Intern. J. of Research in Marketing 32 (2015) 94–112

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Improving the statistical performance of tracking studies based on repeated cross-sections with primary dynamic factor analysis

Rex Yuxing Du [a,*], Wagner A. Kamakura [b,1]

[a] *Bauer College of Business, University of Houston, Houston, TX 77204-6021, United States*
[b] *Jones Graduate School of Business, Rice University, 6100 Main Street (MS 531), Houston, TX 77005, United States*

ARTICLE INFO

Article history:
First received on January 31, 2014 and was under review for 6 months
Available online 15 November 2014

Area Editor: Sandy D. Jap

Keywords:
Tracking study
Repeated cross-sectional survey
State-space model
Dynamic factor analysis

ABSTRACT

Tracking studies are prevalent in marketing research and virtually all the other social sciences. These studies are predominantly implemented via repeated independent, non-overlapping samples, which are much less costly than recruiting and maintaining a longitudinal panel that track the same sample over time. In the existing literature, data from repeated cross-sectional samples are analyzed either independently for each time period, or longitudinally by focusing on the dynamics of the aggregate measures (e.g., sample averages). In this study, we propose a multivariate state-space model that can be applied directly to the individual-level data from each of the independent samples, simultaneously taking advantage of three patterns embedded in the data: a) inter-temporal dependence within the population means of each variable, b) temporal co-movements across the population means of different variables and c) cross-sectional co-variation across individual responses within each sample. We illustrate our proposed model with two applications, demonstrating the benefits of making full use of all the available data. In the first illustration, we have access to all the individual-level purchase data from one large population of grocery shoppers over a span of 36 months. This provides us a testing ground for benchmarking our proposed model against existing approaches in a Monte Carlo experiment, where we show that our model outperforms all the alternatives in inferring population dynamics using data sampled through repeated cross-sections. We find that, as compared with using simple sample averages, our proposed model can improve the accuracy of repeated cross-sectional tracking studies by double digits, without incurring any additional data-gathering costs (or equivalently, reducing the data-gathering costs by double digits while maintaining the desired accuracy level). In the second illustration, we apply the proposed model to repeated cross-sectional surveys that track customer perceptions and satisfaction for an automotive dealer, a situation often encountered by marketing researchers.

1. Introduction

Tracking studies play an important role in monitoring population dynamics for various social, political, economic and business purposes. These studies typically rely on two basic sampling schemes: a) longitudinal panels, where data are gathered from the same sample of individuals over time, and b) repeated cross-sections, where data are gathered from different, independent samples in each period. Although both sampling schemes can be used to track population-level dynamics, only longitudinal panels can capture individual-level dynamics (e.g., within-individual attitude or behavior changes). However, because of the heavier burden on the participants in a longitudinal panel and the resulting challenges in recruiting and retaining panel members, maintaining representative longitudinal panels over an extended period of time is much more costly than recruiting repeated cross-sections (Hsiao, 2007). Indeed, the extra costs of longitudinal panels are in many cases unjustifiable because "relatively few analyses are truly longitudinal" (Tourangeau, 2003, p. 7) in the strict sense of studying individual-level dynamics.

Besides being less costly, repeated cross-sections can often provide a better representation of changing populations than do longitudinal panels. For example, with highly mobile populations such as the younger and less affluent, it is often difficult to track panel members as they move, further eroding sample representativeness. Furthermore, because the typical longitudinal panel maintains a static sample over time, sample size is limited, preventing time aggregation to increase sample representativeness over longer time intervals. In contrast, repeated cross-sections can be aggregated over time, enabling researchers to study small sub-populations over coarser time intervals. As a result, repeated cross-sections are far more common than longitudinal panels in tracking studies (Hsiao, 2007; Tourangeau, 2003).

In light of the above, our focus in this paper is *not* on tracking studies whose main interest is in estimating individual-level dynamics (which

* Corresponding author at: Melcher Hall 375E, 4800 Calhoun Rd., Bauer College of Business, University of Houston, Houston, TX 77204-6021, United States. Tel.: +1 713 743 9277; fax: +1 713 743 4572.
E-mail addresses: rexdu@bauer.uh.edu (R.Y. Du), kamakura@rice.edu (W.A. Kamakura).
[1] Tel.: +1 713 348 6307.

http://dx.doi.org/10.1016/j.ijresmar.2014.10.002
0167-8116/

require longitudinal panels). Rather, we focus on tracking studies whose main interest is in monitoring the 'state of population'. In particular, we focus on repeated cross-sectional surveys whose main goal is to monitor population means for variables of interest that are measured on interval or ratio scales. Such tracking studies are prevalent in marketing and virtually all the other social sciences, e.g., the Consumer Expenditure Survey and American Time Use Survey conducted by the U.S. Bureau of Labor Statistics, the Survey of Consumers conducted by Thomson Reuters-University of Michigan, and numerous syndicated trackers on product consumption, brand health and customer satisfaction that rely on repeated cross-sections drawn from omnibus panels maintained by large marketing research companies (e.g., Kantar Worldpanel, NPD, YouGov, Vision Critical).

The most common approach in dealing with repeated cross-sectional survey data is to pool the responses from all those interviewed in the same period, calculate the sample averages, and use those sample averages as estimates of population means in the corresponding time period. Such an approach is easy to implement but faces a major challenge — random sampling errors are confounded with genuine changes in population means. When one observes two sample averages from two different time periods, one does not know the extent to which the difference between these two sample averages is caused by sample composition differences or changes in population means. The former is purely a function of who were drawn into each sample and is therefore of little interest to the researcher. The latter is what the researcher is truly interested in uncovering.

The above confound is exacerbated when the sample size in each time period is small and the population is heterogeneous, as both factors lead to large random sampling errors. As sampling errors increase, true signals about population means become harder to detect, leading to not only more inaccuracies but also more volatility in the estimates of population means. In this paper, we develop a method that can substantially improve the statistical performance of tracking studies based on repeated cross-sections, by better separating random sampling errors from genuine changes in population means. To accomplish such a goal, we take full advantage of three patterns that are commonly embedded in repeated cross-sectional survey data:

1. Inter-temporal dependence in population means. While individual responses from non-overlapping cross-sections are independent over time, they reflect, collectively, the state of the population in each time period, which is obviously temporally dependent. For example, it should be rare that a brand's health would vary dramatically from one month to the next, even though the perceptions of individuals sampled in one month are independent from the perceptions of individuals sampled in the next month. By formally taking into account inter-temporal dependence in population states, our proposed model borrows information from all time periods in inferring the population means in any given period. This implies that, in estimating population means over time, our model smoothens the raw sample averages by filtering out larger-than-expected fluctuations (with the expected level of smoothness empirically determined), attributing these unusual shifts more to random sampling errors than to changes in population states.
2. Temporal co-movements among population means of multiple measures. In most tracking surveys, researchers gather data on multiple variables, many of which related to the same underlying constructs (e.g., customer attitudes with respect to different aspects of a product). To the extent that population means of these measures are manifestations of the same underlying population state and sample averages are manifestations of population means, the movements of the sample averages should be correlated over time. In other words, by formally taking into account temporal co-variations among sample averages, our proposed model borrows information across the sample averages of all measures in inferring the population means of any given measure. Intuitively, this implies that our model filters out idiosyncratic movements in any given measure's sample averages by triangulating them against how the other measures' sample averages move, with the expected pattern of co-movements empirically determined.
3. Cross-sectional co-variations across multiple measures. Due to factors such as common method bias, halo effect, heterogeneity in scale usage and other respondent characteristics, a respondent's answers to multiple questions from the same survey can be correlated with one another. When respondent-level data are available, such within-sample between-measure correlation will manifest itself and therefore can be uncovered from cross-sectional co-variations. However, if only sample averages were available, due to random differences in sample composition from one time period to another, cross-sectional between-measure co-variations would lead to spurious temporal co-movements among the measures' sample averages, which, unfortunately, would be confounded with genuine temporal co-movements in the measures' population means. In other words, in order to disentangle cross-sectional between-measure co-variations from temporal between-measure co-movements in population means, one needs to take advantage of tracking data at the respondent level, which our proposed model allows us to accomplish.

