



Marketing Science Institute Working Paper Series 2019
Report No. 19-105

A General Method for Estimating Asynchronous Dynamic Models: A Novel Study of 100 Ad Creatives

Edlira Shehu, Daniel Zantedeschi, and Prasad A. Naik

"A General Method for Estimating Asynchronous Dynamic Models: A Novel Study of 100 Ad Creatives" © 2019 Edlira Shehu, Daniel Zantedeschi, and Prasad A. Naik

MSI working papers are distributed for the benefit of MSI corporate and academic members and the general public. Reports are not to be reproduced or published in any form or by any means, electronic or mechanical, without written permission.

**A General Method for Estimating Asynchronous Dynamic Models:
A Novel Study of 100 Ad Creatives**

By

Edlira Shehu*

Associate Professor of Marketing
Department of Marketing
Copenhagen Business School
Solbjerg Plads 3
2000 Frederiksberg, Denmark
Email: es.marktg@cbs.dk

Daniel Zantedeschi

Assistant Professor of Marketing
Fisher College of Business and Translational Data Analytics
The Ohio State University
536 Fisher Hall, 2100 Neil Ave
Columbus, OH 43210
Email: zantedeschi.1@osu.edu

Prasad A. Naik

Professor of Marketing
Graduate School of Management
University of California Davis
One Shields Avenue
3314 Gallagher Hall
Davis, CA 95616
Email: panaik@ucdavis.edu

*Corresponding author

Acknowledgments: The authors thank seminar participants at the McGill University, the Ohio State University, 14th Marketing Dynamics Conference, and 39th ISMS Marketing Science Conference for their helpful suggestions. They gratefully acknowledge the funding and feedback received from the Marketing Science Institute. They also thank the company MetrixLab for providing the data used in this project. The last author acknowledges the financial support received from the UC Davis travel and small research grants program across 2015-18.

A General Method for Estimation and Inference of Asynchronous Dynamic Models: A Novel Study of 100 Ad Creatives

Abstract

Virtually all time series in business and economics evolve asynchronously. This unequal frequency of multiple time series raises the question: how to estimate and infer dynamic models when metrics evolve at different time scales? The unavailability of metrics at regular intervals introduces additional uncertainty, which we incorporate in the proposed method and offer three types of generality: data availability, model specifications, and error distributions. Data availability can follow any pattern from systematic unavailability to occasionally missing observations in multivariate y - or x -variables over time. Model specifications can belong to a broad class of state space models beyond the standard regression, VAR, or dynamic factor models. Error terms can follow any distribution (not just Normal) with finite first two moments. In Proposition 1, we derive the optimal gain factor to estimate and infer asynchronous dynamic models. Simulation studies furnish evidence that the proposed method recovers model parameters accurately and efficiently. Empirically, we analyze 100 video ads, where the ad content evolves once every 5 seconds and ad liking every second. We provide diagnostic information to managers on which video scene(s) to edit and which ad content to modify. The meta-analysis of 100 ad creatives, which is the largest study of its kind, reveals that the ad content effects generalize across industry sectors, while their heterogeneity relates to the narrative elements, known as the plot structures, which moderate the content effects.

1. Introduction

A storyboard visualizes scripts into scenes. Like structural drawings of buildings, it offers a pre-production proof-of-concept of the intended video ads. Ad agencies distill their client's brand strategy, state it in scripts known as creative briefs, create one or more storyboards, and pitches them to the client's brand management team, whose approval then initiates funding for the production of video ads. Here's an example of the scenes from the storyboard and the resulting video ad of Volkswagen: click <https://vimeo.com/10298650>.

Previous studies investigated the effects of an ad's creative elements using moment-to-moment data but ignored analyzing it at the scene-by-scene level. Such analyses proceeded in two steps: in the first step, a few salient features from moment-to-moment liking (e.g., start, end, peak, trough, trend, duration) are selected and, in the second step, a cross-sectional regression model estimated their effects on outcomes such as overall liking (Baumgartner, Sujan and Padgett 1997) or purchase intent (Teixeira, Picard, and Kaliouby, 2014). The two steps can also be combined, via functional data analysis, where the first step integrates out (see Hui, Meyvis, and Assael, 2014, Equation 1, p. 226) the moment-to-moment series instead of pre-selecting certain instants. Yet, these analyses remain inherently *cross-sectional*, relying on the variation across *multiple ads* to estimate the effects. Consequently, copywriters do not get diagnostic information to edit the specific scenes of a *single* video ad. In other words, because previous studies analyze a sample of multiple ads to estimate ad content effects, they cannot recommend diagnostics to edit the focal video ad. Furthermore, a specific scene may be disliked due to multiple characteristics of the scene itself, and so copywriters need to know what it is about that scene that they need to edit: should they make that scene more entertaining or reduce its irritation? Thus, the extant approaches cannot provide diagnostic information on two fronts: (1) the identification of specific scenes for

editing of a focal video ad, and (2) inference of one or more scene characteristics that needs re-visualization.

The main reason why extant approaches cannot yield such diagnostic information is that scenes and liking evolve at unequal frequencies. A scene, represented by multiple ad characteristics $x_l = (\text{entertainment, familiarity, irritation, relevance, stimulation, warmth})'$, furnishes a viewer's reactions at the instants $l = 5^{th}, 10^{th}, 15^{th}, 20^{th}, 25^{th}, 30^{th}$ second of an ad. Whereas moment-to-moment liking, represented by the scalar y_t , are available every second $t = 1, \dots, 30$ of the ad. The standard time series analysis requires observations to arrive at the same frequency $t = l$ for both the x_l and y_t series. In contrast, to conduct a scene-by-scene analysis, we need an approach that tackles *multiple asynchronous* time series with $t \neq l$: some variables move slower or faster than others. In the econometrics literature, Ghysels, Sinko and Valkanov (2007) proposed the mixed data sampling approach to analyzing the impact of fast-moving x -variables (e.g., weekly inflation) on slow-moving y -variables (e.g., quarterly gross domestic product). However, our empirical situation is reverse: regressors (scenes) evolve slower than the outcome (liking). The absence of a method to analyze such time series has impeded the progress in estimating asynchronous scene effects on ad liking.

Besides video ads as an application context, asynchronous time series arise in numerous business and economic contexts. For example, the daily yield curve and monthly consumer confidence may affect the quarterly gross domestic product and the annual recession probability; or monthly advertising spends available from Kantar media drives the weekly ad awareness available from Millward Brown; or the fast-moving online activities influence the slow-moving attitudinal measures; or daily online search may influence weekly purchase intentions which, in turn, affects monthly sales. To accommodate all such possibilities, we develop a general method to analyze multiple asynchronous time series

consisting of (i) slow-moving x -variables (relative to the y -variable), (ii) fast-moving x -variables, (iii) multiple y -variables at same or unequal frequencies (e.g., weekly offline sales and hourly online sales), or (iv) multiple x -variables at same or unequal frequencies (e.g., monthly ad spending, daily click-through rates). In other words, it accommodates *all the possibilities* across multivariate time-series from multiple dependent variables and multiple regressors $(x_{1,l}, x_{2,l'}, y_{1,t}, y_{2,t'})$ with either same or unequal frequencies (l, l', t, t') .

The proposed method nests, i.e., includes as a special case, any dynamic linear model such as vector autoregression models (e.g., Lütkepohl 2005), ARIMA models (e.g., Brockwell and Davis 1996), dynamic regressions (e.g., Biyalogorsky and Naik 2003), dynamic factor models (e.g., Bruce, Peters and Naik 2012) or hierarchical linear models (e.g., Raudenbush and Bryk 2002). Thus, the proposed method estimates and infers a broad class of dynamic models based on multivariate asynchronous time series data.

In summary, this study makes four contributions. First and foremost, we develop a general method to estimate and infer dynamic models using asynchronous time series data. Second, we demonstrate its efficacy in recovering the true parameters via simulation studies, which mimic the empirical setting. The estimated parameters are close to the true values at varying levels of data sparsity (e.g., scene reactions observed at every three- versus five-second intervals). Third, using Google's thirty-seconds ad and Apple's sixty-seconds ads, we illustrate how the proposed method estimates the ad content effects of *an* individual advertisement uncontaminated by the presence of other ads in the estimation sample. The results identify not only the specific scenes in the ad to be edited, but also the significant ad characteristics (e.g., entertainment or irritation) for re-visualization. Last but not least, we conduct a meta-analysis of 100 video ads that not only generalizes the findings across multiple industries, but also sheds light on how ad's plot structure shapes the content effects.

The rest of the paper proceeds as follows. Section 3 derives the asynchronous filter; Section 4 presents the simulation studies; Section 5 applies the asynchronous filter to individual Google or Apple ads and conducts a meta-analysis of 100 video ads. Section 6 concludes, but first we review the extant literature.

2. Literature Review

2.1. Advertising Content and Liking

Advertising agencies use audience reactions to evaluate ads. An ad's content has a complex structure with visual and auditory elements telling a story in 30 seconds of commercial time, and so a few dimensions may not be adequate to understand the full effect (Holbrook and Batra 1987). Consequently, several studies have developed multiple scales to measure different characteristics of an ad. For example, Wells, Leavitt, and McConville (1971) and Schlinger (1979), in collaboration with Leo Burnett agency, used multiple items, which were then factor analyzed to identify ad characteristics. Wells et al. (1971) found six characteristics (humorous, relevant, irritating, sensual, vigorous, and unique), whereas Schlinger (1979) identified seven characteristics, four of which were common to Wells et al. (1971) and the other three were familiar, confusing, and brand reinforcing. Other studies have identified four (e.g., Moldovan 1984) to nine (Aaker and Stayman 1990) characteristics, although many are common across studies such as entertainment, relevance, stimulation, or warmth. More recently, Smit, van Meurs and Neijens (2006) summarize various characteristics from prior research into the following six dimensions: entertaining (e.g., clever, humorous), stimulating (e.g., lively), relevant (e.g., informative, believable, meaningful), warm (e.g., sensual, sensitive), irritating (e.g., tasteless, confusing), and familiar. In our empirical application, we use this set of six characteristics of ad content.

Previous research has also established that ad characteristics influence ad liking. Specifically, warmth and relevance (Zinkham and Fornell 1985; Aaker and Staymann 1990;

Biel and Bridgewater 1990), entertainment (Aaker and Staymann 1990; Smit, van Meurs and Neijens 2006), and stimulation (Zinkham and Fornell 1985; Aaker and Staymann 1990) positively influence ad liking; whereas irritation (Aaker and Stayman 1990; Biel and Bridgewater 1990) and familiarity (Zinkham and Fornell 1985; Aaker and Stayman 1990; Smit, van Neurs and Neijens 2006) negatively influence ad liking.

These studies, however, ignored the dynamics of ad liking and scene-by-scene characteristics. That is, the persistence in the flow of liking reinforces future ad liking. Also, an ad's storyline consists of a sequence of scenes, which may focus on different characteristics (Loewenstein, Raghunathan, and Heath 2011). For example, an ad may begin with an entertaining scene and end up by establishing product relevance. In contrast to our application, previous studies did not capture time-varying scene-by-scene characteristics nor did they measure how moment-to-moment ad liking unfolds over the thirty-second duration of commercials.

2.2. Moment-to-Moment Metrics

Recent studies use moment-to-moment metrics, for example, Baumgartner, Sujan and Padgett (1997) used a computerized "feelings monitor" to collect moment-to-moment liking, and analyze how beginning, peak, average, and the end values affect overall ad liking. They found that that overall liking correlates with peak liking and the liking for the ending sequences. Elpers, Mukherjee and Hoyer (2004) investigated how moment-to-moment humor and surprise affect overall humor perception. They found that overall humor perceptions correlate positively to peak humor and surprise. Madrigal and Bee (2005) collected moment-to-moment fear and hope and related their average and peak values to ad suspense experience from two groups of suspenseful and non-suspenseful ads. They found that both these emotions underpin the suspense experience. Elpers, Wedel, and Pieters (2003) analyze the influence of moment-to-moment information and entertaining values on zapping behavior.

