming that the overall evaluation y is driven by a (temporally) weighted average of $x(t)$, where the weights are defined by a weighting function $h(t)$. We then approximate $h(t)$ using a step function and specify a random walk prior (West and Harrison 1997) on the function value at each interval. This nonparametric specification puts a “smoothness prior” on $h(t)$, which allows us to estimate $h(t)$ flexibly without making any a priori functional form assumptions. Next, we further extend our model to allow for the additional roles of specific patterns in $x(t)$ (e.g., peak, trough). As such, our model nests existing models, such as the peak-end rule and the unweighted average, as special cases.

To ensure that our proposed approach is able to recover the weighting function $h(t)$ and other model parameters given the available data, we conduct a set of simulation studies in §4. Next, we apply our proposed method to analyze data from 3,572 participants who took part in a series of TV show pilot tests conducted by CBS, a major national TV network. We demonstrate that our proposed approach outperforms alternative heuristics-based methods in a holdout prediction study. Our results reveal several empirical insights. First, the last quintile of the show

is weighted around four times as much as each of the first four quintiles. Second, in contrast to findings in the context of commercials (Baumgartner et al. 1997), we find that specific patterns such as the peak or trough do not play substantial roles in driving overall judgments in the context of TV shows. Third, the last quintile is more important for procedural dramas than for serial dramas.

Our paper aims to make both a methodological and a substantive contribution. At a methodological level, we develop a formal Bayesian functional modeling approach to analyze how MTM data are integrated into overall judgments. Our approach is general enough to be directly applied to other applications that involve the collection of MTM data. At a substantive level, we apply our proposed method to a real-world data set that captures MTM dial ratings from a TV show pilot test, allowing us to better understand how viewers’ overall evaluation of a TV show may be improved by changing the temporal structure of the show. In doing so, we contribute to the ongoing debate about the generality and boundary conditions of peak and end effects (e.g., Miron-Shatz 2009). Moreover, our findings also have direct implications for broadcasters seeking to raise advertising revenues. We show that the show likeability ratings are strongly linked to subsequent Nielsen ratings, which are the currency for up-front negotiations regarding ad expenditures on TV shows. Thus, if an improvement in show structure succeeds in improving likeability, and thereby Nielsen ratings, then this has direct cost implications for advertisers and revenue implications for the network.

The remainder of the paper is organized as follows. Section 2 begins with an overview of MTM data and functional data analysis in marketing. In §3, we propose a Bayesian functional linear model to analyze MTM data. In §4, we conduct simulation studies to ensure that our proposed model is able to accurately recover the weighting function $h(t)$ and other model parameters. In §5, we apply our proposed model to analyze actual data from CBS. Finally, §6 concludes with directions for future research.

2. Background: Moment-to-Moment Data and Functional Data Analysis

We provide an overview of previous research that has collected and analyzed MTM data (with a focus on marketing applications), followed by a brief review of functional data analysis.

2.1. Collection and Analysis of MTM Data in Marketing

Marketing researchers have long been collecting MTM data to better understand consumers’ reactions to television commercials. Baumgartner et al. (1997) used

a computerized “feelings monitor” to track participants’ MTM liking of a commercial and showed that overall ad judgment is strongly correlated with the peak emotional experience and the final moment of the series. Similarly, in Polsfuss and Hess (1991), participants used a wireless handheld device to indicate their MTM evaluations while viewing a commercial. The authors then used the resulting data to pinpoint the “key selling seconds” of a commercial and to predict the overall liking of the commercial. They found that the overall liking of a commercial advertisement is closely related to the average of MTM ratings. Finally, Ramanathan and McGill (2007) extended this research to the domain of social influences by studying how MTM evaluations of a commercial are affected by the social presence of other consumers. They found that joint consumption of an experience leads to similar MTM patterns and more positive retrospective evaluations.

MTM data are also used to capture concepts other than subjective liking. For instance, Woltman Elpers et al. (2003) collected MTM ratings of “entertainment” and “information” value as participants viewed a commercial; these authors showed that these two types of MTM ratings have a strong multiplicative effect on consumers’ likelihood to continue watching a commercial. This is conceptually similar to the work by Hughes (1992), who continuously tracked consumers’ affective and cognitive reactions to measure the wearout of a commercial over time. Other MTM measures that have been collected in the previous literature include humor (Woltman Elpers et al. 2004), warmth (Aaker et al. 1986, Vanden Abeele and MacLachlan 1994), and fear and hope (Madrigal and Bee 2005).

Recent technological advances have enabled marketing researchers to measure other types of MTM data beyond self-reported judgments. Using an automated facial-expression detection device, Teixeira et al. (2012) measured the MTM intensity of joy and surprise expressed by participants who were watching Internet video advertisements. The authors then related these MTM emotional intensities to attention and concentration (captured using eye-tracking data) as well as to viewing behavior (the timing of participants “zapping” the ad). Using a dynamic frailty model, they found that the emotions of surprise and joy effectively concentrate attention and retain viewers. Based on these results, Teixeira et al. derived the “optimal emotion trajectories” to aid effective TV advertisement designs.

In most of the above examples, the goal of the research is to study how MTM measures (e.g., humor, liking, warmth) are *integrated* into a scalar dependent variable of interest (e.g., humorousness of the ad, attitude toward the ad). Given that a standard,

off-the-shelf methodology that relates a functional independent variable $x(t)$ to a scalar dependent variable y is generally unavailable, researchers have typically relied on heuristics-based approaches such as computing the overall average or using the peak-end rule. However, analyzing how a continuous function is aggregated into a scalar dependent variable is closely related to an active research area in statistics known as *functional data analysis*, which suggests a more comprehensive alternative approach. We review this research next.

2.2. Applications of Functional Data Analysis in Marketing

Functional data analysis (Jank and Shmueli 2006, Ramsay and Silverman 2005) refers to the statistical analysis of data that are in the form of functions, most commonly functions of time. Functional data arise naturally in many diverse settings such as meteorology (e.g., Ramsay and Li 1998), medicine (e.g., Zhou et al. 2010), human biology (e.g., Ramsay and Silverman 2002), and finance (e.g., modeling of yield curves in Hays et al. 2012), among others.

The issue that is most central to the problem we address in this paper is how to analyze the relationship between a *functional* independent variable $x(t)$ and a *scalar* dependent variable of interest y .¹ For instance, Sood et al. (2009) used a functional principal component approach (fPCA) to analyze penetration (diffusion) curves, and they developed a functional regression model to predict the market penetration of new products. Their approach is based on a decomposition of the observed raw data into functional principal components, which are then used for regression and clustering. Sood et al. demonstrated that their proposed functional data analytic approach outperforms other models such as the Bass model and the Meta-Bass model. Using a similar fPCA approach, Foutz and Jank (2010) proposed a functional shape analysis approach to forecast the box office performance of motion pictures using the dynamic pre-release price patterns from virtual stock markets (e.g., the Hollywood Stock Exchange). Their proposed method automatically identifies a small number of specific “shapes” (which are essentially the functional principal components) from the prerelease price patterns and uses these shapes to predict box office revenue. Functional data analysis has also been applied to the study of online auctions, where it has been used to model price dynamics and to forecast the winning bid (e.g., Reithinger et al. 2008, Wang et al. 2008).

¹ Note that the related problem of relating a functional independent variable $x(t)$ and a functional dependent variable $y(t)$, which corresponds to the data structure of Teixeira et al. (2012), is not considered in this paper.

Note that the fPCA approach discussed above primarily has a predictive focus. Although the fPCA approach also attempts to understand the relationship between a function and a scalar, it does not directly speak to how MTM measures are *integrated* into a scalar measure. Rather, the focus is on identifying the major source of variations across the functional observations (i.e., the functional principal components) and then relating those features to the scalar dependent variable through a regression framework. Later in §5.2, we will compare the fit of our model to the fPCA approach.

In light of this brief review of the previous literature, we can now specify the key methodological and substantive contributions we intend to make with this paper. First, we develop a general Bayesian methodology to study how MTM data are integrated into an overall evaluation. This provides an “off-the-shelf” method that researchers can directly use to analyze the aggregation of functional MTM measures into a scalar dependent variable of interest. Second, and substantively, we apply our method to analyze MTM data from participants in a TV show pilot test. TV shows are, of course, much more complex and longer-lasting than the commercials typically examined in previous studies, which results in findings that are substantially different from those found in previous research, as will be discussed in detail in §5.

3. Analysis of Moment-to-Moment Data Using a Bayesian Functional Linear Model

In this section, we develop a Bayesian functional linear model to analyze MTM data. Since we will be applying our methodology to analyze dial data for TV pilot shows in §5, we use the terminology of TV shows when describing our proposed method. As discussed earlier, our proposed model can be applied to other MTM data that are collected in marketing research.

3.1. Notations and Time Normalization

Throughout this paper, we use the following notations: i ($i = 1, \dots, I$) indexes respondents, and t ($t \in [0, T_i]$) indexes the time during which the i th respondent is viewing a TV show. Thus, $x_i(t)$ denotes the value of the dial data for the i th respondent at time t ; we assume that $x_i(t) \geq 0$ for all t . In the context of TV pilot testing, $x_i(t)$ refers to the MTM subjective liking of the TV show at time t . In most applications, $x_i(t)$ is measured either continuously or at a very high sampling frequency (e.g., second-to-second). The variable z_i denotes the show that the i th respondent watches (in the empirical application in §5, a total of 19 shows are being tested). Finally, y_i denotes the overall rating given by the i th respondent at the conclusion of the experience (i.e., an overall evaluation of the TV show).

We first normalize all T_i to 1 following James et al. (2009); i.e., the time in each show is scaled to the interval $[0, 1]$. Therefore, t refers to the in-show time relative to the duration of the show, rather than absolute time; that is, at the 20-minute mark, $t = 0.5$ for a show that is 40 minutes long and $t = 0.33$ for a show that is an hour long. Our choice to analyze MTM data with respect to relative time instead of absolute time is based on both technical and theoretical considerations. First, by normalizing all T_i to 1, we have essentially “registered” or aligned each dial profile to the same time domain ($t \in (0, 1)$), which is an important and necessary preprocessing step for functional data analysis procedures (Ramsay and Silverman 2005). Furthermore, previous research on MTM data has typically found that participants demonstrate “duration neglect” (e.g., Baumgartner et al. 1997, Fredrickson and Kahneman 1993, Schreiber and Kahneman 2000), i.e., the total duration of the experience typically does not have a significant effect on overall evaluation, which supports conducting statistical analysis based on relative rather than absolute time. We will provide some empirical evidence in support of this assumption in §5.3.