In the rest of the paper, we proceed as follows. We first review existing methods that can potentially be utilized to alleviate the confound between random sampling errors and genuine changes in population means, highlighting how each method leverages one or two of the three data patterns mentioned above. We then present our proposed approach, which, by simultaneously leveraging the three aforementioned patterns that are commonly embedded in repeated cross-sectional survey data, goes beyond all that has been attempted in the literature. To test our methodology and better understand the incremental value of leveraging each of the three data patterns, we conduct a Monte Carlo simulation using data from a known population from which we draw repeated cross-sections. Given that we know the true population means in each time period, we can make equitable comparisons in statistical performance across different approaches. After we thoroughly test the performance of our proposed model against known population means and benchmark models, we illustrate its use by applying it to data gathered through repeated cross-sectional surveys of customer perceptions and satisfaction for an automotive dealer, a situation that is often encountered by marketing researchers.

2. Tracking population dynamics with repeated cross-sectional samples

A seminal study on the analysis of repeated surveys was published by Scott and Smith (1974), who were the first to realize that while the observations from each wave of surveys might be independent, the population means being estimated from the sample averages could in fact be temporally dependent. Depending on the assumed inter-temporal dependency of the population means and the type of repeated cross-sectional sampling (overlapping or not), Scott and Smith (1974) and Scott, Smith, and Jones (1977) suggested different ARIMA models to better infer, from sample averages over time, the trend line of the population mean of a single response variable. They referred to this model-based approach for inferring population means over time as a *secondary analysis* of repeated cross-sectional data, as it relies on sample averages as inputs, as opposed to a *primary analysis* of raw respondent-level data. One recent example in the marketing literature of applying time series models to sample averages is Srinivasan, Vanhuele, and Pauwels (2010), who used a vector-autoregressive (VARX) market-response model in investigating the dynamics between marketing mix, brand sales and consumer mindset metrics, which were gathered through repeated cross-sectional surveys. Their focus, however, is not on better inferring population means from sample averages over time, and their model does not distinguish between random sampling errors and genuine shifts in the state of the population.

Other authors (Feder, 2001; Harvey & Chung, 2000; Pfefferman, 1991; Tam, 1987) have viewed the potential inter-temporal dependency of population means from a state-space perspective (Kalman, 1960), where the population means are treated as latent state variables that evolve over time and the sample averages obtained from repeated cross-sectional surveys are noisy manifestations of the underlying population means. From such a state-space perspective, these authors typically specify an *observation equation* such as,

$$y_t = \mu_t + \varepsilon_t \tag{1}$$

where y_t is a vector of sample averages for M measures at time t, μ_t is a latent vector representing the corresponding population means at time t, and ε_t is a vector of random sampling errors, such that $E(\varepsilon_t) = 0$; $E(\varepsilon_t \varepsilon'_{t-k}) = 0, k > 0$; $E(\varepsilon_t \varepsilon'_t) = V$. Instead of being treated as temporally independent, μ_t (i.e., the population means that are not directly observable) is assumed to evolve from one time period to the next following a state equation such as,

$$\mu_t = \mu_{t-1} + \xi_t, \tag{2}$$

where ξ_t is a vector of latent state shocks, such that $E(\xi_t) = 0$, $E(\xi_t \xi_{t-k}') = E(\xi_t \varepsilon_{t-k}') = 0$, $k > 0$, $E(\xi_t \xi_t') = Q_t$, and $E(\xi_t \varepsilon_t') = 0$.

Like ARIMA and VARX, the above state-space model formulation also falls into the category of secondary analysis, where sample averages are used as inputs, instead of raw respondent-level data. Unlike ARIMA and VARX, the above state-space model formulation makes an explicit attempt at separating genuine changes in population means from random sampling errors. This separation between signal and noise is accomplished by leveraging the fact that there should be inter-temporal dependence in population means and that random sampling errors should be independent from one period to another. In the marketing literature, many studies have applied this idea of filtering and smoothing in separating signals from noises in aggregate time series measures (see Xie, Song, Sirbu, & Wang, 1997 for an application of the Kalman filter and Pauwels & Hanssens, 2007 for an application of the Hodrick–Prescott filter).

More recently, marketing researchers (e.g., Du & Kamakura, 2012; Norris, Peters, & Naik, 2012) have applied dynamic factor analysis (DFA), a special case of the general state-space model, in analyzing multivariate time series data with random measurement errors. Like the state-space formulation delineated in Eqs. (1) and (2), a DFA model makes an explicit attempt at separating genuine changes in the latent state variables from random measurement errors in the manifest variables. Unlike Eq. (2), a DFA model in our context assumes that the M-dimensional (unobservable) population means μ_t can be expressed a linear function of N-dimensional latent factors z_t, which evolves over time following, for example, a random walk:

$$\mu_t = \Lambda z_t \tag{3}$$

$$z_t = z_{t-1} + \xi_t \qquad \xi_t \sim N(0, \Omega), \quad z_0 \sim N(a_0, \Omega_0). \tag{4}$$

Because the dimensionality of the latent factors z_t is typically much smaller than that of the population means μ_t (i.e. $N \ll M$), a DFA model such as the above provides a parsimonious way to capture the temporal co-movement patterns embedded in the evolution of the population means. This is important in the context of tracking studies because most of these studies measure multiple variables, and many of these variables can be driven by common underlying factors (e.g., consumer expenditures across different product categories tend to move together because they are all a function of discretionary income). To the extent that sample averages are manifestations of population means, co-movements in the latter will manifest in temporal co-variations in the former. In other words, by tapping into temporal co-variations in sample averages, a DFA model borrows information from the sample averages of all measures in inferring the population means of any given measure. Intuitively, this implies that a DFA model filters out idiosyncratic movements in any given measure's sample averages by triangulating them against how the other measures' sample averages move over time, following an empirically-determined factor-analytic co-movement pattern.

In short, DFA models currently available in the marketing literature (Du & Kamakura, 2012; Norris et al., 2012) can potentially improve the statistical performance of repeated surveys by simultaneously leveraging two patterns that are commonly embedded in the data: inter-temporal dependence in population means of a single variable, and temporal co-movements among population means of multiple variables. However, to the best of our knowledge, the focus of DFA models in the marketing literature has been on extracting the latent dynamic factors, i.e., z_t in Eqs. (3) and (4) (e.g., for spotting common underlying trends in Du & Kamakura, 2012, or for making causal inferences regarding the relationship between advertising and how consumers think-feel-do about brands in Norris et al., 2012). None of the DFA models in the marketing literature have been applied in the context of the current study; that is, how population means can be better inferred from tracking data based on repeated cross-sections.

One limitation of applying DFA to aggregate tracking data based on repeated cross-sections is that it still falls into the category of secondary analysis, as opposed to *primary* (i.e., respondent-level) analysis (Pfefferman, 1991). When researchers have access to individual-level data, like most vendors of tracking surveys and their clients do, information at the respondent level is ignored (Lind, 2005) in the Secondary DFA. In particular, a respondent's answers to multiple survey questions can be correlated with one another, which in turn would lead to cross-sectional between-measure co-variations. By tapping into respondent-level data, as opposed to sample averages only, researchers can better distinguish between temporal co-movements in population means from cross-sectional co-variations in individual responses, because the latter can be better identified through the pattern of within-sample between-measure co-variations. In other words, population means can potentially be inferred with more accuracy by replacing the Secondary DFA of sample averages with the primary analysis of respondent-level data, as we will demonstrate later.

Extending a state-space model from the secondary analysis to the primary analysis seems straightforward at a first glance. One could simply replace Eq. (1) as follows and keep the rest of the model unchanged:

$$y_{i(t)} = \mu_t + \varepsilon_{i(t)} \qquad \varepsilon_{i(t)} \sim N(0, \Sigma) \tag{5}$$

where $y_{i(t)}$ denotes the M responses from respondent i who is sampled in period t, μ_t denotes the latent population means in period t, and $\varepsilon_{i(t)}$ denotes respondent i's deviations from the population means, which is assumed to be normally distributed with mean 0 and variance-and-covariance Σ.