They apply smoothing functional data analysis to moment-to-moment time series, compute its level and velocity (first derivative), and estimate the hazard probability of discontinuing to view an ad. They found that information and entertainment exert a positive interaction effect on the zapping probability. Similarly, Teixeira, Wedel, and Pieters (2012) measure joy and surprise using the automated extraction of facial expression and attention dispersion using eye fixation data from the ad. They found that surprise and joy effectively concentrate attention and retain viewers. Teixeira, Picard, and Kaliouby (2014) analyze how entertainment, measured by facial recognition, affects zapping behavior and purchase intention. They found an inverted U-shape relationship of entertainment on purchase intention. Hui, Meyvis, and Assael (2014) compute a temporally weighted average of moment-to-moment evaluations of nineteen television programs and find support for the ending but not the peak effect. Finally, Liu, Shi, Teixeira, and Wedel (2018) use facial expression data to measure viewers' emotional response to trailers of comedy movies and derive measures of start, trend, peak, and end for happiness, surprise and disgust time series data. They aggregate high frequency variables to match the frequency of slow-moving viewing intention and box office sales without solving the problem of asynchronous time series estimation.

In sum, this literature relies on pre-specified instants (e.g., peak or end), does not estimate models of a single ad's multiple metrics over time and, most importantly, ignores the asynchronicity that arises when multiple time series are observed at different frequencies. To this end, we next review studies from econometrics on "mixed frequency" estimation.

2.3. Mixed Frequency Estimation

Virtually all time series data are collected at different frequencies by different data sources. For example, we observe quarterly gross domestic product, monthly personal income or non-farm employment, daily or intra-day financial series (e.g., stock prices), hourly clickstream data, weekly brand awareness, monthly ad spending, to mention a few asynchronous

variables. Consequently, analyzing mixed frequency data is an active research area (see the *Special Issue (Volume 193, 2016)* of the *Journal of Econometrics*). Below we report studies using mixed frequency regression-based models, vector autoregression (VAR) models, ARIMA models, and dynamic factor models in the econometrics literature (also see the survey by Foroni and Marcellino 2013).

Consider mixed frequency data on monthly sales S_t and weekly awareness A_l , where the slow-moving $t = 1, \dots, 13$, and the fast-moving $l = t, t - 1, t - 2, t - 3$ denotes the current and past three weeks within the month t . The gist of the idea behind MIDAS regressions (Ghysels, Sinko and Valkanov 2007), an acronym for mixed data sampling, is to (1) construct a variable $X_t = w_0 A_t + w_1 A_{t-1} + w_2 A_{t-2} + w_3 A_{t-3} = \sum_{k=0}^3 w_k A_{t-k}$, where A_{t-k} denotes awareness in week k of the month t , and then (2) estimate the standard regression model $S_t = \alpha + \beta X_t + \epsilon_t$. To uniquely identify β , the weights w_k need to sum to unity. The weights w_k to be estimated proliferate as A_l is observed at a greater granularity (say daily) and so, to conserve the degrees of freedom, a parametric structure such as the Almon polynomial is imposed (Almon 1965).

The main drawbacks of regression-based models pertain to the absence of errors-in-variables or time-varying effects. A single error-prone independent variable can render all parameter estimates inconsistent, which means even asymptotically large sample sizes do not guarantee that the estimated parameters converge to their true values (see Naik and Tsai 2000). Such error-prone variables are common: as Bernanke, Boivin and Elias (2005, p. 403) noted: "... the choice of a specific data series to represent a general economic concept (e.g., industrial production for "economic activity," the consumer price index for "the price level") is often arbitrary to some degree; measurement errors and revisions pose additional problems...".

Equally important, the effect β need not be constant over time; stochastically time-varying effects β_t are difficult to incorporate in regression-based models. To incorporate measurement errors or stochastic time-varying parameters, state space models are eminently suited (see Harvey 1994). For a discussion of mixed-frequency state space and regression-based models, see Bai, Ghysels and Wright (2013).

To incorporate multivariate Y_t , VAR models are formulated, which are then cast into the state space framework for estimation, inference, and forecasting via the Kalman filter (see Harvey 1994). For example, Mittnik and Zdrozny (2004) estimate German GDP at monthly frequencies using quarterly data by applying Kalman filtering and maximum likelihood estimation; Chiu et al. (2011) estimate mixed-frequency VAR by applying Kalman filtering and Bayesian estimation.

Like VAR models, ARIMA models are also nested in the general state space framework that consists of the two equations, an observation equation and the transition equation, where the latter captures the evolution of a dynamic system of unobservable constructs, which are linked to potentially error-prone observed time series data via the observation equation. The celebrated Kalman filtering recursions furnish the prior and posterior means and covariances of the unobservable constructs, assuming model parameters are known. To estimate unknown parameters, Schweppe (1965) was the first study to evaluate the likelihood function for the state space model to estimate parameters and infer significance. Jones (1980) is the first study to incorporate “missing” data for a scalar dependent variable in ARMA models, which was extended by Ansley and Kohn (1983) for multivariate ARMA models with time varying parameters. Harvey and Pierse (1984) then extended these ideas to encompass scalar ARMA models with integrated time series (ARIMA), while Zdrozny (1990) estimated bivariate ARMA models using quarterly GDP and monthly employment time series.

Like VAR and ARIMA models, dynamic factor models are special cases of the state space framework. For example, Mariano and Murasawa (2003) formulate a dynamic factor model, cast it into the state space framework, and create “coincident index” using quarterly GDP and monthly series on employment, income, industrial production, and manufacturing and trade sales. Aruoba, Diebold, and Scotti (2009) formulate a dynamic factor model via the state space representation and estimate it by computing the log-likelihood using Kalman filtering recursions and maximizing the log-likelihood to infer the unobservable construct “business conditions” in real time (i.e., every day). Giannone, Reichlin and Small (2008) estimate the “nowcast” of the quarterly GDP, which denotes the filtered estimate of the quarterly GDP as new information arrives from any of the two hundred time-series indicators of the economy. To tackle mixed frequency, they fix the variances of the error terms of the unavailable data to infinity so that the Kalman gain factor places zero weights when updating the posterior means and covariances (also see Doz, Giannone and Reichlin 2011). Finally, to estimate dynamic factor models with an arbitrary pattern of missing data, Banbura and Modugno (2014) apply expectation-maximization algorithm, which was first developed by Shumway and Stoffer (1982), but it lacks the capability to provide statistical significance.

In sum, the extant literature in econometrics has focused on the special cases (VAR, ARIMA, dynamic factor models) of the state space models, which encompass a much wider class of dynamic linear models (e.g., VAR with time-varying parameters, Dynamic Factor Models with time-varying parameters, Hierarchical Linear Models). Even the dynamic model in our empirical application is not nested in the econometrics models reviewed here. This raises the question: how to estimate *general* state space models—not just its three special cases—using asynchronous time-series data? The asynchronicity from an arbitrary pattern of data availability injects additional uncertainty. How should it be incorporated systematically in the estimation and inference engine? The next section examines these open questions.

3. Optimal Asynchronous Filter

3.1. Asynchronous Dynamic Model of Ad Liking

How do scene-by-scene reactions to an advertisement influence its likeability? As the Volkswagen ad exemplifies in the Introduction, each scene enacts the plot intended in a storyboard. Every scene evokes different intensity of reactions to how entertaining, stimulating, relevant, warm, familiar, or irritating that scene is, and builds the flow of liking towards an advertisement over its span of 30 seconds. Let L_t denote ad liking at instant $t = 1, \dots, T = 30$, and $\{x_{it}\}$ for $i = 1, \dots, 6$ are the ad characteristics such as entertaining or familiarity, and $\{\beta_i\}$ represent their effects on liking as follows:

$$L_{t+1} = \lambda L_t + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \beta_6 x_{6t} + v_t, \quad (1)$$

where the error term v_t follows $N(0, \sigma_v^2)$, and λ measures how the flow of liking builds up over time. Equation (1) can be re-expressed as $\Delta L_t = L_{t+1} - L_t = -\delta L_t + \sum_{i=1}^6 \beta_i x_{it} + v_t$, where the decay rate $\delta = 1 - \lambda$. Then it reveals that the change in liking decays at the rate proportional to its level, and this change is driven by the reactions to the scenes in an ad.

We encounter three novel challenges to estimate equation (1) using observed data. First, the liking L_t is not directly observable because respondents provide error-prone measures of liking y_t . Second, the ad characteristics x_{il} are ratings that also contain measurement errors. Third, the observed liking y_t are available at every second $t = 1, \dots, N = 30$, whereas x_{il} are observed at $l = 5, 10, 15, 20, 25, 30$ seconds because the comprehension of scenes takes time. This last point induces asynchronous time-series: liking y_t observed every second, and scene-by-scene reactions $\{x_{il}\}$ observed at five-second intervals.

Consequently, standard regression cannot be used to estimate the model $y_{t+1} = \lambda y_t + \sum_{i=1}^6 \beta_i x_{it} + v_t$ because it requires $t = l$. Secondly, with $t = l$, the standard regression cannot estimate the model $y_{l+1} = \lambda y_l + \sum_{i=1}^6 \beta_i x_{il} + v_l$ with 8 parameters $(\lambda, \{\beta_i\}, \sigma_v)'$ using $l =$

1, ..., 6 observations from a typical 30-second ad. Finally, the measurement errors in the ratings x_{it} lead to inconsistent estimation (Naik and Tsai 2000). Hence, we next develop a new method that incorporates asynchronicity and measurement errors.

3.2. Deletion Matrix

We first control the presence of errors in the metrics $(y_t, x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}, x_{6t})' = Y_t'$ via the observation equation:

$$\underbrace{\begin{bmatrix} y_t \\ x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \\ x_{6t} \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix}}_Z \underbrace{\begin{bmatrix} L_t \\ \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \\ \mu_{4t} \\ \mu_{5t} \\ \mu_{6t} \end{bmatrix}}_{\alpha_t} + \underbrace{\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \\ \epsilon_{5t} \\ \epsilon_{6t} \\ \epsilon_{7t} \end{bmatrix}}_{\epsilon_t}, \quad (2)$$

where Z is the link matrix, α_t is the state vector, and ϵ_t denotes the errors that follow $N(0, H)$ with the covariance matrix H . Then we incorporate equation (1) in the first row of the transition equation as follows:

$$\underbrace{\begin{bmatrix} L_{t+1} \\ \mu_{1,t+1} \\ \mu_{2,t+1} \\ \mu_{3,t+1} \\ \mu_{4,t+1} \\ \mu_{5,t+1} \\ \mu_{6,t+1} \end{bmatrix}}_{\alpha_{t+1}} = \underbrace{\begin{bmatrix} \lambda & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix}}_T \underbrace{\begin{bmatrix} L_t \\ \mu_{1t} \\ \mu_{2t} \\ \mu_{3t} \\ \mu_{4t} \\ \mu_{5t} \\ \mu_{6t} \end{bmatrix}}_{\alpha_t} + \underbrace{\begin{bmatrix} \nu_{1t} \\ \nu_{2t} \\ \nu_{3t} \\ \nu_{4t} \\ \nu_{5t} \\ \nu_{6t} \\ \nu_{7t} \end{bmatrix}}_{\nu_t}, \quad (3)$$

where T is the transition matrix, and ν_t denotes the error terms that follow $N(0, Q)$ with the covariance matrix Q .

More generally, we represent the above equations in the vector-matrix form:

$$Y_t = Z_t \alpha_t + c_t + \epsilon_t \quad (4)$$

$$\alpha_{t+1} = T_t \alpha_t + d_t + \nu_t \quad (5)$$

The dimension of the observation vector Y_t is $m \times 1$, the link matrix Z_t is $m \times n$, the state vector α_t is $n \times 1$, the transition matrix T_t is $n \times n$, (c_t, d_t) are vectors of dimensions $m \times 1$

and $n \times 1$, respectively, and the error terms follow multivariate normal $\epsilon_t \sim N(0, H_t)$ and $v_t \sim N(0, Q_t)$ with the covariance matrices of dimensions $m \times m$ and $n \times n$, respectively.

Next, we capture asynchronicity in time series: not all the elements of Y_t are observed at every t . For example, in our empirical setting, we observe y_t for every t and $(x_{1t}, \dots, x_{6t})'$ for the specific instants $t = 5, 10, 15, 20, 25, 30$. Note that observations at the instants $t = 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29$ are not missing—they are systematically unavailable because the comprehension of scenes takes about 5 seconds. Also “unavailable” does not imply that a zero value is observed instead. No value, zero or otherwise, is observed for the many instants of the thirty-seconds span.