3.2. Bayesian Functional Linear Model

Recall that the goal of our analysis is to understand how functional MTM data $x_i(t)$ are aggregated into a scalar dependent variable y_i (an overall rating). We begin with a functional linear model (Cardot et al. 1999, 2003; James et al. 2009) that links y_i to the functional inner product of $x_i(t)$ and a coefficient function $h(t)$:

$$y_i = \alpha + \int_0^1 x_i(t)h(t)dt + \delta_{z_i} + \varepsilon_i, \quad (1)$$

where $h(t) \geq 0$ for all t for model identification² and δ_{z_i} is an additive random effects term that captures any unobserved variations across different TV shows.³ Thus, Equation (1) specifies an integral equation (Corduneanu 1991) between scalar y_i and function $x_i(t)$ that serves to estimate the coefficient function $h(t)$.

To understand the intuition behind the integral equation in Equation (1) and to facilitate further interpretation, we reparametrize Equation (1) by defining

² This identifying restriction is necessary because in a functional linear model, $h(t)$ is only identified up to an additive function $g(t)$, where $\int_0^1 x_i(t)g(t)dt = 0$. That is, $\tilde{h}(t) = h(t) + g(t)$ gives the same likelihood as $h(t)$.

³ In Equation (1), we are assuming that the coefficient function $h(t)$ is homogeneous across shows. In §5.3, we present a model extension that allows $h(t)$ to be heterogeneous across TV shows through a hierarchical specification.

$h(t) = \beta h^*(t)$, where $h^*(t) \geq 0$ and $\int_0^1 h^*(t) dt = 1$. Thus, we have

$$y_i = \alpha + \beta \int_0^1 x_i(t) h^*(t) dt + \delta_{z_i} + \varepsilon_i. \quad (2)$$

As can be seen in Equation (2), we are modeling the overall rating y_i through a weighted temporal average of $x_i(t)$, defined by the weighting function $h^*(t)$, which is itself a probability distribution function as it integrates to 1 (Casella and Berger 1992). Thus, through the estimate of $h^*(t)$, we can understand how the different parts of $x_i(t)$ are being weighted and aggregated into an overall rating y_i . For instance, if the later parts of $x_i(t)$ are more heavily weighted, we should find that $h^*(t)$ increases over time. In contrast, if the earlier parts of $x_i(t)$ are more heavily weighted, $h^*(t)$ should decrease over time. Clearly, the special case where y_i is driven by the unweighted average of $x_i(t)$ can be captured by having $h^*(t) = 1$ for all t . Thus, the current approach nests unweighted averaging (e.g., Aaker et al. 1986, Polsfuss and Hess 1991, Thorson and Friestad 1989) as a special case.

3.3. Estimating the Coefficient Function $h(t)$

We now estimate the coefficient function $h(t)$ given the model in Equation (1) (or equivalently, $h^*(t)$ in Equation (2)). First, we note that given any finite number of observations, $h(t)$ is always underdetermined (Ramsay and Silverman 2005); that is, given that both $x(t)$ and $h(t)$ are of very high (or infinite) dimensionality, it would generally be possible for multiple $h(t)$'s to perfectly fit the dependent variables y_i if no restrictions are placed on $h(t)$ (James et al. 2009). In other words, it is necessary to put constraints or prior knowledge on $h(t)$ in order to estimate it. To understand this intuition, consider a discretized case where S (equally spaced) measurements of $x(t)$ are taken over the time horizon ($t \in (0, 1)$); i.e., x_s ($s = 1, 2, \dots, S$), where $x_s = x(s/S)$. By also discretizing $h(t)$ into h_s ($s = 1, 2, \dots, S$), where $h_s = h(s/S)$, Equation (1) can now be written as $y_i = \alpha + \sum_s h_s x_{is} + \delta_{z_i} + \varepsilon_i$, which can then be estimated using a regression approach by treating x_s ($s = 1, 2, \dots, S$) as independent variables. The high dimensionality problem occurs when $S > I$ (the number of observations), which results in the design matrix not having full rank, and hence $h(t)$ is underdetermined.⁴ Another concern is that such an approach does not take into account the natural ordering of the indexes s that results from the functional nature of the data (Cardot et al. 2003, Hastie and Mallows 1993). More specifically, there is no reason for $h(t)$ to have huge discrete

jumps over time, but such prior domain knowledge is not reflected here.

Thus, in the context of our Bayesian framework, the underdetermination of $h(t)$ suggests that it is necessary to specify an informative prior distribution on $h(t)$. We motivate our choice of an appropriate prior specification on $h(t)$ by reviewing the “penalization approach” (James et al. 2009), which estimates parameters through minimizing a “penalized” objective function that incorporates both fit to the observed data and some penalty associated with certain aspects of $h(t)$. The penalization approach is closely related to the ideas of putting a prior distribution on $h(t)$ to represent prior knowledge and of “complexity control” in machine learning (Evgeniou et al. 2007). The most common form of this approach penalizes a function based on the second derivative of $h(t)$, which corresponds to encouraging smoothness in the coefficient function, resulting in a penalized b -spline (e.g., Cardot et al. 2003). Given that, in our application, $h(t)$ represents the weight that consumers apply to the experience at time t , it is reasonable (from a psychological perspective) to assume that $h(t)$ does not have any rapid jumps or discontinuities, which justifies incorporating smoothness into $h(t)$. Thus, we specify a prior on $h(t)$ that directly incorporates smoothness over time, allowing us to accurately estimate $h(t)$.

We propose the following Bayesian approach based on a random walk prior (West and Harrison 1997, Choi et al. 2010). Following the approach of Cardot et al. (2003) and James et al. (2009), we first divide the time domain $t \in (0, 1)$ into a fine grid consisting of M equally spaced intervals, where the j th interval is denoted as $T_j = ((j-1)/M, j/M]$. In both the simulation studies in §4 and the empirical application in §5, we take $M = 100$.⁵ Thus, $h(t)$ is now approximated by a step function that takes the value of $q_j > 0$ in the interval T_j . Given the step-function approximation, the integral $\int_0^1 x(t) h(t) dt$ can now be written as $(1/M) \sum_{j=1}^M \bar{x}_j q_j$, where \bar{x}_j denotes the average value of $x_i(t)$ within the interval T_j . With this representation, Equation (1) now takes the form of a standard linear model, which allows for straightforward posterior computations (to be discussed later).

To incorporate a smoothness prior into the prior specification of $h(t)$, we specify that the function values q_j ($q_j > 0$) are linked together via a random walk prior (West and Harrison 1997); i.e., $\pi(q_j | q_{j-1}) \sim N(q_{j-1}, v^2)$ for $j \geq 1$. Similar to the penalization approach discussed earlier, the random walk prior specification discourages rapid changes in $h(t)$ and encourages it to vary smoothly over time (e.g., Choi

⁴ Of course, we can solve the underdetermination problem by taking $S \ll I$. However, this defeats the purpose of having MTM data, as the approximation of h_s to $h(t)$ will be very crude.

⁵ In Appendix §A.2, we assess the robustness of our approach with respect to the choice of M by taking $M = 50$ and $M = 200$; the results remain substantially unchanged.

et al. 2010). In addition, all the conditional priors are Gaussian, allowing for efficient conjugate computations.

3.4. Incorporating Specific Patterns (Peak and Trough)

So far, we have assumed that the overall rating y_i is driven only by the weighted aggregation of MTM dial ratings $x_i(t)$. Previous literature in psychology and in marketing (Ariely 1998, Baumgartner et al. 1997, Fredrickson and Kahneman 1993, Kahneman et al. 1993, Varey and Kahneman 1992) has often argued that the overall rating y_i is not only driven by the temporal aggregation of $x_i(t)$ but also by specific patterns of $x_i(t)$ that are not (directly) related to time t . For instance, as discussed earlier, Baumgartner et al. (1997) showed that the overall evaluation of an advertisement is dominated by the peak emotional experience as well as the final moment of the advertisement. Redelmeier and Kahneman (1996) showed that retrospective judgments of pain are strongly correlated with the peak intensity of pain and with the intensity of pain recorded during the last moments of the procedure. In a similar vein, Woltman Elpers et al. (2004) found that the overall perception of the humorous nature of a commercial is driven primarily by the peak moments in terms of surprise and humor.

To capture the additional role of specific patterns, we now extend our model by generalizing Equation (2) as follows:

$$y_i = \alpha + \beta \int_0^1 x_i(t) h^*(t) dt + \sum_k \gamma_k \int_0^1 x_i(t) g_k(t, x_i(t), \lambda_k) dt + \delta_{z_i} + \varepsilon_i, \quad (3)$$

where $\int_0^1 g_k(t, x_i(t), \lambda_k) dt = 1$ and $\gamma_k \geq 0$. In Equation (3), the additional terms $\gamma_k \int_0^1 x_i(t) g_k(t, x_i(t), \lambda_k) dt$ capture the role of specific patterns (e.g., peak, trough) that are represented as a function of $x_i(t)$. Specific patterns that are considered in previous research include the peak, trough, beginning, end, and trend/slope (Ariely 1998, Baumgartner et al. 1997, Schreiber and Kahneman 2000). Among these patterns, the beginning and end patterns are related to time and can thus be captured by the time-dependent weighted average and the weighting function $h^*(t)$; we therefore focus on capturing the effect of the peak and the trough in our model. We return to the issue of capturing the trend/slope below.

We represent the peak of the MTM dial data $x_i(t)$ with the following functional (as depicted in the left panel of Figure 2):

$$g_1(t, x_i(t), \lambda_p) = \begin{cases} 1/\lambda_p & \text{if } t \in (t^* - \lambda_p/2, t^* + \lambda_p/2), \\ 0 & \text{otherwise,} \end{cases}$$

where $t^* = \arg \max_t x_i(t)$. (4)

That is, the peak measure of MTM dial data $x_i(t)$ is defined as the average dial rating within a time window centered on the time point where the maximum dial rating occurs.⁶ The length of the time period around the peak during which the peak functional is computed, which we refer to as the *bandwidth parameter*, is represented by λ_p . We assume that λ_p is unknown and sample from its posterior distribution along with other model parameters. Thus, another important feature of our modeling framework is that the definitions of specific patterns are also data-driven, which provides some insight into how patterns should be defined. To reflect the prior knowledge (or intuition) that the peak should represent only a short duration around the maximum dial rating, we specify a Gamma(10, 0.001) prior distribution on λ_p . This prior distribution has a prior mean of 0.01, corresponding to 1% of the show, which is around half a minute, and puts more than 99% of the prior density on λ_p being smaller than 0.02 (i.e., 2% of the show, which is about a minute).