In reality, however, implementing the above state-space model faces the "curse of dimensionality" and is often deemed infeasible because it requires inverting matrices whose dimensionality equals the number of response variables multiplied by the sample size for each time period in each iteration of the Kalman filter (Lind, 2005, p. 4). For studies of realistic sample sizes (say $N_t = 500$) that track, say, $M = 10$ response variables, it would require the inversion of a 5000×5000 matrix for each time period in each iteration of the Kalman filter. Fortunately, Lind (2005), through ingenious matrix algebra, shows how the computations can be simplified considerably, requiring only the inversion of an $M \times M$ matrix in a modified Kalman filter (Eq. (10) on p. 4 and the proof in the Appendix on pp. 8–9 of Lind, 2005). With Lind's simplified Kalman filter, state-space models become computationally feasible in performing the primary analyses of tracking data based on repeated cross-sections, as opposed to only using the sample averages.

Surprisingly, to the best of our knowledge, Lind's (2005) ingenious solution has never been put to practical use in marketing (or, for that matter, in any other disciplines) and Lind's modified Kalman filter has been relatively obscure (with only seven citations in almost 10 years). Building on Lind's work, in the next section we propose our Primary DFA model for inferring population means from respondent-level data, leveraging on two patterns that are commonly embedded in repeated cross-sectional data: inter-temporal dependence in population means and cross-sectional between-measure co-variations. By imposing a parsimonious dynamic factor structure on the latent states, our model can also tap into one additional pattern in the data: temporal co-movements among population means, a source of information that was not efficiently leveraged in previous research, including Lind (2005). Finally, aside from putting Lind's (2005) algorithm to practical use for the first time, we also examine the statistical performance of the Primary DFA. To the best of our knowledge, no prior research has systematically examined the actual empirical performance of the primary analysis using state space models.

3. A state-space model for the primary analysis of repeated cross-sectional data

Let $y_{i(t)}$ denote an $M \times 1$ vector of observed indicators for respondent i in period t in a tracking study based on repeated cross-sectional surveys. The goal is to infer the corresponding population means. The observation equation for our state-space model is,

$$y_{i(t)} = \mu_t + \varepsilon_{i(t)} \qquad \varepsilon_{i(t)} \sim N(0, \Sigma) \tag{6}$$

and the state equations are,

$$\mu_t = \Lambda z_t + B x_t \tag{7}$$

$$z_t = z_{t-1} + \xi_t \qquad \xi_t \sim N(0, \Omega), \quad z_0 \sim N(a_0, \Omega_0) \tag{8}$$

where, $\mu_t(M \times 1)$ denotes the population mean vector in period t, $z_t(K \times 1)$ denotes a vector of latent states, $\Lambda(M \times K)$ maps the latent states z_t into the population means μ_t, $x_t(H \times 1)$ is a vector of observed covariates in period t, and $B(M \times H)$ maps the covariates x_t into the population means μ_t. $\Lambda(M \times K)$, $B(M \times H)$, $\Sigma(M \times M)$, $\Omega(K \times K)$, a_0, and Ω_0 are to be determined empirically.

The observation Eq. (6) relates $y_{i(t)}$ (M responses for respondent i in period t) to the corresponding population means (μ_t) and deviations from the means ($\varepsilon_{i(t)}$), which is assumed to normally distributed with mean zero and covariance Σ. The first state Eq. (7) specifies a factor decomposition of the population means into a K-dimensional latent space defined by the loadings Λ and latent states z_t, along with a regression of the population means on a vector of covariates x_t. The second state Eq. (8) defines the temporal structure of the latent states z_t. In its current form, this state equation specifies a simple random walk, which can be extended to more elaborate structural time series formulations (e.g., Du & Kamakura, 2012).

For the covariance (Σ) of the respondent deviations ($\varepsilon_{i(t)}$) from population means, we estimate a full matrix to capture cross-sectional between-measure co-variations, which can arise due to factors such as common method bias, halo effect, heterogeneity in scale usage and other respondent characteristics. Controlling for co-variations of $\varepsilon_{i(t)}$ across respondents is important because it helps take into account random differences in sample compositions. Such differences arise because, from one period to another, each sample may include disproportionately more or fewer respondents from certain sub-populations. Without properly accounting for these unobservable differences in sample composition, the resulting temporal co-variations in sample averages would be confounded with genuine changes in population means over time.

The covariance (Ω) of the latent state shocks (ξ_t) is fixed to be diagonal. We do so for two reasons. First, in DFA, a diagonal Ω has routinely been assumed for identification purposes (Harvey, 1989, pp. 450–451; Zuur, Fryer, Jolliffe, Dekker, & Beukema, 2003), the rationale of which is the same as in standard factor analysis.[2] Second, a diagonal Ω would imply that the latent states are independent of one another, which is an attractive property because it facilitates interpreting the extracted dynamic factors (Du & Kamakura, 2012, Web Appendix, p. 6). It is important to note that, although the dynamics of each latent factor is independent of one another, these factors together drive the evolution of the population means, resulting in a temporal co-movement pattern in the population means that is governed by Λ, the factor loading matrix.

The specification of our model, as delineated through Eqs. (6), (7) and (8), is deceivingly simple. It in fact combines core ideas from two sophisticated state-space models: DFA for the secondary analysis (Du & Kamakura, 2012; Norris et al., 2012) and Lind's (2005) model for the primary analysis. Compared with DFA, our proposed model can account for cross-sectional between-measure co-variations as it leverages all respondent-level data, as opposed to sample averages only. Compared with Lind's primary analysis model, which specifies a one-to-one relationship between the population means and the latent states, our proposed model accounts for temporal co-movements among population means through a parsimonious dynamic factor structure, an idea borrowed from DFA. We argue that this extension to Lind's model is not trivial because our specification requires potentially a much smaller state space as the same latent factor may contain information related to more than one population mean. Second, the factor structure accounts for the possibility that the different population means might exhibit co-movements over time, allowing more efficient between-measure information sharing.

3.1. Model estimation and interpretation

For model estimation, one can stack the respondent-level data within each time period t:

$$y_t \equiv \left(y_{1(t)}{}' \quad y_{2(t)}{}' \quad \cdots \quad y_{N_t(t)}{}' \right)' \qquad (MN_t \times 1) \tag{9}$$

$$\varepsilon_t \equiv \left(\varepsilon_{1(t)}{}' \quad \varepsilon_{2(t)}{}' \quad \cdots \quad \varepsilon_{N_t(t)}{}' \right)' \qquad (MN_t \times 1) \tag{10}$$

$$J_t \equiv 1_{N_t} \otimes I_M \qquad (MN_t \times M). \tag{11}$$

Given the above, one can rewrite the model delineated in Eqs. (6), (7) and (8) into a general state-space form:

$$y_t = J_t \Lambda z_t + J_t B x_t + \varepsilon_t \qquad \varepsilon_t \sim N\left(0, I_{N_t} \otimes \Sigma\right) \tag{12}$$

$$z_t = z_{t-1} + \xi_t \qquad \xi_t \sim N(0, \Omega). \tag{13}$$

We estimate the above general state-space model via the Expectation–Maximization estimator (Shumway & Stoffer, 2000), where the expectation-step involves Lind's simplified Kalman filter and smoother. The maximization-step has a closed-form solution for each model parameter. Details about the EM estimator that is tailored for our proposed model are provided in Appendix A.

Estimates of Σ capture cross-sectional covariance in respondent deviations from population means. The size of Σ shall highlight the built-in advantage of the primary analysis over the secondary analysis, because

[2] The main difference between DFA and standard factor analysis lies in the fact that DFA is applied to multivariate time series data while the standard factor analysis is applied to multivariate cross-sectional data.

in any secondary analysis of sample averages, the cross-sectional covariance observed within each time period is lost due to aggregation, making it harder to disentangle random sample composition differences from genuine population state changes.

The estimates of Λ can be interpreted in the same way as one would interpret loadings in the standard factor analysis. Population means that have large loadings on the same set of factors tend to exhibit strong co-movements over time. Similarly, the estimates of $z_{t|T}$ can be interpreted in the same way as one would interpret factor scores in the standard factor analysis, except that they represent movements over time in the latent state space. By tracing $z_{t|T}$ over time one can detect the predominant trends in the latent state space defined by Λ, as we will illustrate later. The estimated variance of $z_{t|T}$ (i.e., $\text{var}(z_{t|T})$) determines the uncertainties associated with the latent states. Estimates of population means at time t can be obtained as $\hat{\mu}_t = \hat{\Lambda} z_{t|T} + \hat{B} x_t$.