To incorporate asynchronicity, we define the deletion matrix A_t as follows: when all the elements of Y_t are observed, it equals an identity matrix conformable with Y_t . When some elements of Y_t are not available, it equals an identity matrix, conformable with Y_t , whose rows corresponding to the unavailable elements are deleted. The resulting dimension of A_t is $m_t \times m$, where m_t depends on how many elements in Y_t are available at an instant t .

For example, at $t = 1$, if only the first element is available, say $Y_t = (1, *, *, *, *, *, *)'$, where $*$ denotes unavailable data, the deletion of the rows two through seven of an identity matrix yields A_t as follows:

$$I_7 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow A_t = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

We strikethrough the rows 2, 3, 4, 5, 6 and 7 because those elements in Y_t are not available at $t = 1$ and keep only the first row because the element y_t is available; hence $m_t = 1$. If all the seven elements were available, say at $t = 5$, $Y_t = (1, 2, 3, 4, 5, 6, 7)'$, then $A_t = I_7$.

$$I_7 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

No row is deleted because all the seven elements of Y_t are available; hence $m_t = 7$.

This deletion principle is general: *Delete the rows corresponding to the unavailable elements in the observation vector Y_t .* More generally suppose the 2nd and 5th elements of Y_t were unavailable so that $Y_t = (1, *, 3, 4, *, 6, 7)'$, then the deletion matrix would be

$$I_7 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The resulting A_t does not contain the second and fifth rows. Also, the columns in A_t corresponding to the unavailable elements in Y_t are zeros. Consequently, the deletion matrix maps the observation vector Y_t to the available data vector D_t :

$$D_t = A_t Y_t, \quad (6)$$

where the dimension of D_t is $m_t \times 1$, which varies at each t depending on which elements in Y_t are available at that instant t .

3.3. Optimal Asynchronous Gain Factor

The unavailability of certain elements in Y_t introduces additional uncertainty in the evolution of the state vector α_t , raising the central question: What is the optimal asynchronous gain factor for the general state space model based on the equations (4), (5), and (6)? The following proposition provides the answer.

Proposition 1: *The optimal gain matrix for any asynchronous state space model specified by equations (4), (5) and (6) is given by*

$$G_t^* = P_{t|t-1} Z_t' A_t' (A_t Z_t P_{t|t-1} Z_t' A_t' + A_t R A_t')^{-1}, \quad (7)$$

where G_t^* is $n \times m_t$ matrix, and m_t is the row dimension of the $m_t \times m$ deletion matrix A_t .

Proof. See the Appendix.

A few remarks are in order. First and foremost, the uncertainty due to unavailability of certain elements of Y_t is incorporated, via the deletion matrix A_t , in the mean and covariance of the state vector. Equations (A6) and (A8) in Appendix A reveal that the posterior mean and covariance depend on the gain matrix G_t^* , which depends on the deletion matrix (as shown in equation 7). Second, Table 1 presents the algorithm for estimation and inference based on the derived optimal gain matrix. In Table 1, the asynchronous filter recursions directly yield the log-likelihood, whose maximization furnishes parameter estimates and statistical significance (see Step 3 in Table 1). Third, the proof in Appendix A does *not* make any assumption that the error terms (ϵ_t, v_t) are normally distributed. Hence, the derived asynchronous filter recursions in Table 1 are robust to *any* distribution of the error terms with finite first two moments.

When no observations are available in some time periods, any conformable matrix G_t serves as the gain matrix because the error is $(D_t - \hat{D}_t) = 0$. Consequently, the mean of the state vector $a_t = a_{t|t-1}$ without any theoretical or numerical issues even when *all* the observations at some instants are unavailable. Its main implication is that the data vector D_t need not be equispaced over time. In other words, irregularly observed time series requires no additional theory or algorithms. Fifth, when all observations are available in some time periods, $A_t = I_m$, and so the asynchronous gain matrix in (7) equals the celebrated Kalman gain factor. Thus, the optimal asynchronous filter nests the standard Kalman filter. Last but not least, when *some* of the elements of Y_t are unavailable, the asynchronous filter works because the gain matrix in (7) and the available data vector D_t are conformable in the

measurement update equation: $a_t = a_{t|t-1} + \underbrace{G_t^*}_{n \times m_t} \underbrace{(D_t - \widehat{D}_t)}_{m_t \times 1}$. Intuitively, the asynchronous

filter uses the prediction errors $(D_t - \widehat{D}_t)$ from *whatever available data* rather than halting the recursions in the absence of full availability, as would be the case with the standard Kalman filter.

3.4. Generality

The asynchronous state space model, specified by equations (4), (5) and (6), offers three levels of generality: data availability, model structure, and error distributions. The third remark above clarifies that the error terms in (4) and (5) belong to a broad class, beyond normality, such as gamma, exponential, and so on. For all distributions with the finite first two moments, the derived optimal gain factor in (7) remains the same. In other words, equation (7) generalizes to not only (infinitely) many distributions, but also different distributions for different instants via the time-varying covariance H_t .

Regarding the model structure generality, equations (4) and (5) subsume broad classes of statistical models such as the vector autoregression models (e.g., Lütkepohl 2005), ARIMA models (e.g., Brockwell and Davis 1996), time-varying parameter models (e.g., Koop and Korobilis 2010), dynamic regression (e.g., Biyalogorsky and Naik 2003), dynamic factor models (e.g., Bruce, Peters and Naik 2012) or hierarchical linear models (e.g., Raudenbush and Bryk 2002). Appendix B shows how all these models are special cases of equations (4) and (5).

Most importantly, the deletion matrix A_t is versatile in accommodating diverse possibilities of data availability: a scalar y_t arrives faster (or slower) than a scalar or multivariate $x_{i,l}$ does (as in the empirical application); multivariate $y_{i,t}$ arrive slower (or faster) than another $y_{j,l'}$; multivariate $x_{i,l}$ arrive slower (or faster) than another $x_{j,l'}$; multivariate $(y_{i,t}, y_{j,t'}, x_{m,l}, x_{n,l'})$ arrive slower (or faster) than each other; neither some of

the multivariate $y_{i,t}$ nor some of the multivariate $x_{j,l}$ are available at some of the instants (i.e., irregularly spaced data); all multivariate $y_{i,t}$ and $x_{j,l}$ are available at all t (i.e., standard models are special cases with $A_t = I$). Thus, the proposed approach estimates asynchronous dynamic models with any error distribution, model structure, and data availability pattern. We next illustrate its efficacy in recovering model parameters via simulation studies.

4. Simulation Studies

This section examines (1) the accuracy and efficiency of parameter estimation and (2) the impact of data sparsity on them. We mimic the empirical setting by letting two ad characteristics affect liking as per $y_{t+1} = \lambda y_t + \beta_1 x_{1t} + \beta_2 x_{2t} + v_t$. We set the true parameters $(\lambda, \beta_1, \beta_2)' = (0.5, 1.0, -0.5)'$. We generate x_{it} using independent random walk that starts from initial $\mu_{i0} = 3$ over $t = 1, \dots, N = 30$, with initial liking $y_0 = 0$, and all noise standard deviations of 0.5. A finite-sample bias $(\hat{\theta}_N - \theta)$ measures accuracy, and the variance of the estimated parameter $\sigma_{\hat{\theta}_i}^2$ quantifies efficiency.

We incorporate asynchronicity by making y_t available for every $t = 1, \dots, 30$ and $(x_{1l}, x_{2l})'$ at every 5th instant for $l = 5, 10, 15, 20, 25, 30$. Based on the three asynchronous time series $(y_t, x_{1l}, x_{2l})'$, we estimate equations (2) and (3) using the deletion matrix:

$$A_t = \begin{cases} [1 & 0 & 0] & \text{if } \text{mod}(t, 5) \neq 0 \\ I_3 & \text{otherwise.} \end{cases}, \quad (8)$$

where $\text{mod}(p, q)$ denotes the remainder from the division of p by q .

Table 1 presents the algorithm to implement the proposed method. Applying the steps in Table 1, we run the asynchronous filter recursions, compute the log-likelihood, maximize it to estimate the parameter values, and obtain their standard errors for statistical inference. We conducted this maximization for hundred simulated data sets and attained convergence every time.

Table 2 reports the accuracy and efficiency of estimation. The estimated values are close to their true values, and the finite-sample bias is small. As the sample size increases asymptotically, the finite-sample bias tends to zero due to the unbiasedness property of maximum likelihood estimators. We also repeated the simulation by using $\text{mod}(t, 3) \neq 0$ in equation (8) to assess the accuracy when x_{it} are available at every 3rd instant. Panel B in Table 2 reports the results. Also, we found that the estimated parameters differ from zero significantly, and the true parameters lie within the 95% confidence intervals. Thus, the proposed method estimates and infers the unknown parameters satisfactorily.

Does the accuracy degrade when data are available every 5th rather than every 3rd instant? To understand this impact of data sparsity, we regress the estimated parameter values on the dummy variable $X = \begin{cases} 1 & \text{if } \text{mod}(t, 5) = 0 \\ 0 & \text{if } \text{mod}(t, 3) \neq 0 \end{cases}$. For the carryover effect, the estimated $\hat{\lambda} = \underbrace{0.465}_{0.007} + \underbrace{0.089}_{0.010} X$ with standard errors shown below the estimates. Consequently, $\hat{\lambda}$ is closer to its true value 0.5 when $X = 1$ (i.e., data are available every 5th instant). Also, $\hat{\beta}_1 = \underbrace{1.119}_{0.015} - \underbrace{0.121}_{0.021} X$, and so $\hat{\beta}_1$ is closer to 1.0 when $X = 1$. Similarly, $\hat{\beta}_2 = -\underbrace{0.462}_{0.009} - \underbrace{0.048}_{0.013} X$, and so $\hat{\beta}_2$ is closer to -0.5 when $X = 1$. These results reveal that the accuracy does not degrade with reduced data availability from 33.3% (every third observation available) to 20% (every fifth observation available). This finding holds true because, in general, state space models accurately estimate the state vector α_t of $\dim(\alpha_t) = n$ using *reduced* data from $\dim(Y_t) = m$ when $m < n$.

Counter-intuitively, the accuracy improves marginally when $X = 1$ versus $X = 0$. To understand this finding, we note that two time series can be correlated in finite samples even when none exists theoretically, which is known as “spurious” correlations (Yule 1926). Even *independent* random walks exhibit spurious correlations in finite samples (see Ernst, Shepp

and Wyner 2016). Consequently, when the gap between subsequent observations is small, the sample correlation is large, which reduces the estimation accuracy. In other words, accuracy improves as the gap between subsequent observations increases.

We close this section by presenting the results on efficiency. In principle, the variance of parameter estimates increases as data sparsity increases. That's because a sample of 6 observations when data are observed every 5th instant contains less information than that the sample of 10 observations when data are observed every 3rd instant. Accordingly, the following regression models show that $\sigma_{\lambda}^2 = \frac{0.022}{0.004} + \frac{0.015}{0.006} X$, $\sigma_{\beta_1}^2 = \frac{0.082}{0.006} + \frac{0.055}{0.009} X$, and $\sigma_{\beta_2}^2 = \frac{0.065}{0.004} - \frac{0.041}{0.005} X$. While the first two equations comport with theory, the negative sign in the last equation, suggesting increased efficiency, is likely due to the spurious sample correlation in two independent random walks.

To summarize, the proposed approach recovers model parameters accurately and efficiently. Interestingly, accuracy does not decrease despite the reduction in sample size, whereas efficiency does as expected. We next apply the derived asynchronous filter to estimate the scene-by-scene effects on ad liking of a single video ad and conduct a meta-analysis of the results obtained from 100 such video ads from multiple industry sectors to discover generalizable insights.

5. Empirical Application

5.1. Data

We obtained proprietary data sets on 100 video ads from the five industry sectors: fast moving consumer goods (e.g., Yoplait), technology (e.g., Apple), services (e.g., Geico), Leisure (e.g., Disney), and “other” miscellaneous companies (e.g., Ikea). Figure 1 presents the distribution of ads across these sectors. Most ads have 30 seconds span, while 10% are about a minute long.

I

A total of 20,412 participants evaluated hundred ads. The number of participants per ad varied from a minimum of 62 (Google) and a maximum of 799 (Oral B), depending on the client's budget. Every participant watched and evaluated the ad by using an evaluation slider. They were instructed on how to use the slider by video demonstrations. The slider was positioned at the 0-point of the evaluation slider, which ranged from -5 ("do not like at all") to 5 ("like very much"). The liking scores were tracked every second for each ad per participant.