We similarly define the trough functional as follows (depicted in the right panel of Figure 2):

$$g_2(t, x_i(t), \lambda_R) = \begin{cases} 1/\lambda_R & \text{if } t \in (t^* - \lambda_R/2, t^* + \lambda_R/2), \\ 0 & \text{otherwise,} \end{cases}$$

where $t^* = \arg \min_t x_i(t)$. (5)

As before, the bandwidth parameter λ_R is assumed to be unknown and is given an informative Gamma(10, 0.001) prior distribution; it is sampled along with other model parameters. Putting everything together, we specify our full model as follows:

$$y_i = \alpha + \beta \int_0^1 x_i(t) h^*(t) dt + \sum_k \gamma_k \int_0^1 x_i(t) g_k(t, x_i(t), \lambda_k) dt + \delta_{z_i} + \varepsilon_i, \quad (6)$$

where

$$\varepsilon_i \sim N(0, \sigma^2), \quad (7)$$

$$\delta_{z_i} \sim N(0, \tau^2), \quad (8)$$

$$h(t) = \beta h^*(t) = \sum_{j=1}^M q_j 1(t \in T_j),$$

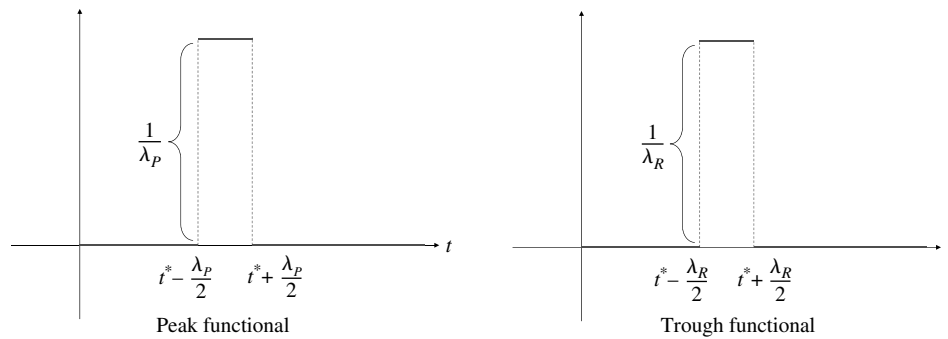
where $T_j = \left(\frac{j-1}{M}, \frac{j}{M} \right]$, (9)

$$\pi(q_j | q_{j-1}) \sim N(q_{j-1}, v^2) \quad \text{for } j \geq 1, q_j \geq 0, \quad (10)$$

$$\lambda_p, \lambda_R \sim \text{Gamma}(10, 0.001). \quad (11)$$

⁶ An alternative operationalization of the peak measure is to compute a sequence of moving averages (each with width λ_p) of the dial rating and define the peak as the maximum of such moving averages. This definition is much more computationally intensive; we leave it for future research.

Figure 2 The Peak and Trough Functionals



Note that in Equation (9), $1(t \in T_j)$ denotes the indicator function, which takes the value of 1 if $t \in T_j$ and 0 otherwise. To complete our Bayesian model specification, we specify conjugate, weakly informative priors on all other model parameters (Gelman et al. 2003). Specifically, α and q_0 ($q_0 \geq 0$) are both given independent, weakly informative $N(0, 100^2)$ prior distributions, and the variance parameters σ^2 and v^2 are both given independent, weakly informative $\text{Inv-}\chi^2(0.001, 1)$ prior distributions. The γ_k 's are given weakly informative truncated normal distributions with mean 0 and variance 100^2 . We sample from the posterior distribution of all model parameters using a standard Markov chain Monte Carlo (MCMC) procedure (Gelman et al. 2003), which was coded in C++; the details of our computational procedure are described in Appendix §A.1.

We now return to the issue of incorporating the trend/slope pattern. We argue that the model described by Equation (6) has already incorporated the time-varying effect of the first derivative $x'(t)$. To see this, assume that there is an additional additive term $\int_0^1 x'_i(t)k(t)dt$ in Equation (6), where $k(t)$ denotes the contribution of trend/slope at time t (i.e., $x'(t)$) that captures the time-varying effect of the trend/slope on the overall rating. Using integration-by-parts, we can write $\int_0^1 x'_i(t)k(t)dt = [x_i(t)k(t)]_0^1 - \int_0^1 x_i(t)k'(t)dt$. Thus, the second term shows that the contribution of the first derivative $x'(t)$ can be absorbed into the term $\int_0^1 x_i(t)h^*(t)dt$. In other words, the time-varying effect of $x'(t)$ cannot be separately identified from the time-varying effect of $x(t)$, and thus our model is in effect estimating the “net contribution” of each time point on the overall rating.

4. Simulation Study

We conduct a set of simulation studies to ensure that our proposed methodology is able to accurately recover all model parameters (in particular, the weighting function $h^*(t)$), given the amount of data available in the data set used in our empirical application in §5. In each simulation, we take $x_i(t)$ to be

the dial ratings we obtained from $I = 3,572$ participants in the actual data set that we will analyze, and we simulate overall ratings y_i using a set of known model parameters and different specifications of the weighting function $h^*(t)$.

We conduct six simulations, each using a different weighting function $h^*(t)$. In each case, we take $\alpha = 1.0$, $\beta = 0.04$, and $\sigma^2 = 0.6$, which are chosen to maximally replicate the actual data in our empirical application. In addition, we set the bandwidth parameters for the peak and trough patterns (λ_P and λ_R , respectively) to be both 0.01, which specifies that, as discussed earlier, the peak (trough) rating is taken to be the average of the dial ratings during the 1% of the show centered around the max (min) value. Furthermore, we set the coefficients for peak and trough to be $\gamma_1 = \gamma_2 = 0.01$.

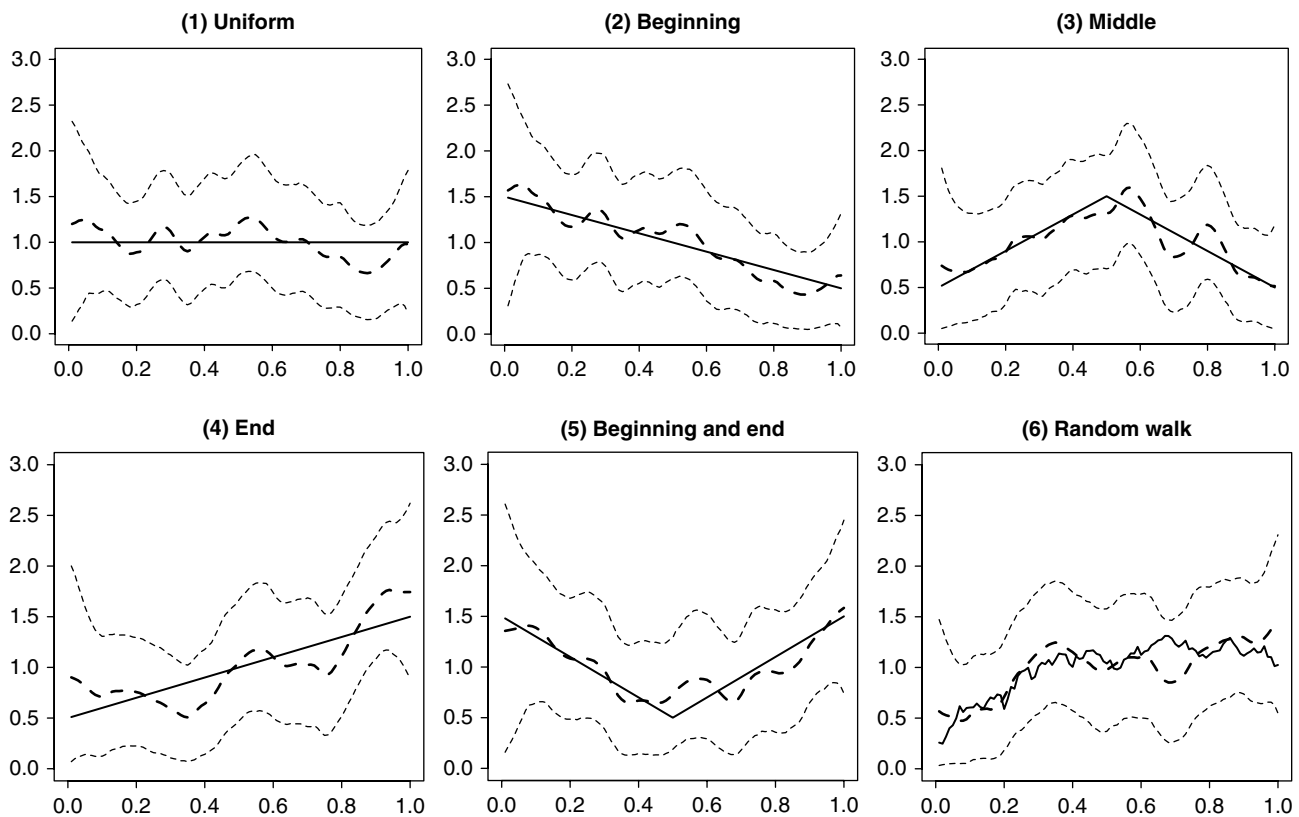
The six different specifications for $h^*(t)$ are shown by the solid lines in Figure 3. In Case 1, participants weigh each moment within the show equally; i.e., $h^*(t) = 1$ for all t . Case 2 represents the situation where participants weigh the beginning of a show more heavily than other parts (Blacker 1998), which is represented by the weighting function $h^*(t) = 1.5 - t$. In Case 3, participants put greater weight on the middle of the show, with $h^*(t) = 0.5 + 2t$ for $t \leq 0.5$ and $h^*(t) = 1.5 - 2(t - 0.5)$ for $t > 0.5$. Next, in Case 4, participants weigh the end of the show more, with a weighting function $h^*(t) = 0.5 + t$. Case 5 is a combination of Cases 2 and 4, where participants weigh both the beginning and end more than the middle, with $h^*(t) = 1.5 + 2t$ for $t \leq 0.5$ and $h^*(t) = 0.5 + 2(t - 0.5)$ for $t > 0.5$. Finally, in Case 6, $h^*(t)$ is simulated from an arithmetic random walk. Note that in all six cases, the normalizing constraint $\int_0^1 h^*(t)dt = 1$ is satisfied.