4. An empirical validation of the primary dynamic factor analysis of repeated cross-sectional data

In order to validate our proposed model, we need a realistic "testing ground" where we have individual-level tracking data for each member of an entire population, so that we know the true population means in each time period. We can draw random samples from the entire population in each period, mimicking tracking data that would have resulted from repeated cross-sectional surveys. By applying our model to the sampled data, we can compare the model estimates with the true population means. By repeating this process numerous times, we can assess the accuracy of our model in inferring population means. Moreover, in building this testing ground, we want to have multiple measures that are correlated both cross-sectionally and longitudinally, which would allow us to investigate the ability of our model in leveraging 1) within-measure inter-temporal dependence, 2) between-measure co-movements in the population means, and 3) between-measure cross-sectional co-variations.

To accomplish the above, we use a dataset provided by a grocery chain, which contains purchase history for all the retailer's loyalty-club members residing in one large metropolitan area in the U.S. In total, this dataset comprises of 153,540 loyalty-club members for whom we have information on monthly spending with the focal retailer, broken down into 15 major product categories over a period of 36 months. We treat this group of 153,540 loyalty-club members as the target population, and the objective is to see how accurately we can infer the mean category expenditure per household in each month by relying on repeated cross-sections randomly sampled from the population. Admittedly, the focal retailer has no need to sample from a population of its own loyalty-club members. However, because we know the true population means over an extended period of time, we have the rare opportunity to conduct a Monte Carlo study based on real-life data, systematically evaluating the statistical performances of simple sample averages vs. secondary analysis vs. primary analysis, and large vs. small sample sizes. In particular, we compare the population mean estimates obtained with our proposed model (hereafter referred as Primary III or Primary DFA) with the following five alternatives, all using the same set of repeated cross-sections in each replication:

1. Sample averages from repeated cross-sections. This approach does not involve any modeling and is probably the most common in practice. By comparing this approach with the others we can see how much one can improve the accuracy of tracking studies by better leveraging various patterns embedded in repeated cross-sectional multi-dimensional data. We refer to this approach hereafter as sample averages.
2. Apply a dynamic factor model (Eqs. (1), (7) and (8)) to the multivariate sample averages directly, assuming that the covariance of the error terms in Eq. (1) (V) is diagonal. This falls into the modeling framework that Du and Kamakura (2012) have proposed for analyzing multivariate time series. Instead of focusing on extracting the common underlying trends, our focus here is on recovering the trends of the individual population means from sample averages. Unlike our proposed model, this approach cannot separate a) between-measure cross-sectional covariance that is caused by unobserved individual heterogeneity within each sample from b) temporal co-movements that is caused by co-evolution of population means over time, because it confounds differences in sample composition with population state changes. We refer to this approach hereafter as Secondary I or Secondary DFA.
3. The same as (2), except that the covariance of the error terms in Eq. (1) (V) is allowed to be non-diagonal. Albeit minor, this extension is new to the literature. Comparing this approach with (2) we can see if allowing V to be non-diagonal can help (at least partially) separate between-measure cross-sectional covariance from temporal co-movements in population means. We refer to this approach hereafter as Secondary II.
4. A special case of our proposed model, with the covariance of the error terms in Eq. (6) (Σ) assumed to be diagonal. This model is novel to the literature in and of itself. Comparing this special case with our proposed more general model can determine how much a full Σ matrix can help further separate between-measure cross-sectional covariance from temporal co-movements in population means. We refer to this alternative hereafter as Primary I.
5. A less parsimonious version of the proposed model, where z_t has the same dimensionality as μ_t and Λ is diagonal (to ensure identification). Comparing this approach with our proposed model helps determine how temporal co-movements among population means can be better leveraged through a dynamic factor structure imposed on the latent state space, as opposed to specifying a one-to-one relationship between the population means and the latent states, a less parsimonious and potentially less efficient approach. We refer to alternative (5) hereafter as Primary II.

4.1. Illustrative results based on a single replication

Before presenting the comprehensive validation tests across 100 replications using three sample sizes (200, 500, and 1000), to illustrate the main features of our proposed model Table 1 presents the parameter estimates from one replication, which draws an independent random sample of 500 customers each month.

One of the key features of our model is the factor structure imposed on the latent space. As in standard or dynamic factor analysis, we need to first determine the "optimal" number of factors. Following Du and Kamakura (2012), we rely on the Bayesian Information Criteria (BIC) in determining the dimensionality of the factor structure, which strikes a balance between fit and parsimony by imposing a penalty on each additional factor. As it turns out, a three-factor structure would be the best when the size of each sample is either 500 or 1000, and a two-factor solution would be the best when the size of each sample is 200. Furthermore, like standard or dynamic factor analysis, the loading matrix Λ in our model is identified up to an orthogonal rotation. Again following Du and Kamakura (2012), we apply a varimax rotation at the end of the M-step after the EM estimator converges. Finally, to control for potential seasonality in household expenditures for each product category, we include 12 monthly dummies as covariates (x_t). The top half of Table 1 reports the estimates of the regression coefficients (B) for the monthly dummies, along with those of the loading matrix (Λ).

The loading estimates in our model can be interpreted in the same way as those from standard factor analysis. However, unlike standard factor analysis, where the interpretability of the estimated factor loadings is of central importance, the loading estimates in our model are merely a means to an end; they are the weights to be used to obtain population mean estimates for each of the manifest variables (i.e., dollar spend per household in each product category in each month). In other

Table 1
Parameter estimates for the proposed model (repeated cross-sections of 500 customers each month).

	Loading			Seasonal regression coefficients											
Category	λ1	λ2	λ3	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Bakery	0.02	0.20	0.04	7.2	7.3	8.5	6.4	6.5	6.4	6.7	7.4	6.2	6.6	6.7	6.5
Beer_wine_liquor	0.04	0.05	0.07	3.0	3.9	5.6	3.0	2.8	3.2	3.0	4.3	3.3	4.6	3.8	3.4
Candy	0.00	0.01	0.01	5.2	3.2	4.7	2.6	3.3	3.7	4.0	3.4	2.8	2.7	3.1	3.1
Dairy	0.49	0.23	0.12	24.8	26.8	29.7	26.4	23.1	25.7	25.7	25.4	22.6	24.7	25.4	22.9
Deli	0.26	0.10	0.10	11.8	11.1	13.7	11.2	10.4	10.9	11.2	12.8	11.5	12.8	12.4	10.6
Frozen	0.21	0.11	−0.06	18.7	18.7	19.7	20.0	17.4	19.3	19.3	20.6	18.7	19.6	18.9	17.8
Grocery_edible	1.04	0.00	0.00	61.8	66.6	71.0	64.6	56.8	63.1	62.5	69.2	62.3	64.0	64.4	59.0
Grocery_inedible	0.42	0.15	0.14	19.3	19.4	23.1	20.6	17.9	19.5	19.6	22.0	20.1	19.7	20.3	18.4
HBC	−0.05	0.18	0.09	12.5	12.4	13.9	13.7	11.8	13.0	13.6	14.2	12.4	12.7	13.3	11.7
Meat	0.22	0.37	−0.02	23.9	24.6	29.0	24.5	21.6	23.5	23.6	28.4	25.3	26.7	25.2	23.4
Pharmacy	0.29	0.12	−0.05	6.5	7.8	8.7	7.8	8.4	6.8	7.0	8.1	6.9	6.1	8.0	7.2
Prepared_foods	−0.06	0.22	0.07	4.1	4.6	4.7	4.7	3.6	4.5	4.2	5.1	4.4	4.5	4.6	4.1
Produce	0.21	0.21	0.12	21.7	21.8	24.3	24.3	21.7	23.7	24.0	27.8	26.7	28.0	23.0	20.3
Seafood	0.05	0.06	0.06	3.5	3.0	5.4	3.6	3.5	4.1	3.2	3.4	3.6	3.9	3.0	3.5
Others	0.04	−0.09	−0.13	6.9	7.4	11.2	6.0	7.4	6.8	8.1	10.2	7.8	6.6	7.0	6.3