Ad content was measured on the six ad characteristics: entertainment, stimulation, relevance, warmth, irritation, and familiarity (Aaker and Stayman 1990; Smit, van Meurs and Neijens 2006). Two raters were instructed and practiced the rating task using test ads (not part of the data set) to develop a shared understanding of the six ad characteristics. During this practice evaluation, different scene lengths (three, four, or five seconds) were experimented. Both raters informed that scenes shorter than five seconds were difficult to comprehend. We also corroborated with experts from the advertising industry, who suggested that five seconds is a reasonable span. Hence, a scene lasts for five seconds. For each scene, we measured the intensity of six ad characteristics on the seven-point scale (1 = "not at all" to 7 = "very much"). The inter-rater reliability was high: Cronbach's alpha equals 0.663 for familiarity, 0.704 for warmth, 0.750 for stimulation, 0.844 for entertainment, 0.869 for relevance, and 0.899 for irritation. So we averaged the rater's scores to obtain the scene-by-scene metrics on ad content.

Overall, the data comprise ad liking at every second and scene-by-scene reactions on the six ad characteristics every five seconds. Figure 2 presents the metrics of ad content and liking for Google (Panel A) and Apple (Panel B) ads. Table 3 displays the descriptive statistics across 100 ads. Average liking across ads equals 0.860 with standard deviation

2.317 ranging from -4.942 to 4.966. As for ad content, relevance attains the highest average rating (Mean = 4.013), followed by familiarity (Mean = 3.755), stimulation (Mean = 3.718), entertainment (Mean = 3.632), warmth (Mean = 3.385), and irritation (Mean = 1.863).

"

5.2. Estimation and inference of the asynchronous scene-by-scene effects

Companies buy data, such as those displayed in Figure 2, with the hope of understanding the impact of ad characteristics on ad liking for *their specific ads*. Currently, they do not acquire this understanding for multiple reasons. First, the extant studies rely on the variation from other ads to estimate the content effects, thereby contaminating the effects. Second, they ignore the presence of serial correlation in ad liking. Indeed, the serial correlation in Figure 2 for ad liking is 0.957 (0.966) for the Google (Apple) ad. Third, the metrics of ad characteristics and liking emerge from the judgments of respondents and so contain measurement errors, which when ignored inject inconsistency that prevents convergence of the estimates to their true values even with asymptotic sample sizes. Finally, managers lack a framework that simultaneously incorporates ad liking dynamics, measurement errors, and the ad characteristics effects. While equations (1) and (2) offer such a framework, the theory (Proposition 1) and algorithm (Table 1) we developed tackles the asynchronicity to enable the estimation and inference of ad-specific content effects on the liking dynamics.

We apply the algorithm in Table 1 to the data in Figure 2 for Google (Panel A) and Apple (Panel B) ads to obtain the results shown in Table 4. For the Google ad, only three out of the six ad characteristics are statistically significant at the 95% confidence level.

Entertainment and familiarity positively influence ad liking, while warmth affects it negatively. The first two effects comport with existing studies (e.g., Aaker and Staymann 1990; Zinkham and Fornell 1985), whereas the warmth effect differs from those in the

previous studies perhaps because they did not control for the dynamics of ad liking and the presence of measurement errors. Thus, these findings can be used to enhance ad liking by reducing warmth levels for this Google ad.

For the Apple ad, all the six ad characteristics are statistically significant at the 95% confidence level. Entertainment, familiarity and warmth increase ad liking, whereas irritation, relevance and stimulation suppress it. The effects of entertainment, familiarity, irritation and warmth comport with the existing studies. However, the adverse effects of relevance and stimulation differ from those in the previous studies perhaps because they did not control for the dynamics of ad liking and the presence of measurement errors. The relevance comprises informational aspects which consumers may find boring. Similarly, stimulation beyond a threshold may exert a negative impact as suggested by Steenkamp, Baumgartner and van der Wulp (1996). Thus, these findings can be used to enhance ad liking for this Apple ad by reducing irritation, stimulation, and relevance levels.

For both the ads, current ad liking drives future ad liking. Both the estimated $\hat{\lambda}$ are statistically significant at the 95% confidence level. At first glance, this persistency of liking may not appear as surprising, yet *no prior study in the extant literature has demonstrated it empirically*. This outcome arises because the extant moment-to-moment studies are inherently cross-sectional, relying on the variation between ads and ignoring temporal variation within an ad. The reason persistency of liking should be included in a model is that it amplifies the impact of ad characteristics. Specifically, the instantaneous effect of an ad characteristic i on ad liking accumulates, from moment to moment, over the span of the ad, to eventually yield the *total effect* given by $\frac{\beta_i}{1-\lambda}$. In other words, the dynamic multiplier $\frac{1}{1-\lambda}$ amplifies the instantaneous effect of an ad characteristic β_i . For the Google ad, this dynamic multiplier is 20 (since $\hat{\lambda} = 0.952$); it is as high as 29 for the Apple ad (since $\hat{\lambda} = 0.966$).

Thus, the extant studies under-represent the total effect of ad characteristics on ad liking because they ignored the dynamics of ad liking.

Finally, when measurement errors are present in the metrics but ignored in the model, the estimated effects are inconsistent (Naik and Tsai 2000). Are measurement errors significantly present in ad characteristics and ad liking? Table 4 shows that the *measurement noise is significant* for both the ads. Specifically, for liking, $\hat{\sigma}_\epsilon = 0.514$ for Google and 0.366 for the Apple ad; the content noise $\hat{\sigma}_v = 0.509$ for Google and 0.273 for Apple. Hence, measurement errors in metrics should be explicitly incorporated in model specifications, as in this study, not just to ensure consistency of the estimators, but even to learn whether ignoring noise is innocuous (if noise estimates are insignificant) or not.

5.3. Editing Video Ads

Companies seek to identify specific scenes of their video ads to edit. Because multiple ad characteristics in a given scene may suppress ad liking, they also need to know which specific ad characteristic(s) to focus: should they edit a particular scene to make it more entertaining or reduce its irritation?

To furnish such diagnostic information, we first inspect the parameter estimates of the ad characteristics and identify the most negative significant characteristic based on their effect sizes. Formally, we find $i^* = \underbrace{ArgMin}_i \{\hat{\beta}_i | \hat{\beta}_i < 0\}$ over the ad characteristics $i = 1, \dots, 6$. Then we identify the specific scene that scored the highest on the characteristic i^* . Formally, we find $l^* = \underbrace{ArgMax}_l \{x_{i^*,l}\}$ over the scenes $l = 1, \dots, L$. The resulting scene-characteristic pair should be modified to reduce the score x_{i^*,l^*} . We illustrate this idea via Google and Apple ads by identifying the scene-characteristic pairs that need editing and re-visualization.

For the Google ad, we first identify the dominant negative ad characteristic. In Table 4, we observe that $\underbrace{ArgMin}_{i \in \{1, \dots, 6\}} \{\hat{\beta}_1 = 0.362, \hat{\beta}_2 = 0.289, \hat{\beta}_3 = -0.678, \hat{\beta}_4 = -0.025, \hat{\beta}_5 = -0.547, \hat{\beta}_6 = 0.085\} = i^* = 3$ (Irritation). Then we identify the scene that scored the highest on the irritation characteristic. Specifically, we inspect $\underbrace{ArgMax}_{l \in \{1, \dots, 6\}} \{x_{31} = 2.5, x_{32} = 2.25, x_{33} = 4, x_{34} = 4, x_{35} = 5, x_{36} = 2.75\}$ to find the scene $l^* = 5$. Thus, to enhance liking, the proposed framework offers specific guidance to reduce irritation in scene 5. Indeed, this editing process can continue recursively to identify the next prominent irritating scenes 3 and 4 and then move to the next dominant negative characteristic (stimulation) and its prominent scenes, and so on. Similarly, for the Apple ad, the ad characteristic under scrutiny based on Table 4 is $i^* = 6$ (Warmth), and its corresponding scenes are $l^* = \{3, 4, 5, 9, 11\}$, which scored the highest (5 points). Thus, decreasing the stimulation in these scenes would increase liking for the Apple ad.

5.4. Meta-analysis of 100 Video Ads

5.4.1. Mean Effects

Applying the algorithm in Table 1, we estimated the ad-specific asynchronous effects of 100 video advertisements. Then, we conducted a meta-analysis of the estimated effects via the R package *metafor* (Viechtbauer, 2010). Table 5 presents the means of random effects. Figure 3 displays the histograms overlaid with nonparametric density. We observe that the mean liking persistence $\hat{\lambda} = 0.92$, and the mean effects of familiarity $\hat{\beta}_2 = 0.223$, irritation $\hat{\beta}_3 = -0.494$, and stimulation $\hat{\beta}_5 = 0.056$ have the correct signs and are statistically significant. The effects of entertainment, relevance, and warmth can be either positive or negative across ads, exhibiting substantial heterogeneity ($I^2 = 98.25$). Hence, for these latter ad characteristics, managers and researchers should use the ad specific results as illustrated in section 5.2. Overall, these findings, first, reassure us that the directional effects in the extant

literature based on a few ads per study comport with this meta-analysis of hundred ads.

Second, the mean liking persistence $\hat{\lambda} = 0.92$ amplifies the content effects by $\frac{\hat{\beta}_i}{1-0.92} = 12.5 \hat{\beta}_i$. This twelve-fold amplification of the impact of ad characteristics is substantial, and so future studies should not ignore the persistence of ad liking in their studies.

"

5.4.2. Meta-Regression Model

What explains the heterogeneity in Figure 3? Ad length? American versus Dutch ads?

Industry sectors? Plot structure? These variables, other than plot structure, are self-evident.

Plot structure captures the narrative elements of an ad (Thorndyke 1977). Textbooks recommend that ads should focus on a “single message instead of one that makes too many points. Focus on a single idea and support it” (Moriarty, Mitchell and Wells, 2017, p. 276).

We label such focused plot structures as “A-only,” where A denotes an ad characteristic. For example Norwegian Cruises implement E-only pattern, where E indicates entertainment; Ikea, Nescafe or Nintendo display F-only pattern, where F denotes familiarity; Bing Clutch implements W-only pattern, where W indicates warmth; Bing NYC and Bing Restaurant exemplify S-only pattern, where S stands for stimulation; Windows Phone and Cheerio build on relevance and display R-only pattern (R denotes relevance), whereas Windows and Activia show I-only pattern (I denotes irritation).

But some ads such as Bing Supermarket differ in their plot structure. They use the “repetition break” pattern (Loewenstein and Heath 2009; Loewenstein; Raghunathan and Heath 2011), where an ad characteristic (relevance for Bing Supermarket) repeats across the first two scenes and then a different ad characteristic (stimulation) breaks the narrative continuity in the remaining four scenes (Rozin, Rozin, Appel and Wachtel 2006). We label this repetition-break pattern as AB plot structure.

When A-only pattern shows one switch to a different ad characteristic in the middle (rather than at the end), we get ABA plot structure (Stern 1994; Deighton, Romer and McQueen 1989). For example, Volvo ad displays relevance-warmth-relevance pattern. When multiple switches between two dimensions exist, we label it as “Alternating” plot structure. For example, Apple’s ad presents this plot structure with the entertaining-familiar-entertaining-familiar pattern.

To measure plot structure, we use ratings data x_{il} and identify the ad characteristic i^* in each scene l via the operation $i_l^* = \underset{i \in \{1, \dots, 6\}}{\text{ArgMax}} \{x_{1l}, x_{2l}, x_{3l}, x_{4l}, x_{5l}, x_{6l}\}$. If multiple x_{il} are identical, we resolve the tie based on $\underset{i \in \{1, \dots, 6\}}{\text{ArgMax}} \{\hat{\beta}_i\}$. We then code the resulting plot structures using dummy variables, where the baseline is defined by plots which do not correspond to any of the patterns outlined above (i.e., A-only, AB, ABA or Alternating). Finally, we use the random effects meta-regression model (DerSimonian and Laird 1986) with six ad characteristics and liking persistence as the dependent variables, and ad length (in seconds), American or Dutch ads dummy, industry dummies, and the plot structure dummies as the regressors. Table 6 reports the regression results.