The estimated model parameters, along with their posterior 95% intervals, are shown in Table 1. The estimated weighting functions $h^*(t)$ are shown in Figure 3, along with pointwise 95% posterior intervals. As can be seen in Table 1, in each case our methodology is able to accurately recover all model parameters,

Table 1 Simulation Study: Estimated Posterior Means of Model Parameters, Along with 95% Posterior Intervals

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 |
|-------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| α | 1.080 (0.434, 1.743) | 1.089 (0.454, 1.720) | 1.074 (0.438, 1.707) | 0.899 (0.248, 1.517) | 0.955 (0.258, 1.595) | 0.896 (0.287, 1.476) |
| β | 0.043 (0.031, 0.547) | 0.043 (0.031, 0.055) | 0.042 (0.031, 0.053) | 0.042 (0.031, 0.054) | 0.040 (0.027, 0.053) | 0.041 (0.032, 0.052) |
| γ_1 | 0.008 (0.004, 0.012) | 0.008 (0.003, 0.012) | 0.008 (0.005, 0.012) | 0.009 (0.005, 0.014) | 0.010 (0.006, 0.014) | 0.010 (0.006, 0.013) |
| γ_2 | 0.008 (0.005, 0.012) | 0.008 (0.005, 0.012) | 0.010 (0.006, 0.013) | 0.010 (0.007, 0.014) | 0.011 (0.007, 0.016) | 0.010 (0.007, 0.013) |
| λ_p | 0.014 (0.003, 0.028) | 0.014 (0.003, 0.028) | 0.011 (0.002, 0.024) | 0.012 (0.002, 0.025) | 0.012 (0.002, 0.028) | 0.012 (0.002, 0.027) |
| λ_T | 0.011 (0.003, 0.024) | 0.012 (0.003, 0.025) | 0.010 (0.002, 0.023) | 0.013 (0.003, 0.029) | 0.013 (0.002, 0.032) | 0.011 (0.003, 0.025) |

Note. True values: $\alpha = 1.0$, $\beta = 0.04$, $\gamma_1 = \gamma_2 = 0.01$, and $\lambda_p = \lambda_T = 0.01$.

Figure 3 Simulation Study: Posterior Estimates of the Weighting Function $h^*(t)$ 

Notes. The solid line depicts the known true $h^*(t)$, and the broken line is the posterior mean estimate of $h^*(t)$. The bounds are pointwise 95% posterior intervals.

and the posterior means of the model parameters are very close to the known true value, which always falls within the 95% posterior intervals. Furthermore, Figure 3 suggests that our proposed approach is also able to recover the weighting function $h^*(t)$ reasonably well given the data in our empirical application.

5. Empirical Application: TV Show Pilot Testing

5.1. Data Overview

We apply our proposed methodology to analyze MTM dial data collected during a series of TV show pilot tests conducted by CBS between April 2002

Table 2 Show-Level Summary Statistics

| ID | Show | Duration (minutes) | Mean dial rating (across participant and time) | Avg. overall evaluation | Nielsen rating (Live + 7, %HHs) |
|----|--------------------------------------|-----------------------|---|----------------------------|------------------------------------|
| 1 | <i>3 lbs</i> ^{OP} | 42.4 | 62.68 | 3.16 | 6.1 |
| 2 | <i>Century City</i> ^{OP} | 43.7 | 59.78 | 3.33 | 5.3 |
| 3 | <i>Close to Home</i> ^{OP} | 44.8 | 71.98 | 3.92 | 6.4 |
| 4 | <i>Clubhouse</i> ^{OS} | 43.8 | 63.46 | 3.50 | 5.2 |
| 5 | <i>Company Town</i> ^{PS} | 39.3 | 65.92 | 3.40 | N/A |
| 6 | <i>Criminal Minds</i> ^{PP} | 40.8 | 74.05 | 3.90 | 8.1 |
| 7 | <i>CSI: Miami</i> ^{PP} | 46.4 | 73.60 | 4.07 | 10.5 |
| 8 | <i>CSI: NY</i> ^{PP} | 44.3 | 71.79 | 3.96 | 8.3 |
| 9 | <i>Ghost Whisperer</i> ^{OP} | 43.6 | 70.09 | 3.77 | 5.9 |
| 10 | <i>Hack</i> ^{PS} | 45.7 | 60.85 | 3.45 | 5.2 |
| 11 | <i>Jericho</i> ^{OS} | 38.4 | 69.37 | 3.58 | 5.0 |
| 12 | <i>Joan of Arcadia</i> ^{OS} | 31.9 | 60.75 | 3.51 | 5.3 |
| 13 | <i>Presidio Med</i> ^{OP} | 45.3 | 61.42 | 3.53 | 5.0 |
| 14 | <i>Queens Supreme</i> ^{OP} | 41.1 | 63.31 | 3.43 | 4.8 |
| 15 | <i>Shark</i> ^{OP} | 44.4 | 67.39 | 3.61 | 7.6 |
| 16 | <i>Smith</i> ^{OS} | 62.0 | 62.09 | 3.16 | 6.7 |
| 17 | <i>The Unit</i> ^{PP} | 44.9 | 63.96 | 3.40 | 7.0 |
| 18 | <i>Threshold</i> ^{OS} | 35.8 | 67.62 | 3.51 | 5.1 |
| 19 | <i>Without a Trace</i> ^{PP} | 40.5 | 68.18 | 3.76 | 9.9 |

Notes. PP, police procedural; OP, other procedural; PS, police serial; OS, other serial. Nielsen ratings show the percentage of households (HHs) that watched the show live or time-delayed (DVR) within seven days of its original broadcast. The show *Company Town* did not air.

and May 2006. As listed in Table 2, 19 different TV show pilots with durations ranging from 32 minutes to 62 minutes, along with their genres (“police” versus “other”) and types (procedural drama versus serial drama), were tested in the CBS Television City Research Center located at the MGM Grand hotel in Las Vegas. Participants were interested volunteers. They did not receive financial compensation for participating in the TV show pilot study.

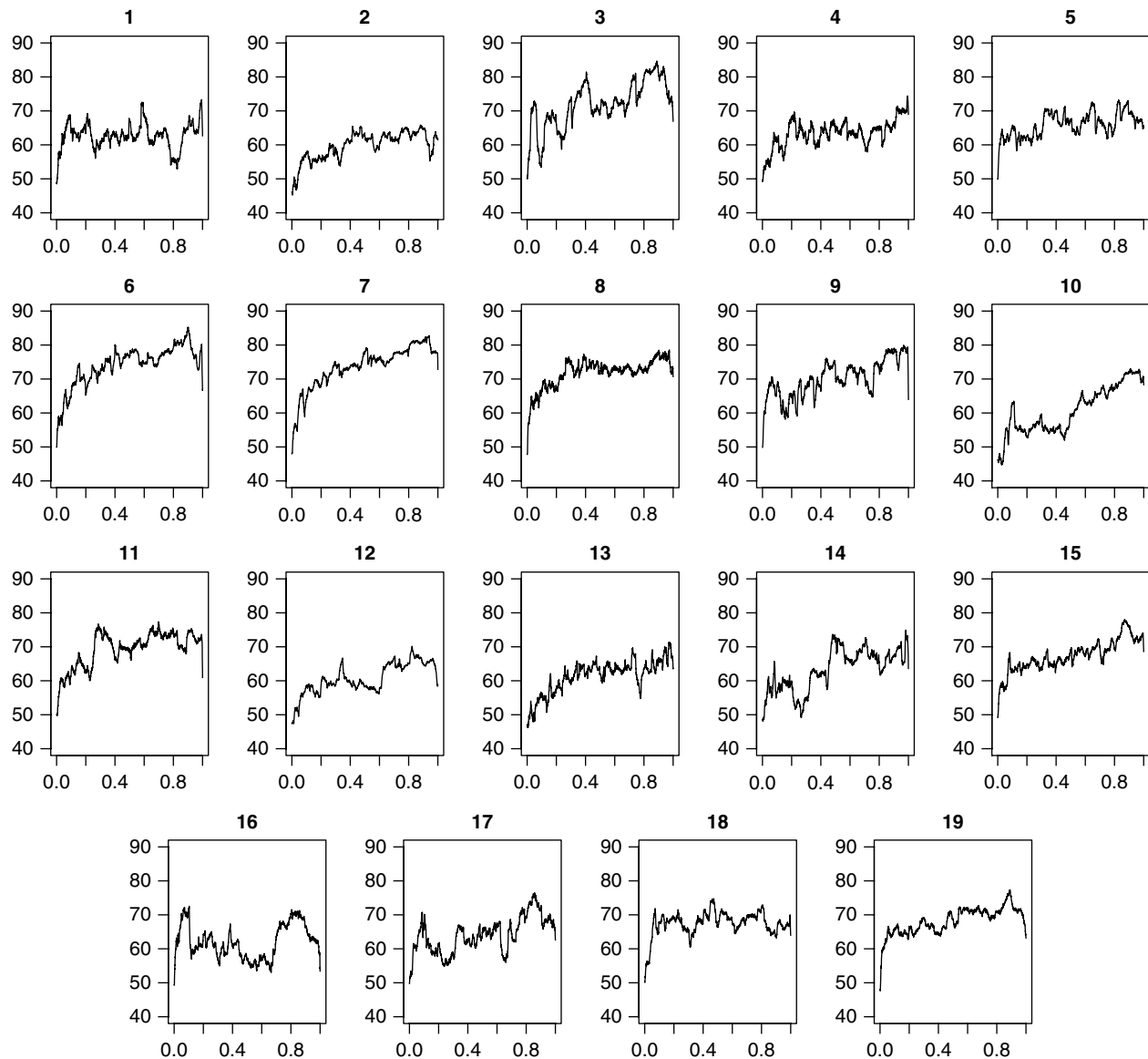
A total of 3,572 participants took part in the TV pilot tests, with each participant watching only one pilot show that lasted, on average, about 45 minutes. While participants were watching the TV show, they recorded their MTM liking of the show on a 101-point scale using an electronic dial system (0 = “wouldn’t watch it or turn off the TV,” 50 = “neutral,” 100 = “extremely interested”). This procedure is analogous to MTM measures used in prior marketing studies, such as electronic dial ratings to indicate affective reactions to a sequence of pictures (Pham et al. 2001), continuous computer mouse movements to indicate feelings elicited by a commercial (Baumgartner et al. 1997), and continuous adjustments of a slider scale to indicate enjoyment of a music video (Nelson et al. 2009). Examples of the resulting data are shown in Figure 1. At the end of the show, each participant filled out an exit survey indicating his or her overall evaluation of the TV show on a five-point scale, where 1 represents “poor” and 5 represents “excellent.”⁷

Figure 4 and Table 2 present several key show-level summary statistics. First, Figure 4 shows the mean MTM dial rating (across participants) over time for each of the 19 TV shows. Consistent with Baumgartner et al. (1997), MTM dial ratings generally increase over time during the show; a linear regression of mean dial rating versus time results in positive and statistically significant coefficients ($p < 0.001$) in all of the 19 shows in our data set. This suggests that participants, in general, gradually liked the TV show more as the show progressed. Our observed mean patterns over time, however, are much more complicated than those shown in Baumgartner et al., which is to be expected, given that TV shows are richer and more layered than the 90-second commercials studied in Baumgartner et al.