Error covariance (Σ)	Bakery	Beer_wine_liquor	Candy	Dairy	Deli	Frozen	Grocery_edible	Grocery_inedible	HBC	Meat	Pharmacy	Prepared_foods	Produce	Seafood	Others
Bakery	147.7	0.04	0.22	0.35	0.36	0.27	0.38	0.27	0.23	0.34	0.07	0.25	0.34	0.14	0.27
Beer_wine_liquor	8.7	269.7	0.03	0.10	0.09	0.08	0.11	0.06	0.05	0.13	0.00	0.03	0.10	0.07	0.06
Candy	22.5	4.5	71.8	0.32	0.26	0.26	0.41	0.27	0.27	0.24	0.06	0.12	0.27	0.08	0.25
Dairy	123.0	45.6	78.5	834.0	0.50	0.56	0.77	0.44	0.44	0.55	0.09	0.18	0.64	0.24	0.33
Deli	84.8	27.6	42.1	277.0	2.10	0.35	0.55	0.31	0.29	0.44	0.07	0.22	0.50	0.22	0.30
Frozen	86.8	33.4	59.1	431.1	177.5	700.4	0.63	0.37	0.35	0.44	0.06	0.15	0.41	0.16	0.24
Grocery_edible	326.0	121.7	243.9	1562.1	744.1	1169.0	4975.8	0.54	0.50	0.61	0.09	0.22	0.63	0.24	0.37
Grocery_inedible	118.5	37.0	82.5	459.4	215.1	347.4	1354.1	1283.6	0.46	0.35	0.09	0.16	0.35	0.14	0.29
HBC	65.5	21.1	54.8	301.5	134.6	217.9	834.2	389.6	565.9	0.29	0.13	0.17	0.39	0.16	0.32
Meat	148.4	148.4	73.5	570.7	307.5	416.8	1549.1	448.2	250.5	1301.6	0.06	0.17	0.49	0.25	0.25
Pharmacy	26.8	−2.0	15.2	77.3	38.6	46.0	197.7	97.5	91.9	67.0	932.9	0.06	0.06	0.02	0.08
Prepared_foods	38.3	5.5	12.4	64.8	53.4	50.7	191.1	71.8	51.8	75.3	23.7	155.3	0.20	0.11	0.19
Produce	137.7	55.7	75.8	606.4	321.4	361.4	1474.7	415.2	307.5	585.8	61.1	82.2	1089.2	0.33	0.33
Seafood	19.6	12.4	7.8	80.6	48.9	47.9	196.2	56.9	43.7	103.9	7.9	15.1	123.7	131.2	0.14
Others	52.7	14.8	31.9	150.1	92.7	100.3	414.5	162.8	121.9	139.9	38.2	37.0	174.0	24.5	249.2

Note: The upper diagonal cells contain correlations, while all the other cells contain variance–covariances.

words, of central importance for our model is how close $\hat{\mu}_t = \hat{\Lambda} z_{t|T} + \hat{B} x_t$ are to the true population means.

The bottom half of Table 1 reports the estimates of Σ, the cross-sectional co-variances of individual deviations from population means. We see a positive correlation in household spend across most of the product categories, indicating that households spending more in some categories (e.g., *bakery* and *dairy*) are likely to spend more on others as well (e.g., *grocery edible* and *meat*). This is intuitive because unobserved household heterogeneity (e.g., income, family size or lifestyle, and loyalty to the focal retailer) can lead to above- or below-average spending across multiple product categories. Applied to individual-level data, our model for the primary analysis captures this pattern through Σ and is thus able to separate 1) temporal co-movements in sample averages that are caused by a combination of cross-sectional co-variation and changing sample composition from 2) temporal co-movements in sample averages that are caused by changing population means, which is captured through the factor loading matrix Λ.

Given the model parameter estimates and the observed data, we can compute $z_{t|T}$, the latent dynamic factor scores, through the Kalman filter and smoother (see Appendix A for details). Given $z_{t|T}$, we can produce estimates of the population means as $\hat{\mu}_t = \hat{\Lambda} z_{t|T} + \hat{B} x_t$. To illustrate, for three randomly selected categories, Fig. 1 plots $\hat{\Lambda} z_{t|T}$ from the same replication that led to the parameter estimates in Table 1. $\hat{\Lambda} z_{t|T}$ can be interpreted as de-seasonalized trends in monthly category expenditure per household. For comparison, Fig. 1 also plots $\hat{\Lambda} z_{t|T}$ from Secondary I, an alternative approach that falls into the modeling framework of Du and Kamakura (2012) and can only be applied to sample averages.

From Fig. 1, we see that our proposed Primary DFA produces trend lines that are much smoother than those produced by the Secondary DFA, which seem to exhibit more random fluctuations from one month to another. This contrast highlights the advantage of the primary analysis over the secondary analysis in filtering out random sampling

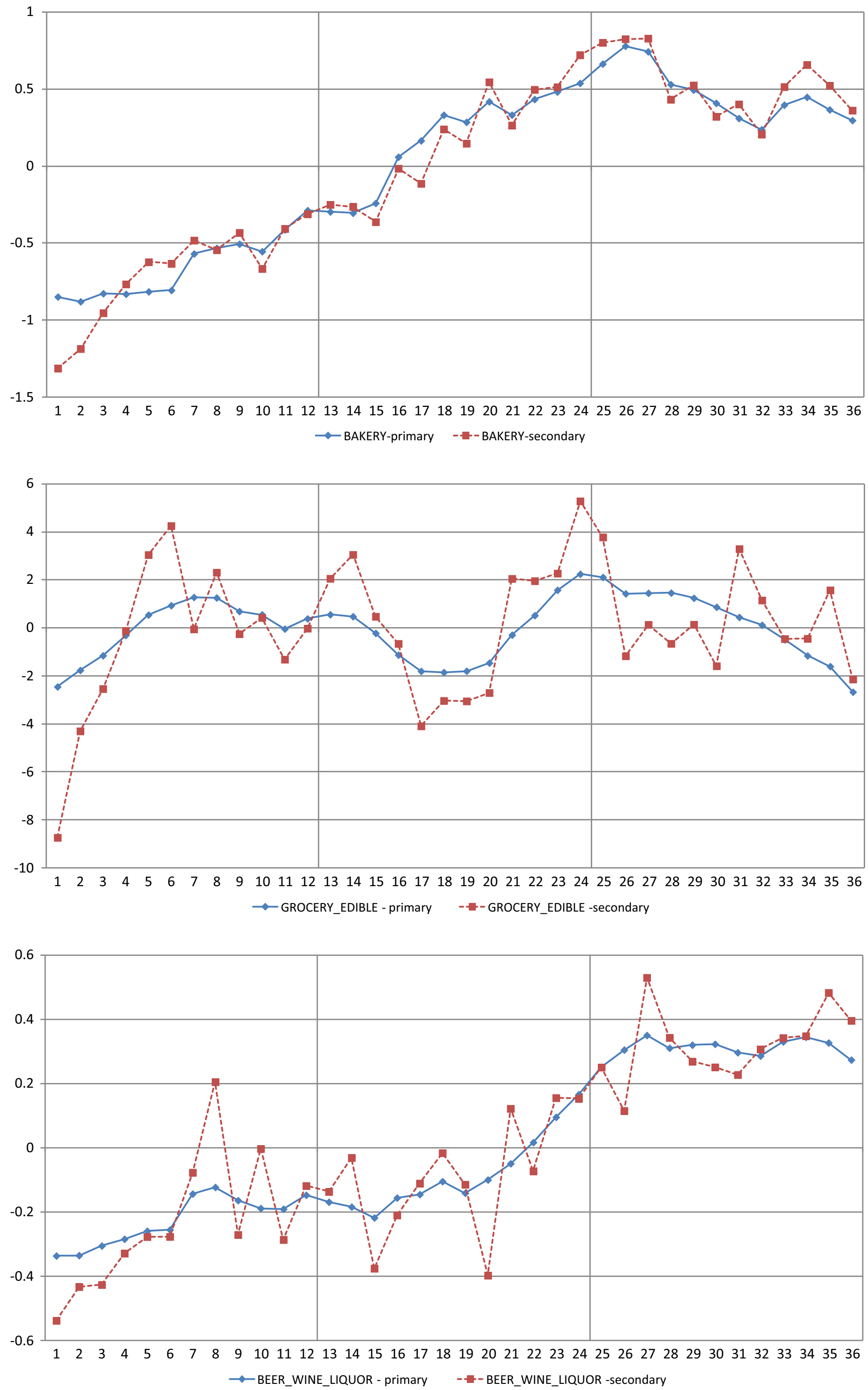
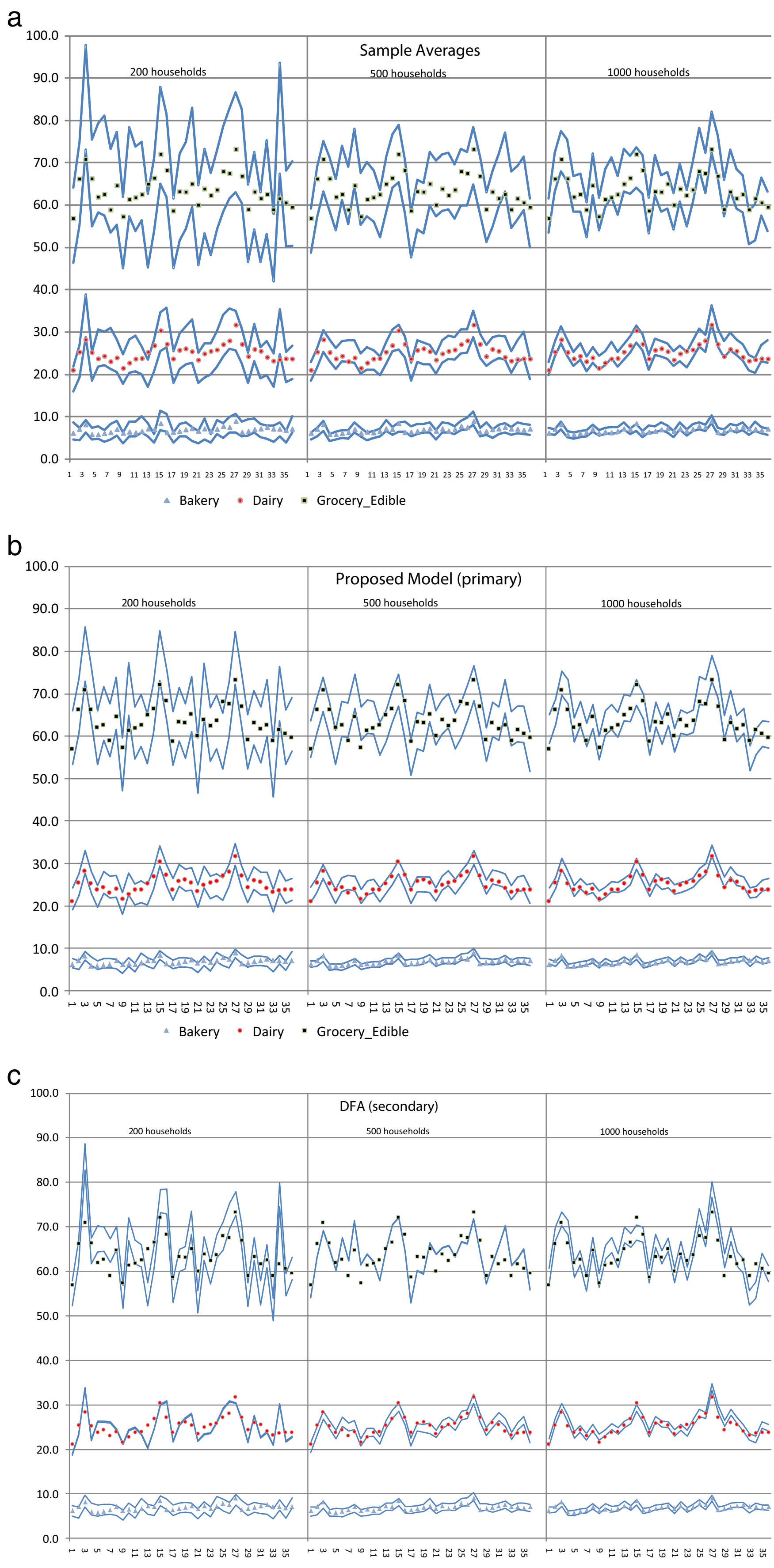