"

5.4.2.1. Liking Persistence

Based on the first equation in Table 6, liking persistence $\hat{\lambda}$ increases by 0.23% ($= 100 \times \frac{0.0020}{0.8695}$) as the ad length increases by one second. This effect seems small: liking persistence increases by 6.9% as the ad length doubles to 60 seconds. Yet this modest increase *disproportionately* amplifies the content effects. To see this point, consider a 30-second ad with $\hat{\lambda}_{30} = 0.9$. It amplifies the content effects by $\frac{1}{1 - \hat{\lambda}_{30}} = 10$. Whereas its 60-second version

yields $\hat{\lambda}_{60} = 0.9 \times 1.069 = 0.96$, and the amplification factor increases to $\frac{1}{1-\hat{\lambda}_{60}} = 25$. Thus, we learn that the dramatic impact of longer ads dwells in the persistence of the flow of liking.

The Dutch creatives amplify the ad content effects more than the US ads do, as evidenced by the effect of the indicator variable $I_{US} = 1$ (0 for Dutch ads). Furthermore, stimulation-only plot suppresses the liking persistence by 4.26% ($= 100 \times \frac{0.0370}{0.8695}$). Lastly, all the industry dummies are insignificant, and so the liking persistence generalizes across the industry sectors.

5.4.2.2. Entertainment Effect

The second equation in Table 6 expresses that warmth-only or stimulation-only plot structure decreases the entertainment impact on liking. *This result sheds new light that warm and stimulating ads may be hazardous in building liking.* Table 5 shows that the mean entertainment effect is -0.058 with the 95% confidence interval $(-0.131, 0.015)$ across ads. Consequently, warm or stimulating creatives work for some ads (e.g., Bing Clutch or Yoplait), but not others (e.g., Oral B). Since ad agencies don't know such outcome *ex-ante*, they should use the proposed approach to first test their creatives, as illustrated in section 5.2, to determine systematically whether *their specific* creative generates positive entertainment effect on ad liking or not. Lastly, all the industry dummies are insignificant, and so the entertainment effect also generalizes across the industry sectors.

5.4.2.3. Plot Structure Effects

Table 6 reveals that A-only plot structure significantly impacts the liking persistence and every ad characteristic. Warmth-only plot structure exerts both positive and negative effects: it reduces the entertainment effect and increases the irritation effect. Similarly, S-only plot increases the stimulation effect, but decreases liking persistence and entertainment effects. Table 5 shows that the mean stimulation effect is 0.056 with the 95% confidence interval $(-0.006, 0.119)$ across ads. Because a particular ad's $\hat{\beta}$ can be positive, ad agencies can

make stimulating creatives, but they need to be mindful of its adverse impact on the entertainment effect. Furthermore, as expected, A-only plots using familiarity, warmth, relevance, irritation or stimulation increases its own effectiveness.

Besides A-only plots, the single-switch (AB and ABA) and multiple switches (Alternating) plots exert significant impacts on the relevance and warmth coefficients. While alternating plot enhances the relevance coefficient, AB and ABA exert mixed effects: they increase relevance, but adversely affect warmth. *This empirical result offers evidence that a single, focused theme need not be the only way to build liking for ads*, a finding that differs from current beliefs (e.g., Moriarty, Mitchell, Wells 2017, p. 276). Specifically, the alternating plot works too, as Apple's ad exemplifies. Overall, this study is the first one to empirically document these moderating effects of plot structure on ad liking.

5.4.2.4. Industry Effects

The mean effects in Table 5 generalize across the industry sectors, except for the familiarity effect that's larger for leisure industry. In other words, none of the other effects in Table 6 depend on industry sectors. Managers and agencies believe that their brands, their companies, their industries differ from everyone else's. Consistent with this belief, each ad's estimated effects are unique. Yet these effects belong to the *same distribution* with a common mean invariant to industry sectors, reflecting a broad meta-analytic generalization.

5.4.2.5. Summary

This meta-analysis is based on the largest sample of ads hitherto. So, what have we learned?

1. Ad length does not, by itself, significantly impact ad characteristics. But it impacts the liking persistence, which increases by about 7% as the ad length doubles from 30 to 60 seconds. The resulting amplification of content effects may increase from a factor of 10 to 25.
2. The flow of liking typically amplifies the content effects twelve folds. Given this magnitude, future studies should not ignore the persistence of ad liking.
3. The US ads are less irritating and more stimulating, whereas the Dutch ads exhibit marginally greater liking persistence.

4. The impact of entertainment, relevance, or warmth on liking can be positive or negative for a given creative. Because managers don't know that *ex ante* for their specific creative, they should test their creatives individually by estimating the magnitude, direction and significance uncontaminated by the presence of other ads (see section 5.2).
5. Ads focused on a single theme need not be the only way to build liking, in contrast to textbook suggestions (e.g., Moriarty, Mitchell, Wells 2017, p. 276). Alternating plot structure with multiple switches also works (e.g., Apple ad).
6. Warm, stimulating or familiar ads may be hazardous in building liking.
7. Except that familiarity matters in the leisure sector, the industry effects are insignificant: so, these findings generalize.

6. Conclusions

This study makes substantive and methodological contributions. Substantively, we analyze 100 advertisements, the largest study of ad creatives in the literature, to estimate the effects of ad content on liking. The main estimation challenge is posed by the asynchronous availability of metrics: ad characteristics are available once every 5 seconds, whereas ad liking evolves every second. As Google or Apple ads illustrate, these content effects are estimated for a particular ad, uncontaminated by the presence of other ads in the estimation sample.

Consequently, specific scenes of a given ad and specific ad characteristics of the identified scene(s) furnish diagnostic information for editing and re-visualization. Analyzing a hundred ad creatives, we discover seven findings that generalize across industry sectors (see section 5.4.2.4). For example, the dramatic impact of 60-second ads dwells in the persistence of the flow of liking. Furthermore, the heterogeneity in content effects relates to the narrative elements of plot structures. The plot structure A-only —stick to one theme— is the most common plot, but it need not be the only way to build liking.

Methodologically, we advance the practice of advertising from collecting data, as in Figure 2, to analyzing it to: (1) understand the ad-specific impact of content effects on liking and (2) identify specific scenes and their ad characteristics for editing. Specifically, the

general method we developed tackles the asynchronous time series data, allowing for all possibilities of asynchronicity such as x -variables are slow-moving relative to y -variables, or y -variables are fast moving, or multivariate x -variables possess unequal frequencies, or multivariate y -variables exhibit mixed frequencies. The unavailability of certain elements of time series data introduces additional uncertainty, which we incorporate in the estimation and inference of dynamic models. Specifically, Appendix A presents the derivation of the optimal asynchronous gain factor in Proposition 1, and Table 1 offers the algorithm for the estimation and inference of dynamic models using asynchronous time series data.

The proposed theory (Proposition 1) and algorithm (Table 1) offer three types of generality: data availability, model structure, and error distributions. Data availability can follow any pattern from systematic unavailability (e.g., observed once every 5 seconds) to occasionally missing observations in multivariate x -variables or y -variables evolving at irregular frequencies. Model specifications can follow a broad class of state space models, as described in Appendix B, going beyond the special cases of regression, VAR, ARIMA, or factor models. Error distributions need not be multivariate Gaussian and can follow any distribution with finite means and covariance matrix.

We close by emphasizing that the empirical scope of the proposed method goes well beyond this novel study of 100 ad creatives because virtually every time series in business and economics is asynchronous: daily yield curve, monthly consumer confidence, quarterly gross domestic product, annual recession probability; or daily online search, weekly awareness and sales, monthly advertising spends, quarterly price increases, annual salesforce re-sizing, not to mention occasional missing data even for regularly observed time series. We hope researchers and managers use the proposed approach to analyze multiple asynchronous time series.

References

- Aaker, D. and D. Stayman (1990), "Measuring Audience Perceptions of Commercials and Relating Them to Ad Impact," *Journal of Advertising Research*, 30(4), 7-18.
- Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica*, 33 (1), 178-196.
- Ansley, C. F. and R. Kohn (1983), "Exact Likelihood of Vector Autoregressive-Moving Average Process with Missing or Aggregated Data," *Biometrika*, 70 (1), 275-280.
- Aruoba, S. B., F. X. Diebold, and C. Scotti (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27 (4), 417-427.
- Bai J., E. Ghysels, J. and H. Wright (2013), "State Space Models and MIDAS Regressions," *Econometric Reviews*, 32(7), 779-813.
- Banbura, M., and M. Modugno (2014), "Maximum Likelihood Estimation of Factor Models On Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Economics*, 29 (1), 133-160.
- Baumgartner, H., M. Sujaan, and D. Padgett (1997), "Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgments," *Journal of Marketing Research*, 34 (2), 219-232.
- Bernanke, B.S., J. Boivin, and P. Elias (2005), "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (favar) Approach," *Quarterly Journal of Economics*, 120 (1), 387-422.
- Biel, A. L., and C. Bridgewater (1990), "Attributes of Likable Television Commercials," *Journal of Advertising Research*, 30 (3), 38-44.
- Biyalogorsky, E. and P. A. Naik (2003), "Clicks and Mortar: The Effect of On-line Activities on Off-line Sales," *Marketing Letters*, 14 (1), 21-32.
- Brockwell, P. J., and R. A. Davis RA (1996), *Introduction to Time Series and Forecasting*, Springer, New York.
- Bruce, N. I., K. Peters, and P. A. Naik (2012), "Discovering How Advertising Grows Sales and Builds Brands," *Journal of Marketing Research*, 49(6), 793-806.
- Chiu, C. W., B. Eraker, A. T. Foerster, T. B. Kim, and H. D. Seoane (2011), "Estimating VAR's Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach," *Federal Reserve Bank of Kansas City*, working paper, ISSN 1936-5330.
- Deighton, John, Daniel Romer, and Josh McQueen (1989), "Using Drama to Persuade," *Journal of Consumer Research*, 16 (3), 335-43.
- DerSimonian, R., and N. Laird (1986), "Meta-analysis in Clinical Trials," *Controlled Clinical Trials* 7 (1), 177-188.

- Doz, C., D. Giannone, and L. Reichlin (2011), “A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering,” *Journal of Econometrics*, 164, 188–205.
- Durbin, J. and S. J. Koopman (2001), *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series: London, UK.
- Elpers, J., A. Mukherjee, and W. D. Hoyer (2004). “Humor in Television Advertising: A Moment-to-Moment Analysis,” *Journal of Consumer Research*, 31 (3), 592-98.
- Elpers, J., M. Wedel, and R. G.M. Pieters (2003), “Why Do Consumers Stop Viewing Television Commercials? Two Experiments on the Influence of Moment-to-Moment Entertainment and Information Value,” *Journal of Marketing Research*, 40 (4), 437-53.
- Ernst, P., L. Shepp, and A. Wyner (2016), “Yule’s “Nonsense Correlation” Solved!” *arXiv*, 1608.04120v2.
- Foroni, C. and M. Marcellino (2013), “A Survey of Econometric Methods for Mixed-Frequency Data,” *Norges Bank, working paper*, ISSN 1502-8143.
- Ghysels, E., A. Sinko, and R. Valkanov (2007), “MIDAS Regressions: Further Results and New Directions,” *Econometric Reviews*, 26 (1), 53-90.
- Giannone, D., L. Reichlin, and D. Small (2008), “Nowcasting: The real-time informational content of macroeconomic data,” *Journal of Monetary Economics*, 55, 665-676
- Harvey, A. C. and R. G. Pierse (1984), “Estimating Missing Observations in Economic Time Series,” *Journal of the American Statistical Association*, 79 (385), 125-131.
- Harvey, A. C. (1994), *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge University Press, New York, NY.
- Holbrook, M. and Batra, R. (1987), “Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising,” *Journal of Consumer Research*, 14(3), 404-420.
- Hui, S., Meyvis, T. and Assael, H. (2014), “Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing,” *Marketing Science*, 33(2), 222-240.
- Jones, R. H. (1980), “Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations,” *Technometrics*, 22 (3), 389-395.
- Koop G. and D. Korobilis (2010), “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics,” *Foundations Trends Econometrics* 3 (4) 267-358.
- Liu, X., Shi, S. W., Teixeira, T. and Wedel, M. (2018), “Video Content Marketing: The Making of Clips,” *Journal of Marketing*, 82(4), 86-101.
- Loewenstein, J. and Heath, C. (2009), “The Repetition-Break Plot Structure: A Cognitive Influence on Selection in the Marketplace of Ideas,” *Cognitive Science*, 33(1), 1-19.