Next, Table 2 shows the average MTM dial rating (computed by averaging across participants and across time) for each TV show, along with its average overall evaluation across participants who viewed the show. As expected, at the show level, the correlation between average dial rating and overall evaluation is positive and statistically significant ($r = 0.857$, $p < 0.001$), providing some face validity for the MTM measurements. Next, we examine the relationship with the Nielsen ratings, which are available for all but one of the shows (as shown in the last column

⁷ We treat the 1-to-5 scale as if the data came from a continuous interval scale, which is consistent with the analysis performed

in the previous literature (Wirtz et al. 2003). In Appendix §A.3, we develop a “cutpoint” formulation based on the ordered probit model (Greene 2011). The results remain substantially unchanged.

Figure 4 Mean Dial Ratings Across Participants in Each of 19 Pilot Shows

of Table 2).⁸ We find that the average overall rating (across respondents) is significantly correlated with the resulting Nielsen rating after the show was aired ($r = 0.61$, $p = 0.007$), which suggests that an improvement in likeability ratings may increase subsequent exposure to the show.

We now turn to the individual-level summary statistics in Table 3. As shown in the table, the mean overall evaluation is 3.59, with a standard deviation of 1.17. Across participants, the mean of the average MTM dial rating is 66.40, with a standard deviation of 19.95. As discussed earlier, we also compute three important features of each dial profile: the peak (maximum

dial rating), the trough (minimum dial rating), and the end (final position of the dial). Across all participants, the peak, trough, and end are, on average, 91.34 ($sd = 14.03$), 25.42 ($sd = 20.34$), and 64.63 ($sd = 27.95$), respectively. Furthermore, Table 4 presents a correlation matrix that explores the relationships between

Table 3 Individual-Level Summary Statistics from the TV Pilot Show Data

| | Mean | sd | Min | Max |
|---|-------|-------|------|--------|
| Overall evaluation (1-to-5 scale) | 3.59 | 1.17 | 1.00 | 5.00 |
| Avg. dial rating | 66.40 | 19.95 | 0.00 | 99.45 |
| Max dial rating (peak) | 91.34 | 14.03 | 0.00 | 100.00 |
| Min dial rating (trough) | 25.42 | 20.34 | 0.00 | 91.00 |
| Final position of the dial rating (end) | 64.63 | 27.95 | 0.00 | 100.00 |

⁸ One of the 19 shows did not air; hence its Nielsen rating is unavailable.

Table 4 Correlation Matrix Between the Overall Evaluation, Average Dial Rating, Peak, Trough, and End Dial Rating

| | Evaluation | Average | Peak | Trough | End |
|------------|------------|---------|------|--------|------|
| Evaluation | 1.00 | 0.71 | 0.44 | 0.47 | 0.56 |
| Average | 0.71 | 1.00 | 0.64 | 0.55 | 0.68 |
| Peak | 0.44 | 0.64 | 1.00 | 0.12 | 0.43 |
| Trough | 0.47 | 0.55 | 0.12 | 1.00 | 0.46 |
| End | 0.56 | 0.68 | 0.43 | 0.46 | 1.00 |

overall evaluation, average rating, peak, trough, and end. As can be seen, the overall evaluation is most strongly correlated with the average dial rating ($r = 0.71$), followed by the end dial rating ($r = 0.56$), the trough ($r = 0.47$), and the peak ($r = 0.44$); all four of the correlations are statistically significant ($p < 0.001$).

5.2. Empirical Results and Comparison to Benchmark Methods

We apply our proposed model to actual data and compare its performance, both in- and out-of-sample, vis-à-vis several benchmark methods. Discussions with executives at CBS reveal that one critical goal of collecting MTM data is to understand the key determinants of overall evaluations and thus identify the specific enjoyment patterns that will produce more favorable evaluations. In other words, ideally the model should predict how overall enjoyment changes if the show is altered based on feedback using MTM measures (e.g., moving an early action sequence to a later part of the show or selectively investing to improve either the opening sequence or ending). Given this goal, an out-of-sample test where an entire TV show is being held out seems to be the most appropriate test given the available data, as it is in the spirit of predicting the liking of new shows unseen in the data. Thus, when conducting an out-of-sample prediction, we hold out entire TV shows one by one using leave-one-out cross-validation, a widely used method for accurately estimating prediction error (Hastie et al. 2009).

For both the in-sample fit and out-of-sample prediction, we compare our model against the following models that are representative of the analyses performed in the previous literature:

i. *Temporal aggregation*: The overall evaluation y is modeled as a linear function of the unweighted average dial rating: $y_i = \alpha + \beta \bar{x}_i(t) + \delta_{z_i} + \varepsilon_i$, where $\bar{x}_i(t) = \int_0^1 x_i(t) dt$.

ii. *Peak-end rule (equal weights)*: The overall evaluation y is modeled as a function of an unweighted average of the maximum and final position of the dial rating: $y_i = \alpha + \beta(\frac{1}{2} \max x_i(t) + \frac{1}{2} x_i(T)) + \delta_{z_i} + \varepsilon_i$.

iii. *Peak-end rule (unequal weights)*: We also consider a generalization of (ii) where unequal weights are allowed for the peak and end ratings: $y_i = \alpha + \beta_1 \max x_i(t) + \beta_2 x_i(T) + \delta_{z_i} + \varepsilon_i$.

Table 5 Results for In-Sample Fit and Out-of-Sample Prediction

| Method | In-sample MSE | Out-of-sample MSE |
|---|---------------|-------------------|
| Temporal aggregation | 0.665 | 0.681 |
| Peak-end rule (equal weights) | 0.828 | 0.860 |
| Peak-end rule (unequal weights) | 0.827 | 0.860 |
| Functional PCA | 0.638 | 0.667 |
| Proposed model without peak and trough patterns | 0.630 | 0.646 |
| Proposed model with peak and trough patterns | 0.629 | 0.647 |

iv. *Functional PCA*: Following Sood et al. (2009) and Foutz and Jank (2010), a smoothing spline is first applied to the raw MTM data. The overall evaluation of y is then modeled as a linear function of the first two functional principal components of $x(t)$ and the first two functional principal components of the first derivative $x'(t)$.

The fit of each benchmark model in terms of both in- and out-of-sample mean squared error (MSE), together with the results of our proposed model (with and without including peak and trough patterns), is shown in Table 5. Our results clearly indicate that the proposed model provides superior in-sample (0.629) and out-of-sample (0.647) MSE compared with all other methods. Among the four benchmark models, the functional PCA method performs the best, as expected, with an in-sample MSE (0.638) that is only slightly higher than that of the proposed method ($p < 0.05$) but with an out-of-sample MSE (0.667) that is substantially worse compared with the proposed method ($p < 0.001$).⁹ This suggests that fPCA, which is based on a decomposition of the available data, is able to adequately summarize the temporal patterns in the existing shows, yet it does not do as well for predicting new shows that are not in the sample because those patterns are unobserved in the available data. In contrast, our proposed Bayesian functional linear model directly models the “contribution” of each moment without having to rely on observed patterns (i.e., functional principal components) in the functional data. Thus, our proposed model is not pattern dependent and hence is more suitable for predicting the overall rating of new shows.

Among the other three benchmark methods, the temporal aggregation model performs better than the methods that are based on the peak-end rule. Given the complexity and multilayered nature of modern TV shows (Johnson 2005), this result is consistent with Ariely and Zauberman’s (2000) finding that overall judgments of segmented (rather than continuous) hedonic experiences rely more on the average intensity of the experience and less on specific patterns.

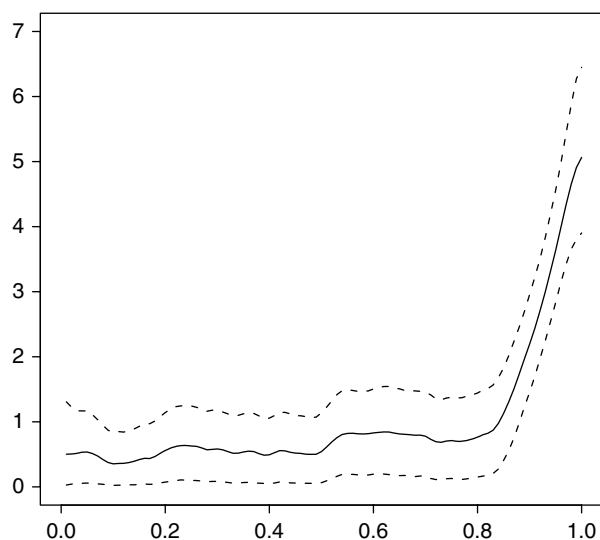
⁹ The p -values here are two-sided, computed using the paired t -test.

Table 6 Posterior Estimates for Model Parameters in the Proposed Model

| | Posterior mean | 95% posterior interval |
|-------------|----------------|------------------------|
| α | 0.856 | (0.365, 1.312) |
| β | 0.038 | (0.031, 0.046) |
| γ_1 | 0.000 | (0.000, 0.001) |
| γ_2 | 0.002 | (0.000, 0.005) |
| λ_P | 0.011 | (0.005, 0.019) |
| λ_T | 0.011 | (0.004, 0.021) |

This hypothesis is further supported by the observation that the predictive performance of the proposed model with peak and trough patterns is virtually indistinguishable from the performance of the model without those patterns, indicating that these patterns do not meaningfully contribute to the prediction of the overall evaluation. Stated more formally, the log Bayes factor in favor of the reduced model (i.e., without peak and trough effects), as computed using the Savage–Dickery density ratio (Rossi et al. 2005), is +34.0, which corresponds to “decisive” evidence in favor of the reduced model (Kass and Raftery 1995).

We now turn to the parameter estimates for our proposed model, which are shown in Table 6. The posterior mean estimate for α is 0.856, with a 95% posterior interval of (0.365, 1.312). The posterior mean estimate for β is 0.038, with a 95% posterior interval of (0.031, 0.046), indicating that, as expected, the overall evaluation is significantly related to a weighted average of the dial ratings. The posterior mean of the normalized weighting function $h^*(t)$ is shown by the solid line in Figure 5, and the broken lines show the pointwise 95% posterior intervals. As can be seen in the figure, participants clearly place more weight on the later part of the show compared with

Figure 5 Estimated Normalized Weighting Function $h^*(t)$ 

Note. The solid line shows the posterior mean, and the broken lines are pointwise 95% posterior intervals.

the earlier parts, with weights increasing dramatically in the last quintile of the show. To further understand this pattern, we compute the relative weight of each quintile by computing the value of the integrals $\int_0^{0.2} h^*(t) dt$, $\int_{0.2}^{0.4} h^*(t) dt$, $\int_{0.4}^{0.6} h^*(t) dt$, $\int_{0.6}^{0.8} h^*(t) dt$, and $\int_{0.8}^{1.0} h^*(t) dt$. The relative weights are 0.102, 0.120, 0.131, 0.151, and 0.496 for the first to fifth quintiles, respectively. In other words, the last quintile of the show is approximately four times as important as each of the first four quintiles.