Fig. 1. De-seasonalized trends in category expenditure per household produced by the proposed model (Primary DFA) and Secondary DFA.

errors and producing more stable population mean estimates. The source of the advantage resides in the fact that the primary analysis uses information contained in the entire data, as opposed to just the sample averages.

Another way to appreciate the differences in statistical performance of the different approaches is via Fig. 2, which plots the confidence intervals for the estimated population means obtained from 1) the Primary DFA, 2) sample averages, and 3) Secondary DFA. In these plots, the dots represent the actual population means, while the solid lines show the 95% confidence intervals of the estimates. We see that our proposed primary analysis produces interval estimates of the population means that are tighter than those produced by simple sample averages.

Fig. 2. a. Confidence intervals for the estimated population means based on simple sample averages. b. Confidence intervals for the estimated population means from the proposed model (Primary DFA). c. Confidence intervals for the estimated population means based on the Secondary DFA.



Note: For 200 households Dairy and 500 households Grocery Edible, the upper and lower bounds of the confidence intervals are overlapping.

Table 2
Mean absolute deviations from the population mean, averaged across 36 months and 100 replications.

Category	Sample size	MAD	% MAD reduction over sample averages				
		Sample averages	Secondary I	Secondary II	Primary I	Primary II	Primary III
Average across All categories	200	1.423	−19.2%	−29.4%	−26.1%	−33.9%	−34.0%
	500	0.904	−16.1%	−21.7%	−22.7%	−26.6%	−27.9%
	1000	0.649	−13.9%	−16.9%	−17.8%	−16.9%	−22.5%
Bakery	200	0.725	−27.3%	−31.3%	−30.7%	−33.2%	−36.1%
	500	0.448	−23.2%	−25.4%	−27.0%	−29.1%	−31.4%
	1000	0.323	−20.4%	−22.3%	−23.0%	−21.1%	−26.7%
	200	0.920	−35.3%	−34.0%	−36.6%	−38.2%	−37.2%
Bear_wine_liquor	500	0.601	−31.6%	−31.0%	−34.3%	−34.4%	−34.2%
	1000	0.436	−29.7%	−29.9%	−32.5%	−32.8%	−33.3%
	200	0.399	−23.3%	−28.1%	−26.2%	−32.7%	−30.9%
Candy	500	0.255	−14.8%	−18.0%	−19.4%	−24.8%	−22.5%
	1000	0.177	−6.2%	−9.3%	−9.1%	−12.1%	−12.9%
	200	1.650	−7.2%	−24.1%	−18.4%	−30.7%	−33.1%
Dairy	500	1.032	−4.0%	−12.8%	−13.6%	−18.7%	−23.6%
	1000	0.731	−1.0%	−5.8%	−5.3%	−5.4%	−17.1%
	200	1.064	−19.9%	−30.0%	−25.7%	−32.2%	−33.8%
Deli	500	0.674	−15.6%	−22.4%	−22.5%	−22.3%	−27.9%
	1000	0.484	−14.8%	−18.4%	−17.5%	−11.2%	−22.1%
	200	1.504	−16.8%	−30.2%	−24.4%	−33.9%	−33.5%
Frozen	500	0.903	−14.8%	−19.4%	−20.7%	−21.8%	−24.4%
	1000	0.671	−10.5%	−12.4%	−14.8%	−10.0%	−17.5%
	200	3.991	−9.2%	−28.9%	−20.7%	−33.7%	−32.9%
Grocery_edible	500	2.477	−6.0%	−17.5%	−16.2%	−23.1%	−23.9%
	1000	1.812	−6.2%	−11.9%	−12.2%	−9.1%	−17.2%
	200	2.064	−24.4%	−32.3%	−29.2%	−36.8%	−34.9%
Grocery_inedible	500	1.318	−22.2%	−25.3%	−26.7%	−32.0%	−30.6%
	1000	0.939	−19.1%	−22.6%	−23.4%	−25.0%	−26.8%
	200	1.285	−22.3%	−30.7%	−27.7%	−34.5%	−35.2%
HBC	500	0.845	−20.0%	−26.3%	−26.4%	−31.4%	−32.5%
	1000	0.599	−17.0%	−22.3%	−21.3%	−24.6%	−28.5%
	200	1.949	−18.4%	−28.1%	−25.8%	−32.1%	−34.1%
Meat	500	1.266	−14.1%	−21.2%	−20.6%	−24.1%	−27.0%
	1000	0.891	−13.0%	−16.3%	−16.6%	−13.7%	−21.6%
	200	0.868	−21.2%	−26.3%	−25.4%	−29.6%	−30.1%
Others	500	0.564	−14.0%	−16.4%	−19.7%	−23.3%	−23.0%
	1000	0.404	−5.1%	−5.6%	−7.9%	−12.9%	−12.0%
	200	1.709	−33.2%	−33.2%	−35.3%	−39.1%	−35.4%
Pharmacy	500	1.122	−29.9%	−28.9%	−34.2%	−39.2%	−34.1%
	1000	0.819	−29.6%	−29.8%	−33.3%	−37.8%	−33.1%
	200	0.708	−24.4%	−22.6%	−31.3%	−33.0%	−33.2%
Prepared_foods	500	0.460	−20.7%	−17.8%	−26.0%	−28.1%	−28.7%
	1000	0.321	−20.7%	−16.4%	−20.1%	−23.3%	−28.5%
	200	1.859	−14.1%	−28.7%	−22.8%	−31.7%	−34.5%
Produce	500	1.185	−14.4%	−23.7%	−22.5%	−25.2%	−30.8%
	1000	0.841	−12.7%	−16.6%	−16.2%	−11.2%	−22.4%
	200	0.652	−30.1%	−31.6%	−32.4%	−35.1%	−35.0%
Seafood	500	0.405	−23.6%	−24.3%	−25.6%	−26.7%	−27.4%
	1000	0.287	−16.6%	−16.6%	−18.1%	−17.2%	−20.1%