- Loewenstein, J., Raghunathan, R. and Heath, C. (2011), "The Repetition-Break Plot Structure Makes Effective Television Advertisements," *Journal of Marketing*, 75(5), 105-19.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer: Berlin.
- Madrigal, R. and C. Bee (2005), "Suspense As an Experience of Mixed Emotions: Feelings of Hope and Fear While Watching Suspenseful Commercials," in *Advances in Consumer Research*, 32, 561-567.
- Mariano, R. S. and Y. Murasawa (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Economics*, 18, 427-443.
- Moldovan, S. E. (1984), "Copy Factors Related to Persuasion Scores," *Journal of Advertising Research* 24(6), 16-22.
- Mitnik, S. and P. Zadrozny (2004), "Forecasting Quarterly German GDP at Monthly Intervals Using Monthly IFO Business Conditions Data," *CESIFO, Working Paper*, Nos. 1203.
- Moriarty, S., N. D. Mitchell and W. D. Wells (2017), *Advertising and IMC: Principles and Practice*, 10th Edition, Pearson Education: Noida, India.
- Naik, P. A. and Tsai, C.-L. (2000), "Controlling Measurement Errors in Models of Advertising Competition," *Journal of Marketing Research*, 37 (1), 113-124.
- Raudenbush, S. W. and A. S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd Edition, Sage: Thousand Oaks, CA.
- Rozin, P., A. Rozin, B. Appel and C. Wachtel (2006), "Documenting and Explaining the Common AAB Pattern in Music and Humor: Establishing and Breaking Expectations," *Emotion*, 6 (3), 349-355.
- Schlinger, M. J. (1979), "A Profile of Responses to Commercials," *Journal of Advertising Research*, 19(2), 37-46.
- Schweppe, F. C. (1965), "Evaluation of Likelihood Function for Gaussian Signals," *IEEE Transactions of Information Theory*, 11 (1), 61-70.
- Shumway, R. H. and D. S. Stoffer (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3(4), 253-264.
- Smit, E. G., L. V. Meurs, and P. C. Neijens (2006), "Effects of Advertising Likeability: A 10-Year Perspective," *Journal of Advertising Research*, 46(1), 73-83.
- Stern, B. (1994), "Classical and Vignette Television Advertising Dramas: Structural Models, Formal Analysis, and Consumer Effects," *Journal of Consumer Research*, 20(4), 601-15.

- Teixeira, T., R. Picard, and R. Kaliouby (2014), "Why, When, and How Much to Entertain Consumers in Advertisements? A Web-based Facial Tracking Field Study," *Marketing Science*, 33(6), 809-827.
- Teixeira, T., M. Wedel and R. Pieters (2012), "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 49(2), 144-59.
- Thorndyke, P. W. (1977), "Cognitive Structures in Comprehension and Memory of Narrative Discourse," *Cognitive Psychology*, 9 (1), 77-110.
- Viechtbauer, W. (2010), "Conducting Meta Analyses in R with the Metafor Package," *Journal of Statistical Software*, 36 (3), 1-48. <http://www.jstatsoft.org/v36/i03/>
- Wells, W., C. Leavitt and M. McConville (1971), "A Reaction Profile for TV Commercials," *Journal of Advertising Research*, 11(6), 11-18.
- Yule, G. U. (1926), "Why Do We Sometimes Get Nonsense-Correlations Between Time-Series? A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, 89 (1), 1-63.
- Zadrozny, P. (1990), "Estimating a Multivariate ARMA Model with Mixed Frequency Data: An Application to Forecasting US GNP at Monthly Intervals," *Center for Economic Studies*, U.S. Bureau of the Census, Washington, DC, *working paper*, 90-5.
- Zinkham, G. M., and C. Fornell (1985), "A Test of Two Consumer Response Scales in Advertising," *Journal of Marketing Research*, 22(4), 447-452.

Table 1
Recursions, Estimation, and Inference for Asynchronous Dynamic Models

Step 1	<ul style="list-style-type: none"> • Setup the parameter vector θ of appropriate length • Using θ, create time-invariant system matrices $\{Z, T, c, d, H, Q\}$
Step 2	<ul style="list-style-type: none"> • Initialize the state mean a_0 using one of the elements of θ • Specify $P_0 = \kappa I$, where κ equals about 3 times the values of a_0, and I is an identity matrix
Step 3	<p>Asynchronous Filter Recursions</p> <ul style="list-style-type: none"> • For $t = 1, \dots, N$ <ul style="list-style-type: none"> ◦ Create the remaining subset of time-varying system matrices $\{Z_t, T_t, c_t, d_t, H_t, Q_t\}$ using θ. <p>Time Update</p> <ul style="list-style-type: none"> ◦ Compute $a_{t t-1}$ using (A4) in Appendix A ◦ Compute $P_{t t-1}$ using (A5) in Appendix A ◦ Compute the optimal gain factor using (7) in Proposition 1 <p>Likelihood contribution</p> <ul style="list-style-type: none"> ◦ Compute \hat{D}_t using (A2) in Appendix A ◦ Compute the forecast error $e_t = D_t - \hat{D}_t$. ◦ Compute F_t using (A3) in Appendix A ◦ Compute $l_t = -0.5[\ln(\det(F_t)) + e_t' F_t^{-1} e_t]$ <p>Measurement Update</p> <ul style="list-style-type: none"> ◦ Update a_t using (A6) in Appendix A ◦ Update P_t using (A8) in Appendix A <ul style="list-style-type: none"> • Do next t • Return $LL = \sum_{t=1}^N l_t$.
Step 4	<p>Estimation</p> <ul style="list-style-type: none"> • Specify the starting values θ_0 <ul style="list-style-type: none"> ◦ Use derivative-free methods such as the simulated annealing or genetic algorithm to maximize LL and use its solution as θ_0 ◦ Use EM algorithm to maximize $E[LL]$ (see Shumway and Stoffer 1982) and use its solution as θ_0. • Specify tolerance of 10^{-5} for convergence • Use BFGS to numerically maximize LL with respect to θ <ul style="list-style-type: none"> ◦ BFGS returns the optimal solution: the maximized LL^*, the parameter values θ^*, and the Hessian \mathcal{H} at convergence • Check convergence <ul style="list-style-type: none"> ◦ Is the largest slope $\frac{\partial LL}{\partial \theta_i}$ across all the elements of θ smaller than the tolerance? If so, convergence is attained. ◦ If convergence failed, rescale (Y_t, X_t) so that the various elements of θ take similar magnitudes during BFGS iterations • Check that different starting values do not yield larger LL^* <p>Inference</p> <ul style="list-style-type: none"> • Compute $se(\theta^*) = \text{sqrt}(\text{diag}(-\mathcal{H}^{-1}))$ • Use $se(\theta^*)$ to obtain the confidence intervals, t-values, and p-values.

Table 2. Accuracy and Efficiency

		<i>Panel A</i> <i>Every 5th Observation</i>			<i>Panel B</i> <i>Every 3rd Observation</i>		
Parameters		λ	β_1	β_2	λ	β_1	β_2
True values		0.5	1.0	-0.5	0.5	1.0	-0.5
Repetitions	1	0.527	1.019	-0.531	0.544	0.978	-0.471
	2	0.505	1.085	-0.552	0.487	1.095	-0.539
	3	0.416	1.142	-0.603	0.350	1.322	-0.633
	4	0.564	1.005	-0.468	0.472	1.134	-0.540
	5	0.515	1.102	-0.529	0.535	0.988	-0.311
	\vdots		\vdots			\vdots	
	\vdots		\vdots			\vdots	
	96	0.566	0.956	-0.423	0.451	1.091	-0.327
	97	0.528	1.057	-0.584	0.462	1.119	-0.428
	98	0.667	0.831	-0.396	0.481	1.085	-0.436
	99	0.410	1.295	-0.608	0.581	0.913	-0.351
	100	0.600	0.943	-0.482	0.504	1.031	-0.365
Average Estimates		0.554	0.998	-0.510	0.465	1.119	-0.462
Bias (Accuracy)		0.054	-0.002	-0.010	-0.035	0.119	0.038
Variance (Efficiency)		0.037	0.137	0.024	0.022	0.082	0.065

Table 3. Descriptive Statistics

Variables	N	Mean	Standard Deviation	Minimum	Maximum
<i>Ad Liking</i>	20412	0.860	2.317	-4.942	4.966
<i>Ad Characteristics</i>					
Entertainment	100	3.632	0.701	1.964	5.167
Familiarity	100	3.755	0.748	2.417	5.286
Irritation	100	1.863	0.916	1.000	4.428
Relevance	100	4.013	0.724	2.250	5.800
Stimulation	100	3.718	0.642	2.500	5.357
Warmth	100	3.385	0.896	1.833	5.500

Table 4. Scene-by-scene Effects on Moment-to-Moment Ad Liking

	Google Ad		Apple Ad	
	<i>Estimates</i>	<i>Std. Errors</i>	<i>Estimates</i>	<i>Std. Errors</i>
Ad Liking, λ	0.952	0.006	0.966	0.002
Entertainment, β_1	0.226	0.088	0.362	0.013
Familiarity, β_2	0.323	0.107	0.289	0.013
Irritation, β_3	-0.026	0.018	-0.678	0.023
Relevance, β_4	0.064	0.115	-0.025	0.010
Stimulation, β_5	-0.142	0.081	-0.547	0.027
Warmth, β_6	-0.398	0.047	0.085	0.031
Liking Noise, σ_ϵ	0.514	0.007	0.366	0.003
Ad Content Noise, σ_v	0.509	0.006	0.273	0.001

Bold estimates are statistically significant at the 95% confidence interval

Table 5. Meta Analysis Results

<i>Effects</i>	<i>Means</i>	<i>Std. Errors</i>
Ad Liking, $\hat{\lambda}$.920	.002
Entertainment, $\hat{\beta}_1$	-.058	.037
Familiarity, $\hat{\beta}_2$.223	.032
Irritation, $\hat{\beta}_3$	-.494	.054
Relevance, $\hat{\beta}_4$	-.014	.021
Stimulation, $\hat{\beta}_5$.056	.032
Warmth, $\hat{\beta}_6$	-.001	.036

Bold estimates are statistically significant at the 95% confidence level.

Table 6. Meta-Regression Results

<i>Coefficients</i>	<i>Estimated Equations</i>
Persistence	$\hat{\lambda} = \underbrace{0.8695}_{0.0180} + \underbrace{0.0020}_{0.0004} Ad\ Length - \underbrace{0.0175}_{0.0088} I_{US} - \underbrace{0.0370}_{0.0154} S_{only}$
Entertainment	$\hat{\beta}_1 = -\underbrace{0.4913}_{0.1817} W_{only} - \underbrace{0.5151}_{0.1954} S_{only}$
Familiarity	$\hat{\beta}_2 = \underbrace{0.3524}_{0.1665} Leisure + \underbrace{0.6119}_{0.1464} F_{only}$
Irritation	$\hat{\beta}_3 = -\underbrace{0.5001}_{0.1493} I_{US} + \underbrace{1.8179}_{0.4059} I_{only} + \underbrace{0.6511}_{0.3095} W_{only}$
Relevance	$\hat{\beta}_4 = \underbrace{0.1854}_{0.1062} AB + \underbrace{0.3033}_{0.1159} ABA + \underbrace{0.1983}_{0.1135} Alternating + \underbrace{0.6281}_{0.1234} R_{only}$
Stimulation	$\hat{\beta}_5 = \underbrace{0.2016}_{0.1062} I_{US} + \underbrace{0.6712}_{0.1909} S_{only}$
Warmth	$\hat{\beta}_6 = -\underbrace{0.2428}_{0.1443} AB - \underbrace{0.3026}_{0.1575} ABA - \underbrace{0.3469}_{0.1491} F_{only} + \underbrace{0.7050}_{0.1917} W_{only}$

Standard errors are shown below the estimates.