Next, we find that the estimated values of γ_1 and γ_2 , which respectively correspond to the magnitude of the peak and trough effects, are very small. The posterior mean of γ_1 (the coefficient for the peak effect) is 0.0003, with a 95% posterior interval of (0.0000, 0.0012), and the posterior mean of γ_2 (the coefficient of the trough effect) is 0.0022, with a 95% posterior interval of (0.0002, 0.0048). Thus, these results indicate that the peak and trough effects are much smaller than the effect of the weighted temporal aggregation. More specifically, the relative contribution of the unweighted temporal aggregation is $\beta/(\beta + \gamma_1 + \gamma_2) = 93.7\%$, and the relative contribution of the peak and the trough is 0.8% and 5.5%, respectively.

In sum, we find evidence for a very pronounced end effect but little evidence for any incremental contribution from the peak of the experience. The absence of a peak effect (and the finding that the trough is actually a better predictor than the peak) is intriguing given that peak effects have been observed in many previous studies using MTM ratings (e.g., Fredrickson and Kahneman 1993, Redelmeier and Kahneman 1996), including a similar study by Baumgartner et al. (1997) in which participants indicated their enjoyment while watching a TV commercial. As such, our findings suggest that peak experiences may be less distinct and memorable for longer, multilayered stimuli such as TV shows than for shorter, less complicated stimuli such as TV commercials. On the other hand, the observance of a disproportionate end effect is in line with previous findings (though it is exceptionally pronounced in this case). Yet whereas the end effect has often been demonstrated, many recent studies have shown that it fails to obtain for a variety of stimuli and in a variety of situations (e.g., Montgomery and Unnava 2009, Robinson et al. 2011), including complex, multilayered experiences (Miron-Shatz 2009). As such, that we observe such a pronounced end effect for TV shows adds to this ongoing debate and provides support for prior speculation that end effects may also occur for complex experiences, but only when this experience forms a cohesive unit (Miron-Shatz 2009).

5.3. Model Extension: Incorporating Heterogeneity of $h(t)$ Across Shows

Up to this point, we have made the simplifying assumption that $h(t)$ is homogeneous across all shows. The homogeneity of $h(t)$ is arguably a reasonable assumption when consumers' response is instinctive and the context is less influential (e.g., patients' reactions to pain across medical procedures), but in the TV setting, $h(t)$ may be different depending on the genre (e.g., police versus medical) or type (e.g., procedural drama versus serial drama) of show. We therefore introduce a model extension that incorporates heterogeneity of $h(t)$ across shows through a hierarchical specification. We change our notation slightly in this section: i denotes the respondent, k denotes the show, and j denotes the time within a show. Also, as discussed above, because the reduced model (without peak and trough effects) provides a better fit to the data than the full model (based on Bayes factor), we start with the reduced model and incorporate heterogeneity into that specification. Formally stated, we specify the following model:

$$y_{ik} = \alpha + \int_0^1 x_{ik}(t) h_k(t) dt + \delta_k + \varepsilon_{ik}, \quad (12)$$

$$h_k(t) = \sum_{j=1}^M q_{kj} I(t \in T_j), \quad \text{where } T_j = \left(\frac{j-1}{M}, \frac{j}{M} \right], \quad (13)$$

$$q_{kj} | q_{k,j-1} \sim N(q_{k,j-1} + \tilde{\Delta}_{j-1}, \omega^2) \quad (\text{for } j \geq 2), \quad (14)$$

$$q_{k1} \sim N(\bar{q}_1, \omega^2), \quad (15)$$

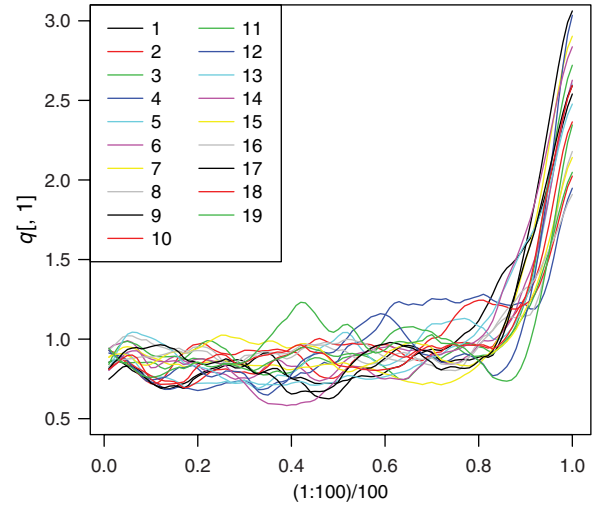
$$\tilde{\Delta}_j \sim N(0, v^2). \quad (16)$$

Equations (14) and (15) allow for heterogeneity of $h(t)$ across shows. More specifically, for each show k , the sequence $(q_{k1}, q_{k2}, \dots, q_{kM})$ follows a random walk specification where the mean increment $(\tilde{\Delta}_j)$ in each period is the same across shows, with variations across shows captured by the variance parameter ω^2 . Note that if $\omega^2 \rightarrow 0$, the above specification in Equations (12)–(16) reduces to the specification in Equations (6)–(11), where $h(t)$ is assumed to be homogeneous across shows.

After estimating the model with MCMC, we compute the posterior mean of $h_k^*(t)$ for each show. As can be seen in Figure 6, although there are some minor differences across shows, the general pattern that the final quintile carries the highest weight holds across all 19 shows. The shape of the normalized weighting function for each show is also very similar, thus providing some support for the homogeneity assumption (which provided more statistical power to estimate $h(t)$ earlier).

Our new results also allow us to assess the validity of our normalization procedure in §3.1, which scales time in each show to the interval $[0, 1]$. Specifically,

Figure 6 Estimated Normalized Weighting Function $h^*(t)$ Across All 19 Shows Under the Specification That Allows for Heterogeneity



we can look at the correlation between the aggregate weight put on the final quintile of each show $(\int_{0.8}^1 h_k^*(t) dt)$ and the duration of each show (shown in Table 2). We find that the correlation is statistically insignificant ($r = -0.067$, $p = 0.78$): the disproportionate impact of the final quintile does not depend on the absolute length of this quintile. In contrast, although the impact of this relative time interval does not depend on the length of the show, the impact of absolute intervals does. For instance, show length is negatively correlated with the importance of the last 5 minutes ($r = -0.62$, $p = 0.005$) as well as the last 10 minutes ($r = -0.65$, $p = 0.002$). In other words, the last 5 (or 10) minutes matter less if they are part of a longer show and thus constitute a relatively smaller part of the show. Together, this supports the assumption behind our normalization procedure—i.e., that the impact of a show segment depends on relative time rather than absolute time.

Although the last quintile has a disproportionate effect in each of the 19 shows, there is some variation in the magnitude of this effect. Therefore, another interesting question is how this end effect varies across TV shows of different genres (police versus other) and types (procedural drama versus serial drama), as denoted by the superscripts in Table 2. Similar to the above analysis, we estimate a regression with the final quintile of each show $(\int_{0.8}^1 h_k^*(t) dt)$ as the dependent variable and with the genre and type of the show as independent variables. The results are shown in Table 7. Our results indicate that the ending (i.e., final quintile) is even more important for procedural dramas than for serial dramas ($\beta = 0.029$, $p < 0.1$). Since procedural dramas (where a conflict is resolved in each stand-alone episode) could be seen as

Table 7 Regression of the Weight of the Final Quintile on Genre and Type of TV Show

| | Estimated coefficient | se | p-value |
|-------------------------|-----------------------|-------|---------|
| <i>Intercept</i> | 0.291* | 0.013 | 0.000 |
| <i>Genre: Police</i> | −0.015 | 0.016 | 0.365 |
| <i>Type: Procedural</i> | 0.029^ | 0.016 | 0.085 |
| R^2 | 0.194 | | |

* $p < 0.05$; ^ $p < 0.10$.

more cohesive than serial dramas (where each episode is part of a larger storyline), this provides further support for earlier suggestions that the cohesiveness of the experience is a key condition for obtaining end effects for complex experiences (Miron-Shatz 2009).

6. Discussion and Conclusion

In this paper, we develop a Bayesian functional analysis approach to analyze the relationship between MTM data and overall evaluations. Specifically, we develop a functional linear model to capture how MTM data $x_i(t)$ are aggregated into an overall rating y_i through a weighting function $h(t)$. We extend our model to allow for the additional roles of specific patterns (peak and trough) in driving overall evaluations, and we apply our proposed model to a MTM dial data set collected from 3,572 participants during pilot TV show tests conducted by CBS. We find that our proposed model outperforms other approaches used in the previous literature, in terms of both in- and out-of-sample MSE. Furthermore, we demonstrate that (i) the final quintile of a TV show is weighted about four times as much as any of the other quintiles, (ii) the role of specific patterns (peak and trough) in driving overall evaluations is limited in the context of TV shows, and (iii) the final quintile is particularly important for procedural dramas. Interestingly, although previous research in the “thin slice” literature has shown that the popularity of, for instance, politicians (Todorov et al. 2005) and musical performances (Tsay 2013) can be well predicted from first impressions or minimal information, our findings indicate that the popularity of more complex narrative stimuli such as television shows is better predicted by the reaction to the end of the experience.

That being said, we want to raise a note of caution regarding any studies inferring end effects based on relationships between MTM ratings and overall evaluations (including past research as well as the current study). Specifically, the stronger relationship between later MTM ratings and the overall evaluation may be caused by not only later segments having a greater impact on evaluations but also later MTM ratings already partly reflecting overall evaluations. That is, as the show progresses, the MTM ratings may

be influenced not only by what is happening on the screen at the moment but also by what has happened on the screen so far. However, we would argue that, in this particular study, this issue is unlikely to fully account for the end effect since the increase in weight does not occur gradually over time (as one would expect if the MTM ratings reflect the continuously updated liking of the show), but rather is restricted to the final quintile.

Another potential alternative explanation of our findings is that the final dial rating is likely to be the most salient at the moment that the overall evaluation is made, which attributes the observed correlation to measurement artifacts. However, this recency effect explanation has been addressed by prior research (Baumgartner et al. 1997), which obtained similar results when the measure of the overall rating was delayed. Furthermore, the differences in weights of the final quintile for procedural drama versus serial drama indicate that this greater weighting of the end does not just reflect a recency effect but depends on the meaning of the end. Of course, the most definitive way of ruling out any measurement issues and establishing the greater causal impact of the final show fragments would be to compare two different cuttings of the same show where different parts are relocated based on the MTM dial profile and then compare their overall evaluations to validate the model predictions.