Moreover, the plots for the secondary analysis in Fig. 2c show that modeling sample averages directly (therefore ignoring the covariances within each of the cross-sections) gives a false sense of precision. Because there are far fewer degrees of freedom, the secondary analysis over-fits the sample averages, leading to misleadingly small standard errors for the estimates. As shown in Fig. 2c, some of the 95% confidence intervals obtained by the secondary analysis are very narrow, and in many cases fail to encompass the true population means.

4.2. Monte Carlo experiment across 100 replications

For a more thorough performance comparison of the alternative approaches, we draw 100 random samples of sizes 200, 500 and 1000 each month, which allows us to simulate 100 times the typical repeated cross-sectional sampling process utilized in tracking studies. We then apply our proposed model, along with the five alternatives discussed earlier, to each of the datasets generated by this design.

We use two measures of accuracy: a) the mean absolute deviation (MAD) from the known population mean in each month (μ_t), assessing the accuracy in estimating population-wide expenditure level in each month, and b) the mean absolute deviation in month-to-month change ($\Delta_t = \mu_t - \mu_{t-1}$), assessing the accuracy in estimating the rate of change in population-wide expenditure from month to month. These accuracy measures, averaged across 36 months and 100 replications are reported in Tables 2 and 3.

Looking first at estimating the population means over time (Table 2), the full version of our proposed Primary DFA (Primary III) performs considerably better than the simple sample averages, reducing MADs (across 15 categories, 36 months, and 100 replications) by 34.0% when monthly sample size is 200, 27.9% when monthly sample size is 500, and 22.5% when monthly sample size is 1000 (see the last column and top three rows of Table 2). It is remarkable that even when monthly sample size reaches 1000, the overall estimation error can still be reduced by over 20%. Another way to look at this is: applying our model to data gathered from a monthly sample of 500 can have about the same level of estimation error as using simple sample averages based on a monthly sample of 1000 (in other words, in the context of our empirical test, as compared with using simple sample averages as

Table 3
Mean absolute deviations from month-to-month shifts in the population mean, averaged across 36 months and 100 replications.

Category	Sample size	MAD	% MAD reduction over sample averages				
		Sample averages	Secondary I	Secondary II	Primary I	Primary II	Primary III
Average across All categories	200	2.012	−22.4%	−36.6%	−31.0%	−37.9%	−38.0%
	500	1.279	−19.8%	−30.2%	−28.7%	−32.4%	−33.2%
	1000	0.920	−17.6%	−23.9%	−23.8%	−24.9%	−26.5%
Bakery	200	1.004	−29.9%	−36.3%	−33.8%	−37.5%	−37.8%
	500	0.624	−26.5%	−31.5%	−31.4%	−33.6%	−34.3%
	1000	0.456	−24.1%	−27.7%	−28.7%	−28.5%	−30.0%
	200	1.288	−37.9%	−39.8%	−39.6%	−40.7%	−41.0%
Bear_wine_liquor	500	0.875	−35.8%	−37.8%	−38.3%	−39.5%	−39.3%
	1000	0.632	−34.4%	−37.3%	−37.5%	−38.3%	−38.6%
	200	0.563	−23.6%	−30.5%	−27.5%	−31.7%	−31.7%
Candy	500	0.364	−14.4%	−20.4%	−19.7%	−22.7%	−22.9%
	1000	0.251	−2.7%	−7.1%	−5.9%	−8.9%	−9.1%
	200	2.323	−9.7%	−35.5%	−25.1%	−37.1%	−37.9%
Dairy	500	1.449	−7.5%	−25.9%	−22.5%	−26.6%	−31.2%
	1000	1.024	−4.5%	−17.0%	−15.8%	−16.2%	−22.4%
	200	1.503	−23.4%	−36.0%	−31.0%	−37.0%	−37.3%
Deli	500	0.957	−19.9%	−30.1%	−29.3%	−31.9%	−32.9%
	1000	0.686	−20.1%	−24.5%	−24.5%	−24.0%	−26.3%
	200	2.137	−19.6%	−34.9%	−28.9%	−36.2%	−36.1%
Frozen	500	1.278	−17.3%	−26.6%	−24.9%	−28.6%	−28.7%
	1000	0.951	−12.2%	−17.8%	−17.8%	−19.6%	−20.2%
	200	5.624	−11.9%	−36.3%	−27.1%	−37.7%	−37.6%
Grocery_edible	500	3.459	−9.1%	−27.2%	−24.0%	−29.3%	−30.1%
	1000	2.571	−9.6%	−18.9%	−18.6%	−19.4%	−21.4%
	200	2.971	−28.2%	−39.1%	−34.3%	−40.0%	−39.8%
Grocery_inedible	500	1.865	−26.3%	−33.0%	−32.7%	−35.2%	−35.0%
	1000	1.335	−23.3%	−29.6%	−28.5%	−30.7%	−31.1%
	200	1.830	−25.1%	−37.5%	−32.1%	−38.8%	−38.9%
HBC	500	1.203	−24.4%	−35.0%	−32.1%	−37.9%	−37.9%
	1000	0.858	−23.1%	−32.7%	−29.5%	−35.0%	−35.3%
	200	2.738	−21.2%	−37.4%	−31.4%	−38.8%	−39.3%
Meat	500	1.788	−16.6%	−29.1%	−26.5%	−31.3%	−32.2%
	1000	1.256	−15.3%	−22.1%	−22.4%	−22.8%	−24.5%
	200	1.226	−26.3%	−32.0%	−29.5%	−33.2%	−33.1%
Others	500	0.805	−17.7%	−21.9%	−22.5%	−24.5%	−24.8%
	1000	0.576	−11.5%	−12.0%	−13.7%	−14.9%	−14.9%
	200	2.411	−37.1%	−39.9%	−39.2%	−41.0%	−40.7%
Pharmacy	500	1.612	−35.1%	−38.1%	−39.1%	−40.7%	−40.2%
	1000	1.137	−33.8%	−36.5%	−37.0%	−38.5%	−37.9%
	200	1.009	−32.8%	−34.3%	−35.8%	−38.0%	−38.1%
Prepared_foods	500	0.665	−30.3%	−31.8%	−35.2%	−37.5%	−37.8%
	1000	0.459	−27.6%	−29.0%	−33.0%	−33.0%	−35.6%
	200	2.629	−16.7%	−35.5%	−27.1%	−36.5%	−37.0%
Produce	500	1.655	−16.8%	−31.5%	−27.6%	−33.5%	−35.1%
	1000	1.196	−15.7%	−23.3%	−22.6%	−23.0%	−26.1%
	200	0.919	−32.8%	−36.2%	−35.4%	−37.0%	−37.3%
Seafood	500	0.582	−28.9%	−29.9%	−30.8%	−31.3%	−31.8%
	1000	0.406	−19.6%	−18.0%	−20.0%	−18.5%	−19.7%

population mean estimates, by applying our model one can produce estimates that have the same level of error with only half the cost[3]).