Figure 1. Distribution of Ads

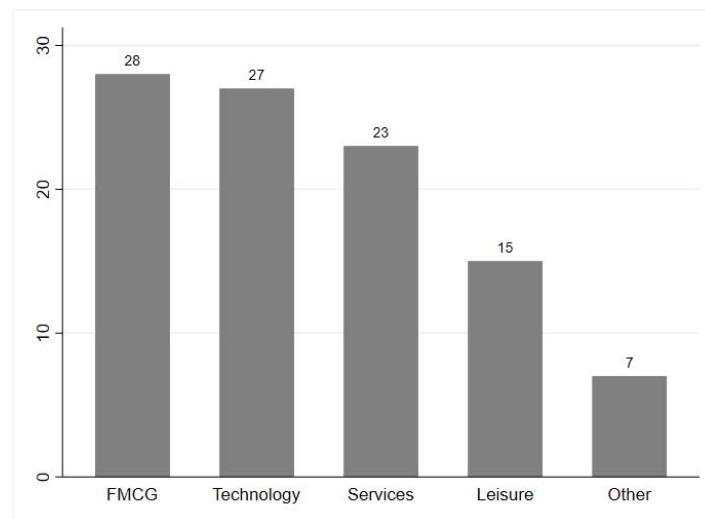
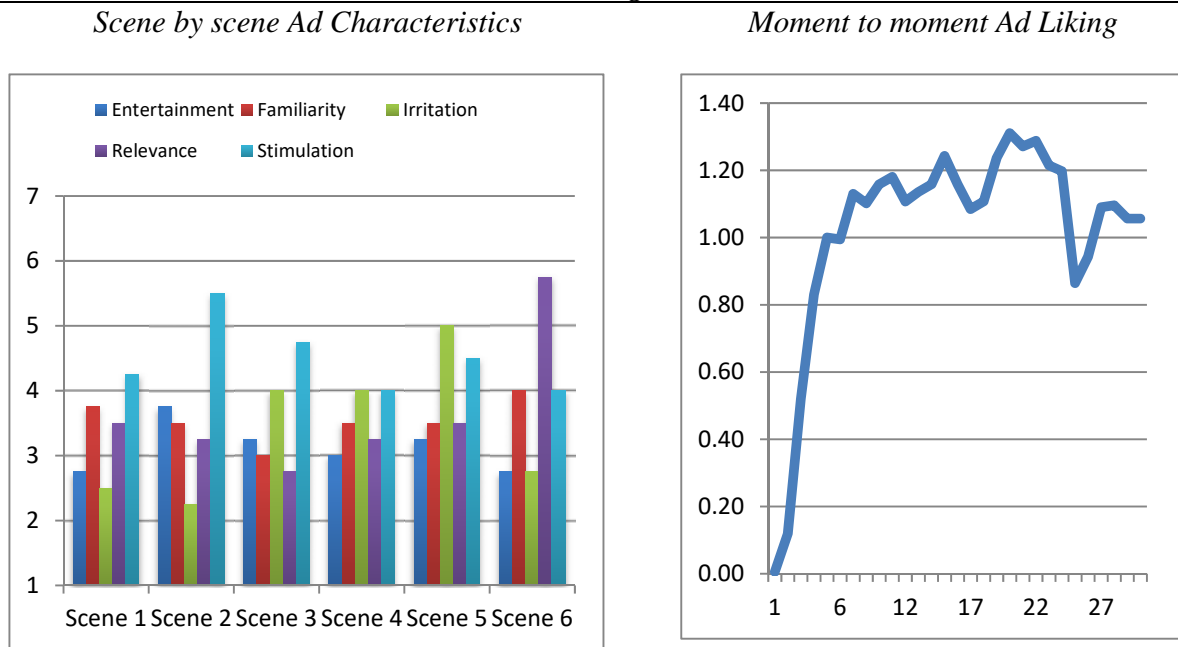


Figure 2. Ad Characteristics and Ad Liking

Panel A. Google Ad



Panel B. Apple Ad

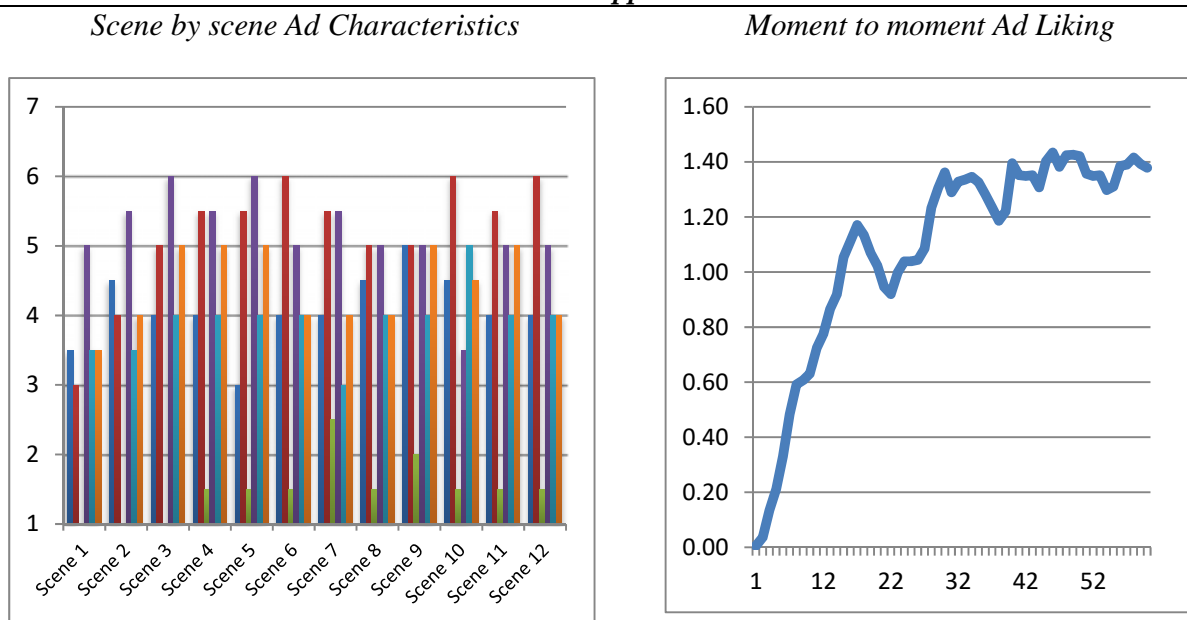
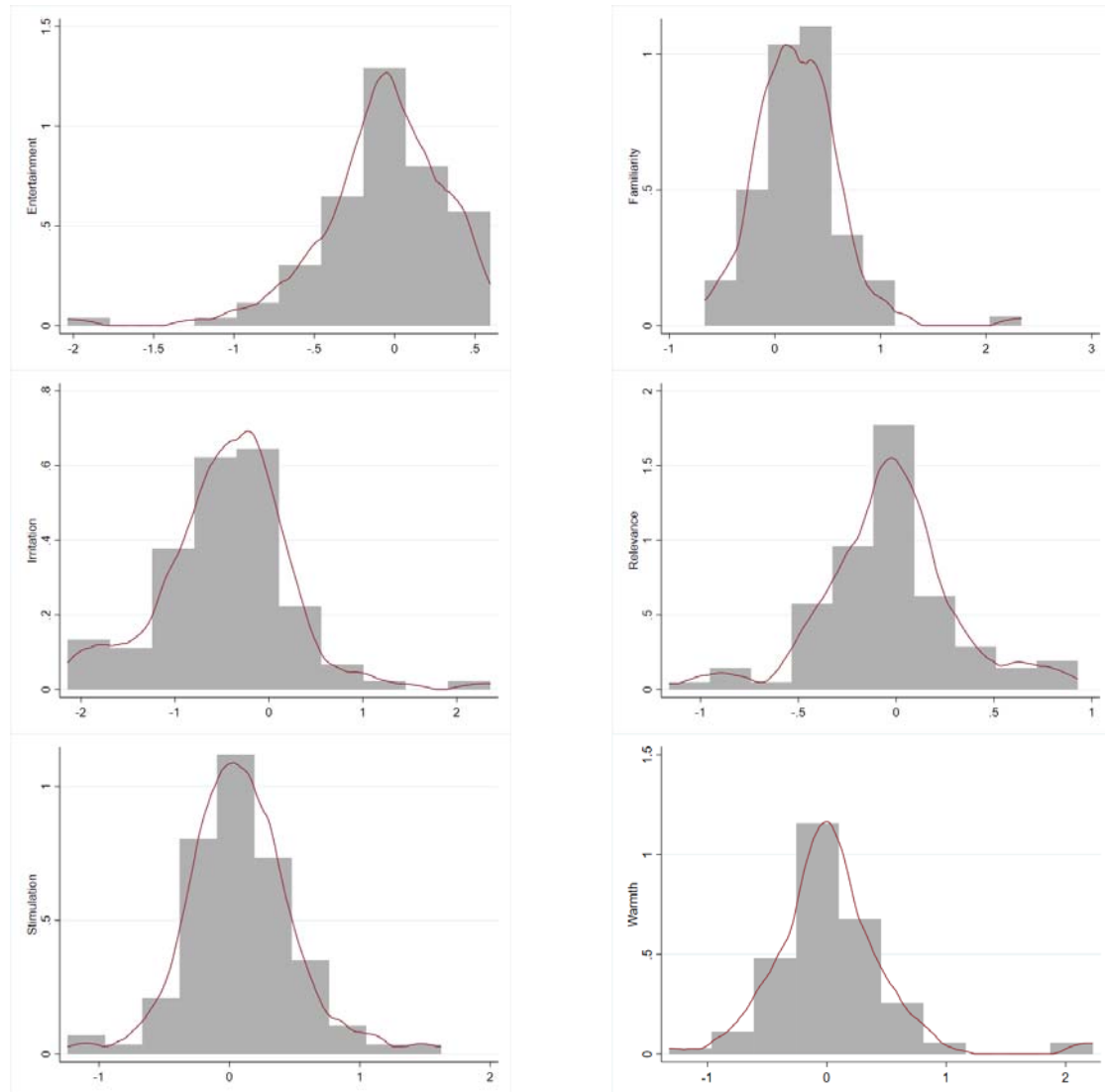


Figure 3. Heterogeneity of Ad Characteristics Effects



Appendix A. Derivation of the Optimal Asynchronous Filter

Based on equation (6), the new observation equation is

$$\begin{aligned} D_t &= A_t Y_t \\ &= A_t (Z \alpha_t + c_t + \epsilon_t) \\ &= A_t Z \alpha_t + A_t c_t + A_t \epsilon_t \end{aligned} \quad (A1)$$

Let $\mathfrak{I}_{t-1} = \{D_1, D_2, \dots, D_{t-1}\}$ denote the information available at time $(t-1)$, and $\mathfrak{I}_t = \mathfrak{I}_{t-1} \cup D_t$ denote the information available after observing data at instant t . Based on \mathfrak{I}_{t-1} and equation (4), the mean and covariance of D_t are given by

$$\hat{D}_t = E[D_t | \mathfrak{I}_{t-1}] = A_t Z_t a_{t|t-1} + A_t c_t \quad (A2)$$

$$F_t = \text{Var}[D_t | \mathfrak{I}_{t-1}] = A_t Z_t P_{t|t-1} Z_t' A_t' + A_t H_t A_t'. \quad (A3)$$

Based on \mathfrak{I}_{t-1} and equation (5), the mean and covariance of α_t are given by

$$a_{t|t-1} = E[\alpha_t | \mathfrak{I}_{t-1}] = T_t a_{t-1} + d_t \quad (A4)$$

$$P_{t|t-1} = \text{Var}[\alpha_t | \mathfrak{I}_{t-1}] = T_t P_{t-1} T_t' + Q_t. \quad (A5)$$

After observing D_t , based on \mathfrak{I}_t , the prediction errors $(D_t - \hat{D}_t)$ are used to update the mean of the state vector α_t as follows:

$$a_t = a_{t|t-1} + G_t (D_t - \hat{D}_t), \quad (A6)$$

where $a_t = E[\alpha_t | \mathfrak{I}_t]$, and G_t is an *arbitrary* gain matrix whose “optimal” value is to be determined.

To determine the optimal gain matrix G_t^* , we seek the mean state vector a_t to be the “closest” to the state vector α_t . To this end, we evaluate the expected sum of squared errors as

$$\begin{aligned} J_t &= E[(\alpha_{1t} - a_{1t})^2 + \dots + (\alpha_{nt} - a_{nt})^2] \\ &= E[e_t' e_t] = E[\text{Tr}(e_t' e_t)] = E[\text{Tr}(e_t e_t')] \\ &= E[\text{Tr}(P_t)] \end{aligned}$$

where $e_t = (e_{1t}, \dots, e_{nt})'$, $e_{jt} = \alpha_{jt} - a_{jt}$, $j = 1, \dots, n$. The third equality follows from the fact that $e_t' e_t$ is a scalar, the fourth from the cyclic permutation $\text{Tr}(AB) = \text{Tr}(BA)$, and the final from the definition of the covariance of the state vector.