With the above caveats in mind, our results suggest that in the context of a television show, the final quintile is the most important and specific patterns have little impact on the overall evaluation. What are the implications of these findings for TV show producers seeking to maximize viewers’ overall evaluation (given its positive correlation with Nielsen ratings)? First, our results imply that producers should not allocate resources to provide a particularly liked peak experience (e.g., a particularly exciting yet expensive car chase sequence). In fact, although prior research has focused on peak effects rather than trough effects, our results indicate that the latter are more important (albeit both effects are relatively small). That is, in the context of TV shows, it may be more important to avoid exceptionally disliked segments than to insert an exceptionally liked segment. However, although peak effects did not materialize, end effects turned out to be very substantial, especially for procedural dramas. This suggests that show makers may want to allocate more resources toward improving the quality of the show’s ending. For instance, they could move sequences around so that the best sequences (e.g., funniest jokes, most exciting action sequence) appear toward the end of the show, or, alternatively, invest more resources in the final resolution of the plot rather than in making the preceding scenes more thrilling or entertaining.

As the first paper in marketing that introduces a Bayesian functional analysis framework to analyze MTM data, we have only scratched the surface of this active and growing research area—both in terms of methodological development and in terms of the substantive managerial insights that can be extracted from MTM data. Below, we conclude by discussing a few promising directions for future research in this area.

First, in this paper, we have focused on two specific patterns, a single peak and a single trough. Future research may look into the role of *multiple* peaks and troughs in MTM data. For instance, if there is a secondary peak in a dial profile, how much weight would be given to it compared with the primary peak? Also, does it matter when the peak occurs? Future research may consider incorporating these patterns into the functional analysis framework developed here.

Second, all participants in the current study watched the entire TV show from start to finish, as they were not given the option to stop and leave in the middle of a show. To more closely mimic TV viewing conditions in real life, TV networks may consider allowing participants to leave in the middle of a show (i.e., change the channel) if they do not wish to finish viewing the show. In such a setting, our model would need to be supplemented by a separate model that captures the attrition process (i.e., if a participant's utility drops below a certain threshold, she would leave the show at that point), perhaps adopting an approach similar to that of Teixeira et al. (2012). From a substantive perspective, we would speculate that our current finding that troughs matter more than peaks would be even more pronounced if people can choose to stop or switch channels, since viewers are more likely to notice the presence of a terrible segment than the absence of a great one.

Finally, the methodology developed in this paper is fully general and can be directly applied to other substantive areas, which includes all the areas mentioned in §§1 and §2.1. For instance, our method can be useful for the Home Shopping Network, which frequently runs moment-to-moment evaluations on its televised sales pitches of various products (Dialsmith 2013). In particular, the model extension that incorporates heterogeneity can be useful for capturing differences in weighting functions $h(t)$ across different product categories. In addition, our methodology can be applied to analyze the relationship between a pain profile and an overall assessment of pain during a colonoscopy or the likelihood of returning for the next scheduled screening procedure. Similarly, our proposed method can also be used to analyze televised political debates and help formulate campaign strategies. In summary, we hope that our paper will

stimulate research in this area and that it provides a springboard for future research using MTM data.

Acknowledgments

The authors thank CBS and David F. Poltrack, CBS Executive Vice President, Planning and Research, for contributing the TV pilot testing data used in the analysis.

Appendix

A.1. MCMC Computation

We describe the MCMC procedure we used to estimate the Bayesian functional linear model described by Equations (6)–(11). We draw from the full conditional distributions of model parameters in the following order: α , γ_1 , γ_2 , δ , q , λ_P , λ_R , v^2 , τ^2 , and σ^2 . Each step follows directly from standard Bayesian posterior computations (Gelman et al. 2003) and is outlined as follows.

Step 1. Drawing α : Recall that \bar{x}_j denotes the average value of $x_i(t)$ in the time interval T_j ; α can be drawn using standard conjugate normal posterior computation (Gelman et al. 2003), where the prior is $N(0, 100^2)$ and the data are defined by

$$y_i - \frac{1}{M} \sum_{j=1}^M \bar{x}_j q_j - \gamma_1 \int_0^1 x_i(t) g_1(t, x_i(t), \lambda_P) dt \\ - \gamma_2 \int_0^1 x_i(t) g_2(t, x_i(t), \lambda_R) dt - \delta_{z_i}.$$

Step 2. Drawing γ_1 and γ_2 : To sample from the full conditional distribution of γ_1 , we first compute the data quantities $(1/\int_0^1 x_i(t) g_1(t, x_i(t), \lambda_P) dt)(y_i - (1/M) \sum_{j=1}^M \bar{x}_j q_j - \gamma_2 \int_0^1 x_i(t) g_2(t, x_i(t), \lambda_R) dt - \delta_{z_i})$, where the i th observation has precision proportional to $(\int_0^1 x_i(t) g_1(t, x_i(t), \lambda_P) dt)^2$. We then proceed with standard (truncated) normal posterior computations to draw from the full conditional distribution of γ_1 . Similarly, γ_2 can be sampled using an analogous procedure.

Step 3. Drawing δ 's: Similar to α , δ_l can be drawn using standard conjugate normal posterior computation, where the prior is given by $N(0, \tau^2)$ and the data are given by

$$y_i - \alpha - \frac{1}{M} \sum_{j=1}^M \bar{x}_j q_j - \gamma_1 \int_0^1 x_i(t) g_1(t, x_i(t), \lambda_P) dt \\ - \gamma_2 \int_0^1 x_i(t) g_2(t, x_i(t), \lambda_R) dt, \quad \text{where } z_i = l.$$

Step 4. Drawing q_j 's: First, note that given the random walk prior specification in Equation (10), q_j has the following conditional prior distribution (see Choi et al. 2010, West and Harrison 1997):

$$\text{For } j = 1: q_1 \sim N(q_2, v^2). \\ \text{For } 1 < j < M: q_j \sim N\left(\frac{q_{j-1} + q_{j+1}}{2}, \frac{v^2}{2}\right). \\ \text{For } j = M: q_M \sim N(q_{M-1}, v^2).$$

Next, we compute the data quantities

$$\frac{1}{\bar{x}_j} y_i - \frac{1}{M} \sum_{l \neq j} \bar{x}_l q_l - \gamma_1 \int_0^1 x_i(t) g_1(t, x_i(t), \lambda_P) dt \\ - \gamma_2 \int_0^1 x_i(t) g_2(t, x_i(t), \lambda_R) dt - \delta_{z_i},$$

which have precisions proportional to $(\bar{x}_j)^2$. We then proceed with standard (truncated) normal posterior computations to sample from the full conditional distribution of q_j .

Step 5. Drawing λ_p and λ_R : Because standard conjugate computations are not available for the bandwidth parameters λ_p and λ_R , we use a random walk Metropolis-Hastings algorithm to sample from their posterior distributions. For each parameter, we use a Gaussian random walk proposal distribution with the mean centered on the previous draw; the variance of the proposal distribution is adjusted to achieve an acceptance rate close to 50% (Gelman et al. 2003).

Step 6. Drawing v^2 , τ^2 , and σ^2 : Since, as stated earlier, all the variance parameters are given conjugate, weakly informative $\text{Inv-}\chi^2(0.001, 1)$ prior distributions, we can draw directly from their fully conditional distribution using standard conjugate computations of variance parameters (Gelman et al. 2003).

After sampling from the joint posterior distribution of all model parameters, we postprocess the posterior samples by dividing each q_j by the normalizing constant $(1/M) \sum_{j=1}^M q_j$ and setting β equal to the normalizing constant. This allows us to separate the coefficient function $h(t)$ into a normalized weighting function $h^*(t)$, as depicted in Figure 5, and the slope coefficient β . We coded the above MCMC procedure in C++, using the GNU Scientific Library, and ran 40,000 iterations. The first 20,000 iterations were discarded as burn-in samples, leaving the last 20,000 draws to summarize the posterior distribution (Gelman et al. 2003). Standard diagnostics confirmed that convergence had been reached. The full C++ code and posterior draws are available from the authors upon request.

A.2. Robustness Check with Respect to M

We explore the robustness of our results with respect to the choice of M by estimating the model with $M = 50$ and $M = 200$. Given that the peak and trough effects are

very small and that the reduced model (without peak and trough effects) provides a better fit to the data than the full model (based on Bayes factor), we focus on the specification that excludes peak and trough effects. The estimated $h^*(t)$ under $M = 50$ and $M = 200$ are shown in the Figure A.1.

As can be seen in Figure A.1, the overall patterns are very similar to the one presented in the paper (with $M = 100$). We further compute the total weight of each quintile under $M = 50$ and $M = 200$. The results, which clearly indicate that our findings are robust with respect to M , are summarized in the following table.

| | First quintile | Second quintile | Third quintile | Fourth quintile | Fifth quintile |
|-----------|----------------|-----------------|----------------|-----------------|----------------|
| $M = 50$ | 0.096 | 0.120 | 0.130 | 0.159 | 0.495 |
| $M = 100$ | 0.102 | 0.120 | 0.131 | 0.151 | 0.496 |
| $M = 200$ | 0.108 | 0.121 | 0.132 | 0.150 | 0.488 |

A.3. Ordered Probit Formulation

We developed a “cutpoint” formulation based on the ordered probit model (Greene 2011) to specifically account for the five-point Likert scale used to collect the overall ratings y . This allows us to assess the robustness of our analysis that treats the five-point scale as a continuous variable. Formally stated, we replace y_i in Equation (6) with latent variables u_i , where

$$y_i = \begin{cases} 1 & \text{if } u_i \leq 1.5, \\ 2 & \text{if } 1.5 < u_i \leq c_2, \\ 3 & \text{if } c_2 < u_i \leq c_3, \\ 4 & \text{if } c_3 < u_i \leq 4.5, \\ 5 & \text{if } u_i > 4.5. \end{cases}$$

The cutpoints (c_2, c_3) , along with the latent variables u_i , are then sampled together with other model parameters using