To analytically evaluate P_t , we first note that

$$\begin{aligned} d_t &= \alpha_t - a_t \\ &= \alpha_t - [a_{t|t-1} + G_t (D_t - \hat{D}_t)] \\ &= (\alpha_t - a_{t|t-1}) - G_t (D_t - A_t Z_t a_{t|t-1} - A_t c_t) \\ &= d_{t|t-1} - G_t ((A_t Z_t \alpha_t + A_t c_t + A_t \epsilon_t) - A_t Z_t a_{t|t-1} - A_t c_t) \\ &= d_{t|t-1} - G_t A_t Z_t \alpha_t - G_t A_t \epsilon_t + G_t A_t Z_t a_{t|t-1} \end{aligned}$$

$$\begin{aligned}
&= d_{t|t-1} - G_t A_t Z_t (\alpha_t - a_{t|t-1}) - G_t A_t \epsilon_t \\
&= (I - G_t A_t Z_t) d_{t|t-1} - G_t A_t \epsilon_t \\
&= L d_{t|t-1} - G_t A_t \epsilon_t
\end{aligned} \tag{A7}$$

where we denote $L = I - G_t A_t Z_t$ for notational brevity in what follows. Next, by definition, we obtain

$$\begin{aligned}
P_t &= E[d_t d_t'] \\
&= E[(L d_{t|t-1} - G_t A_t \epsilon_t)(L d_{t|t-1} - G_t A_t \epsilon_t)'] \\
&= E[L d_{t|t-1} d_{t|t-1}' L' + L d_{t|t-1} (-\epsilon_t' A_t' G_t') + (-G_t A_t \epsilon_t) d_{t|t-1}' L' + G_t A_t \epsilon_t \epsilon_t' A_t' G_t'] \\
&= L E[d_{t|t-1} d_{t|t-1}'] L' + 0 + 0 + G_t A_t E[\epsilon_t \epsilon_t'] A_t' G_t' \\
&= L P_{t|t-1} L' + G_t A_t H_t A_t' G_t'.
\end{aligned} \tag{A8}$$

Then, noting that $\partial \text{Tr}(ABA') = 2AB\partial A$ for a symmetric B , and that $L = I - G_t A_t Z_t$ depends on G_t , we differentiate $\text{Tr}(P_t)$ to obtain the first order condition (FOC):

$$\frac{\partial \text{Tr}(P_t)}{\partial G_t} = 2L P_{t|t-1} (-A_t Z_t)' + 2G_t A_t H_t A_t'. \tag{A9}$$

By setting (A9) to zero, we derive the optimal gain matrix G_t^* :

$$\begin{aligned}
2(I - G_t^* A_t Z_t) P_{t|t-1} (-A_t Z_t)' + 2G_t^* A_t H_t A_t' &\equiv 0 \implies \\
(I - G_t^* A_t Z_t) P_{t|t-1} Z_t' A_t' &= G_t^* A_t H_t A_t' \\
P_{t|t-1} Z_t' A_t' &= G_t^* A_t Z_t P_{t|t-1} Z_t' A_t' + G_t^* A_t H_t A_t' \\
P_{t|t-1} Z_t' A_t' &= G_t^* (A_t Z_t P_{t|t-1} Z_t' A_t' + A_t H_t A_t') \\
\therefore G_t^* &= P_{t|t-1} Z_t' A_t' (A_t Z_t P_{t|t-1} Z_t' A_t' + A_t H_t A_t')^{-1}
\end{aligned} \tag{A10}$$

This completes the proof of Proposition 1.

Appendix B. Generality of Model Equations (4) and (5)

We show how broad classes of statistical models are special cases of equations (4) and (5), which can be estimated using one algorithm given in Table 1 rather than create model-specific algorithms for various models (which would be the case otherwise).

1. VAR. Consider the general vector autoregression model:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1} & \phi_{m2} & \dots & \phi_{mm} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ \vdots \\ Y_{mt-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{bmatrix}, \quad \epsilon_t \sim N(0, \Sigma). \quad (\text{B1})$$

Equation (B1) can be expressed in the state space form (4) and (5) by setting $Z_t = I, T_t = \Phi = \{\phi_{ij}\}, d_t = 0, H_t = 0, Q_t = \Sigma$. Note that $Z_t = I$ and $H_t = 0$ ensures $\alpha_t = Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{mt})'$.

2. VAR-X. Consider the vector autoregression model with exogenous variables:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1} & \phi_{m2} & \dots & \phi_{mm} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ \vdots \\ Y_{mt-1} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \dots & \beta_{mk} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{bmatrix}, \quad (\text{B2})$$

where $\epsilon_t \sim N(0, \Sigma)$. It can be expressed in the state space form by setting $Z_t = I, T_t = \Phi, H_t = 0, Q_t = \Sigma$, as before, and $c_t = BX_t$, where $B = \{\beta_{ij}\}$.

3. TVP-VARX. Consider VAR-X model with time-varying β_{ij} as follows:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1} & \phi_{m2} & \dots & \phi_{mm} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ \vdots \\ Y_{mt-1} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \dots & \beta_{mk} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{bmatrix}, \quad (\text{B3})$$

where $\epsilon_t \sim N(0, \Sigma)$, and

$$\theta_t = \theta_{t-1} + \tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \sim N(0, \tilde{\Sigma}), \quad (\text{B4})$$

where $\theta_t = \text{vec}(B_t)$, which stacks the rows of B_t to obtain the column vector θ_t . The random walk in (B4) permits a flexible, non-monotonic, and non-stationary pattern. The state space form subsumes TVP-VARX by setting $Z_t = [I \ 0], T_t = \begin{bmatrix} \Phi & 0 \\ 0 & I \end{bmatrix}, c_t = \begin{bmatrix} B_t X_t \\ 0 \end{bmatrix}, H_t = 0, Q_t = \begin{bmatrix} \Sigma & 0 \\ 0 & \tilde{\Sigma} \end{bmatrix}$,

and $\alpha_t = (Y_t', \theta_t')'$, which augments the state vector to include the time-varying parameters.

4. ARIMA Errors. Consider, for example, the regression model with ARMA(2,1) error term:

$$y_t = X_t' \beta + \epsilon_t, \quad (B5)$$

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + v_t + \theta_1 v_{t-1}, \quad v_t \sim N(0, \sigma_v^2). \quad (B6)$$

Then we cast (B6) as the state vector $\tilde{\alpha}_t = \begin{bmatrix} \epsilon_t \\ \phi_2 \epsilon_{t-1} + \theta_1 v_t \end{bmatrix}$ so that the transition equation in (B6) becomes

$$\begin{bmatrix} \epsilon_{t+1} \\ \phi_2 \epsilon_t + \theta_1 v_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \phi_2 \epsilon_{t-1} + \theta_1 v_t \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} v_{t+1},$$

where the first row reproduces (B6) with t replaced by $(t+1)$, and the second row is an identity. To complete the state space form, we express the state vector $\alpha_t = (\beta_t, \tilde{\alpha}_t)'$, the link matrix $Z_t = [X_t' \quad 1 \quad 0]$, $H_t = 0$, $\omega_t = \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} v_{t+1}$, $Q_t = \sigma_v^2 \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} \begin{bmatrix} 1 & \theta_1 \end{bmatrix}$. Thus, regression models with ARMA(2,1) errors, as illustrated here, and *any* ARIMA(p, d, q) can be cast into equations (4) and (5). For details, see Durbin and Koopman (2001, sections 3.3 and 3.5).

5. Dynamic Factor Models. Consider the dynamic factor model by Bruce, Peters and Naik (2012), which not only incorporates dynamics in the hierarchy of effects model of advertising, but also allows advertising to simultaneously trigger all the three stages of cognition (think), affect (feel), and experience (do) to influence brand sales:

$$\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} \gamma_{11} & & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & & \gamma_{24} \\ & & \gamma_{33} & \gamma_{34} \\ & \gamma_{42} & & \gamma_{44} \end{bmatrix} \begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} \beta_1 g(u_t) \\ \beta_2 g(u_t) \\ \beta_3 g(u_t) \\ \beta_4 g(u_t) \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{bmatrix} \quad (B7)$$

Then they extract the three factors by using observed data from the battery of n mindset metrics in each week t , denoted by x_{it} , along with the observed sales volume y_t with the mean sales level S_t as follows:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ \vdots \\ x_{nt} \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \lambda_{n3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{1t} \\ \vdots \\ \vdots \\ \epsilon_{nt} \\ \epsilon_{n+1,t} \end{bmatrix} \quad (B8)$$

The rows with zeros and unity impose identification restrictions. We express (B7) and (B8) in the state space form as follows: $Y_t = (x_{1t}, x_{2t}, \dots, x_{nt}, y_t)'$, $\alpha_t = (C_t, A_t, E_t, S_t)'$, $d_t = (\beta_1 g(u_t), \dots, \beta_4 g(u_t))'$, and so on.

6. Dynamic Regression. Consider a regression model with parameters varying more “smoothly” than that in the random walk in (B4). To instill smoothness, we specify small variations in the second-order differences $\nabla^2 \beta_t = \nu_t$. Then we get $\nabla^2 \beta_t = \nabla(\nabla \beta_t) = \nabla(\beta_t - \beta_{t-1}) = \nabla \beta_t - \nabla \beta_{t-1} = (\beta_t - \beta_{t-1}) - (\beta_{t-1} - \beta_{t-2}) = \beta_t - 2\beta_{t-1} + \beta_{t-2}$, which implies $\beta_t - 2\beta_{t-1} + \beta_{t-2} = \nu_t$. We cast the resulting smooth time-varying parameters model in the state space form as follows:

$$y_t = X_t' \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \text{ and} \quad (\text{B9})$$

$$\begin{bmatrix} \beta_t \\ \beta_{t-1} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{t-1} \\ \beta_{t-2} \end{bmatrix} + \begin{bmatrix} \nu_t \\ 0 \end{bmatrix}, \quad \nu_t \sim N(0, \sigma_\nu^2). \quad (\text{B10})$$

The corresponding system matrices are given by $Z_t = [X_t' \ 0]$, $c_t = 0$, $\alpha_{t+1} = (\beta_t, \beta_{t-1})'$, $H_t = \sigma_\epsilon^2$, $T_t = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$, $d_t = 0$, $Q_t = \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & 0 \end{bmatrix}$. More importantly, note that the second row in (B10) appends an identity ($\beta_{t-1} = \beta_{t-1}$), thus representing the scalar *second-order* lag model (namely, $\beta_t - 2\beta_{t-1} + \beta_{t-2} = \nu_t$) as the vector *first-order* lag model in (B10). Analogously, the vector first-order transition equation in equation (5) subsumes *any* p^{th} order linear dynamics.

7. Hierarchical Linear Models. Suppose we observe sales, S_{it} , over time $t = 1, \dots, T_i$ and individuals (or cities) $i = 1, \dots, N$ as well as corresponding exogenous data on individual (or city) characteristics. Then, a hierarchical linear model explains within-individual (or -city) sales variation in the first stage and between-individual (or -city) variation in the second stage as follows:

$$S_{it} = X_{it} \beta_i + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_\epsilon^2) \quad \text{Stage 1} \quad (\text{B11})$$

$$\beta_i = W_i \gamma + \nu_i, \quad \nu_i \sim N(0, \sigma_\nu^2) \quad \text{Stage 2} \quad (\text{B12})$$

Neither (B11) nor (B12) exhibit “dynamics” in the usual sense of inter-temporal dependence, yet dependence exists between the two stages due to the common β_i .

To express (B11) and (B12) in the state space form, consider the individual-level state space form with the observation equation $y_{it} = Z_{it} \alpha_{it} + v_{it}$, $v_{it} \sim N(0, H_{it})$, and the transition equation $\alpha_{it} = T_{it} \alpha_{i,t-1} + d_{it} + \omega_{it}$, $\omega_{it} \sim N(0, Q_{it})$. We then establish equivalence by setting $Z_{it} = X_{it}$, $T_{it} = 0$ (no dynamics), $d_{it} = W_i \gamma$, $\text{diag}(Q_{it}) = \sigma_\nu^2$, so that $\alpha_{it} = \beta_i$.

In sum, the equations (4) and (5) subsume broad classes of disparate statistical models, offering not only a wide generality of model structures, but also a unified theory (Proposition 1) and a common algorithm (Table 1) for estimation and inference of asynchronous dynamic models.