Figure A.1 Robustness Check with Respect to $M = 50$ and $M = 200$

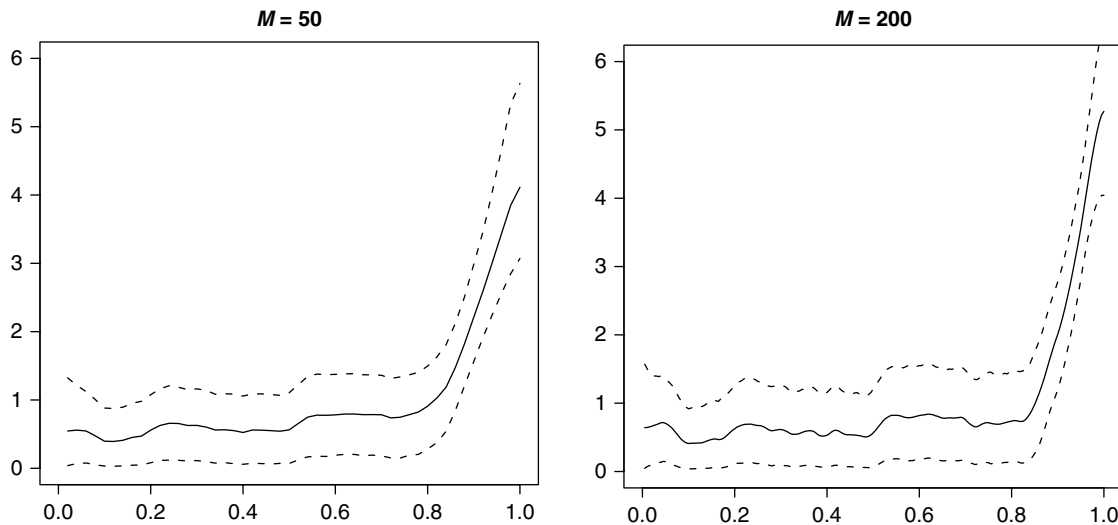
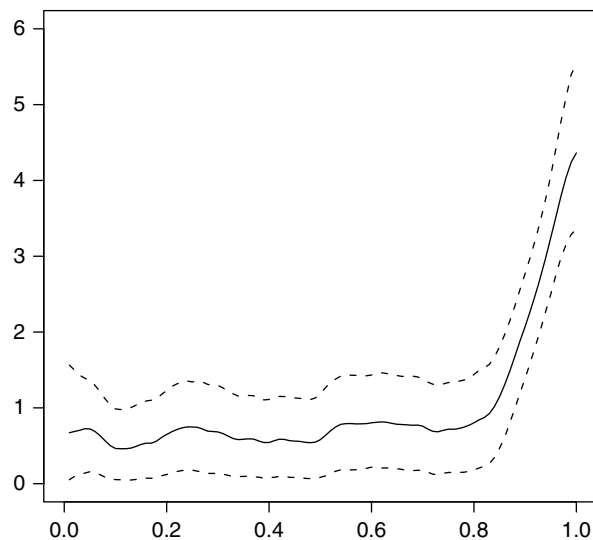


Figure A.2 Robustness Check Using an Ordered Probit Model



the MCMC procedure described in Albert and Chib (1993). The estimated $h^*(t)$ under the cutpoint model is shown in Figure A.2.

Furthermore, we also computed the aggregate weights on each quintile; the aggregate weights of the first to fifth quintiles are 0.117, 0.130, 0.133, 0.151, and 0.469, respectively. Clearly, all the results remain substantially unchanged.

References

- Aaker DA, Stayman DM, Hagerty MR (1986) Warmth in advertising: Measurement, impact, and sequence effects. *J. Consumer Res.* 12(4):365–381.
- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* 88(422):669–679.
- Ariely D (1998) Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *J. Behav. Decision Making* 11(1):19–45.
- Ariely D, Zauberman G (2000) On the making of an experience: The effects of breaking and combining experiences in their overall evaluation. *J. Behav. Decision Making* 13(2):219–232.
- Baumgartner H, Sujan M, Padgett D (1997) Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgement. *J. Marketing Res.* 34(2):219–232.
- Blacker IR (1998) *The Elements of Screenwriting* (Macmillan, New York).
- Cardot H, Ferraty F, Sarda P (1999) Functional linear model. *Statist. Prob. Lett.* 45(1):11–22.
- Cardot H, Ferraty F, Sarda P (2003) Spline estimators for the functional linear model. *Statistica Sinica* 13:571–591.
- Casella G, Berger RL (1992) *Statistical Inference*, 2nd ed. (Duxbury, Pacific Grove, CA).
- Choi J, Hui SK, Bell DR (2010) Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. *J. Marketing Res.* 47(1):75–89.
- Corduneanu C (1991) *Integral Equations and Applications* (Cambridge University Press, Cambridge, UK).
- Correll J, Spencer SJ, Zanna MP (2004) An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *J. Experiment. Soc. Psychol.* 40(3):350–356.
- Dialsmith (2013) Essentials of moment-to-moment research. Report, Dialsmith, Beaverton, OR. http://dialsmith.com/_docs/Dialsmith-Essentials-of-MtM-eBook-070813.pdf.
- Evgeniou T, Pontil M, Toubia O (2007) A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Sci.* 26(6):805–818.
- Foutz NZ, Jank W (2010) Prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Sci.* 29(3):568–579.
- Fredrickson BL, Kahneman D (1993) Duration neglect in retrospective evaluations of affective episodes. *J. Personality Soc. Psychol.* 65(1):45–55.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2003) *Bayesian Data Analysis*, 2nd ed. (Chapman & Hall/CRC Press, Boca Raton, FL).
- Greene WH (2011) *Econometric Analysis*, 7th ed. (Prentice Hall, Upper Saddle River, NJ).
- Hastie T, Mallows C (1993) [A statistical view of some chemometrics regression tools]: Discussion. *Technometrics* 35(2):140–143.
- Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York).
- Hays S, Shen H, Huang JZ (2012) Functional dynamic factor models with application to yield curve forecasting. *Ann. Appl. Statist.* 6(3):870–894.
- Hughes GD (1992) Realtime response measures redefine advertising wearout. *J. Advertising Res.* 32(3):61–77.
- James GM, Wang J, Zhu J (2009) Functional linear regression that's interpretable. *Ann. Statist.* 37(5A):2083–2108.
- Jank W, Shmueli G (2006) Functional data analysis in electronic commerce research. *Statist. Sci.* 21(2):155–166.
- Johnson SB (2005) *Everything Bad Is Good for You: How Today's Popular Culture Is Actually Making Us Smarter* (Riverhead, New York).
- Kahneman D, Fredrickson B, Schreiber C, Redelmeier D (1993) When more pain is preferred to less: Adding a better end. *Psych. Sci.* 4(6):401–405.
- Kass RE, Raftery AE (1995) Bayes factor. *J. Amer. Statist. Assoc.* 90(430):773–795.
- Kirk R, Shril D (2011) A digital agora: Citizen participation in the 2008 presidential debates. *Amer. Behav. Scientist* 55(3):325–347.
- Lewis B, Lewis D, Cumming G (1995) Frequent measurement of chronic pain: An electronic diary and empirical findings. *Pain* 60(3):341–347.
- Madrigal R, Bee C (2005) Suspense as an experience of mixed emotions: Feelings of hope and fear while watching suspenseful commercials. *Adv. Consumer Res.* 32:561–567.
- Miron-Shatz T (2009) Evaluating multiepisode events: Boundary conditions for the peak-end rule. *Emotion* 9(2):206–213.
- Montgomery NV, Unnava HR (2009) Temporal sequence effects: A memory framework. *J. Consumer Res.* 36(1):83–92.
- Nelson LD, Meyvis T, Galak J (2009) Enhancing the television viewing experience through commercial interruptions. *J. Consumer Res.* 36(August):160–172.
- Pham MT, Cohen JB, Pracejus JW, Hughes GD (2001) Affect monitoring and the primacy of feelings in judgment. *J. Consumer Res.* 28(September):167–188.
- Polsfuss M, Hess M (1991) Liking through moment-to-moment evaluation: Identifying key selling segments in advertising. Holman RH, Solomon MR, eds. *Advances in Consumer Research*, Vol. 18 (Association of Consumer Research, Provo, UT), 540–544.
- Ramanathan S, McGill AL (2007) Consuming with others: Social influence on moment-to-moment and retrospective evaluations of an experience. *J. Consumer Res.* 34(4):506–524.
- Ramsay JO, Li X (1998) Curve registration. *J. Roy. Statist. Soc. Ser. B* 60(2):351–363.

- Ramsay JO, Silverman BW (2002) *Applied Functional Data Analysis: Methods and Case Studies* (Springer, New York).
- Ramsay JO, Silverman BW (2005) *Functional Data Analysis*, 2nd ed. (Springer, New York).
- Redelmeier DA, Kahneman D (1996) Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66(1):3–8.
- Redelmeier DA, Katz J, Kahneman D (2003) Memories of colonoscopy: A randomized trial. *Pain* 104(1–2):187–194.
- Reithinger F, Jank W, Tutz G, Shumeli G (2008) Smoothing sparse and unevenly sampled curves using semiparametric mixed models: An application to online auctions. *J. Roy. Statist. Soc. Ser. C* 57(2):127–148.
- Robinson E, Blissett J, Higgs S (2011) Peak and end effects on remembered enjoyment in low and high restrained eaters. *Appetite* 57(1):207–212.
- Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, Chichester, UK).
- Schreiber CA, Kahneman D (2000) Determinants of the remembered utility of aversive sounds. *J. Experiment. Psych.: General* 129(1):27–42.
- Sood A, James GM, Tellis GJ (2009) Functional regression: A new model for predicting market penetration of new products. *Marketing Sci.* 28(1):36–51.
- Teixeira T, Wedel M, Pieters R (2012) Emotion-induced engagement in Internet video advertisements. *J. Marketing Res.* 49(2):144–159.
- Thorson E, Friestad M (1989) The effects of emotion on episodic memory for television commercials. Cafferata P, Tybout A, eds. *Cognitive and Affective Responses to Advertising* (Lexington Books, Lexington, MA), 305–325.
- Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence from faces predict election outcomes. *Science* 308(5728):1623–1626.
- Tsay C-J (2013) Sight over sound in the judgment of music performance. *Proc. Natl. Acad. Sci. USA* 110(36):14580–14585.
- Vanden Abeele P, MacLachlan DJ (1994) Process tracing of emotional responses to TV ads: Revisiting the warmth monitor. *J. Consumer Res.* 20(4):586–600.
- Varey C, Kahneman D (1992) Experiences extended across time: Evaluation of moments and episodes. *J. Behav. Decision Making* 5(3):169–185.
- Wang S, Jank W, Shmueli G, Smith P (2008) Explaining and forecasting online auction prices and their dynamics using functional data analysis. *J. Bus. Econom. Statist.* 26(2):144–160.
- West M, Harrison J (1997) *Bayesian Forecasting and Dynamic Models*, 2nd ed. (Springer, New York).
- Wirtz D, Kruger J, Scollon CN, Diener E (2003) The role of predicted on-line, and remembered experience in future choice. *Psych. Sci.* 14(5):520–524.
- Woltman Elpers JLCM, Mukherjee A, Hoyer WD (2004) Humor in television advertising: A moment-to-moment analysis. *J. Consumer Res.* 31(3):592–598.
- Woltman Elpers JLCM, Wedel M, Pieters R (2003) Why do consumers stop viewing television commercials? Two experiments on the influence of moment-to-moment entertainment and information value. *J. Marketing Res.* 40(4):437–453.
- Zhou L, Huang J, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ (2010) Reduced rank mixed effects models for spatially correlated hierarchical functional data. *J. Amer. Statist. Assoc.* 105(489):390–400.