Comparing the last five columns, we can see that the full version of our proposed model (Primary III) also outperforms all the other model-based alternatives, i.e., Secondary I and II and Primary I and II. In terms of reducing overall estimation errors (i.e., the top three rows), except for Primary II with a monthly sample of 200 or 500, the improvements of Primary III over the other alternatives are all statistically significant ($p < .05$). Moreover, by comparing Primary I with Secondary I, and Primary II with Secondary II, we can clearly see the advantage of the primary analysis over the secondary analysis, holding the basic model formulation the same. Finally, comparing Primary III with Primary II, we can see they perform equally well when the monthly sample size is relatively small (200 or 500); however, Primary III significantly outperforms Primary II when monthly sample size reaches 1000, indicating the advantage of imposing a dynamic factor structure on the latent space becomes more salient as the sample size increases.

Table 3 compares the different alternatives on their ability to estimate month-to-month shifts in population means. This is a more stringent task because month-to-month shifts will be affected by random sampling errors from two consecutive months, doubling the amount of noise. Again, we see that the full version of our proposed Primary DFA (Primary III) performs considerably better than the simple sample averages, reducing MADs in the estimated population mean shifts (across 15 categories, 36 months, and 100 replications) by 38.0% when monthly sample size is 200, 33.2% when monthly sample size is 500, and 26.5% when monthly sample size is 1000 (see the last column and top three rows of Table 3). Similarly, comparing the last five columns of Table 3, we see that the full version of our proposed model outperforms all the other model-based alternatives. In terms of reducing overall estimation errors (i.e., the top three rows), except for Primary II with a monthly sample of 200, all the improvements of Primary III are statistically significant ($p < .05$).

[3] $.904 * (1–27.9\%) = .652 \approx .649$.

5. Tracking customer perceptions and satisfaction with the primary dynamic factor analysis

The previous application served as a realistic testing ground because we had access to the entire population of the focal retailer's loyalty program in the selected region and therefore, were able to use the true population means as the evaluation criteria. The main goals of the first application were in 1) validating our proposed model for conducting the primary analysis with repeated cross-sectional data and 2) comparing its statistical performance with simple sample averages and other model-based alternatives.

The purpose of the application below is different. Here we want to illustrate how a marketing researcher could apply the proposed model to make better use of repeated cross-sectional data currently gathered to track customer perceptions and satisfaction. This illustration typifies repeated surveys conducted by most service-oriented enterprises (hotel chains, banks, retail chains, etc.) in monitoring their service quality and customer satisfaction. In particular, we use monthly tracking data gathered by a distributor of a brand of luxury automobiles in an emerging market. The survey is designed to measure customer perceptions and satisfaction with the distributor's repair and maintenance services. Tables 4a and 4b present summary statistics for the data, with monthly samples of about 100 customers who used the distributor's maintenance and repair services in the month of survey. Because they consider each monthly sample in isolation, these summary statistics do not show any distinct temporal pattern, and are not very informative about trends in customer perceptions and satisfaction. Monthly changes in the sample averages are volatile, but vary within a range due to regression to the mean, something that is typical with customer satisfaction trackers. Moreover, the sample sizes displayed in Table 4b show large variations across the 13 survey items, because respondents were not forced to answer all the questions in the survey.

Table 5 displays within-sample between-item correlations (averaged across the 24 monthly samples), where one can see substantial multicollinearity, potentially due to halo effects (i.e., (dis)-satisfied customers tend to rate all items accordingly). These strong cross-sectional covariations must be isolated from longitudinal co-movements in customer perceptions and satisfaction, highlighting the challenge of disentangling common-method biases (manifested cross-sectionally) from genuine changes over time. Obviously, just tracking sample averages or secondary analyses ignores this problem, running the risk of “throwing the baby out with the bath water” by ignoring all the higher moments of each cross section, resulting in more volatile estimates of population trends, as shown later.

Fig. 3 displays the standardized (to the unit circle) factor loadings obtained from the Secondary and Primary DFA respectively, showing the main direction for each observed variable in the latent state space. The factor structure produced by the Secondary DFA (top panel of Fig. 3) suggests clearly unidimensional service perceptions/satisfaction, with the 13 indicators narrowly pointing in the E/NE direction. The factor structure produced by the Primary DFA (bottom panel of Fig. 3) shows a more nuanced picture, with the 13 indicators spanning the entire top-right quadrant. The unidimensionality of the Secondary DFA suggests a strong halo effect in the sample averages. The halo effect is still present in the Primary DFA (due to inherent common-method biases in customer surveys), but the within-sample co-variance helps further distinguish the information contained in the 13 indicators. As in any other factor model, factor interpretation is highly subjective (and somewhat arbitrary, because of rotation invariance), but one can see that the second (vertical) factor is related to final outcomes, while the first (horizontal) factor is more closely related to specific aspects of the service experience. Most importantly, these latent factors serve as an orthogonal basis for producing more reliable and stable population estimates of customer perceptions and satisfaction over time, as we will show next.

Table 6 shows the respective error-covariance matrices (Σ) of the Secondary and Primary DFA, which provide a cue for the differences in uncovered factor structure between the aggregate and respondent-level approaches. Because it ignores all the cross-sectional co-variances observed within each sample, the Secondary DFA cannot capture any respondent-level information, and tends to over fit the data, because of the much smaller degrees of freedom due to the aggregation within each period. For this reason, the estimated error variances are very small, and all observed (longitudinal) covariances must be captured by the latent factors. In contrast, our proposed model takes into account all the information available within each monthly sample, capturing the cross-sectional covariance through the error-covariance matrix (Σ), and leaving only the longitudinal covariances to be captured by the latent factors. This covariance decomposition leads to a more meaningful latent state-space representation and less volatile estimates of the population state at each point in time. This distinction is made clear in Fig. 4, which displays the monthly scores representing the state of the population at each point in time on the two-dimensional latent space, according to the Secondary DFA and Primary DFA respectively. The top panel of Fig. 4 (Secondary DFA) shows a highly volatile path, ending at a state that is quite close to the initial one. In contrast, the bottom panel (Primary DFA) shows a more defined path, starting at the (undesirable, low-satisfaction) bottom-left quadrant, moving upwards in the direction of “would recommend” and ending in the first (top-right) quadrant in the direction of “satisfied,” suggesting an improvement in perceived service quality and customer satisfaction from the first to the 24th month in our data.

While the latent state space in Fig. 3 and the longitudinal paths shown in Fig. 4 might be meaningful to the knowledgeable analyst and manager, their main purpose is to produce more reliable estimates of perceived service quality and satisfaction for the population of customers over time. We show in Fig. 5 the estimates produced by the Secondary DFA and Primary DFA for two important indicators (“would recommend the dealer for repairs and maintenance,” and “satisfaction with the service department”) for the 24 months in our sample. Fig. 5 shows considerable volatility in the simple sample averages, which confounds random sampling errors with genuine shifts in customer satisfaction. It is clear from this same figure that the Secondary DFA over-fits the aggregate data, producing estimates that mimic the sample averages. In contrast, by leveraging on information available across respondents within each sample, over time and, most importantly, across variables within and between samples, the Primary DFA produces more stable population estimates, providing more reliable/meaningful insights about the evolution of customer perceptions and satisfaction over time.

In summary, the results shown in Figs. 4 and 5 suggest that with our proposed approach, researchers and managers can better detect trends in customer satisfaction that might otherwise be hidden behind volatile sample averages. In addition to looking at trends for individual survey items as shown in Fig. 5, researchers and managers might also want to track the general state of their business in terms of customer perceptions and satisfaction, as shown in the bottom of Figs. 3 and 4.

6. Concluding remarks

With the growing emphasis on business analytics, managers are relying more heavily on data for their decision making. While there has been considerable growth on data about what consumers *do* (e.g., scanner panels, customer databases, and search and clickstream data), information remains limited on how consumers *think* and *feel*; as more activities compete for their time, and they are exposed to more survey requests, consumers are becoming less likely to participate in surveys. Consequently, response rates continue to decline and survey costs continue to grow, forcing researchers to make the most effective use of the data on hand. Despite the widespread use of survey-based trackers by the marketing research industry, the marketing literature is surprisingly scarce on methods that can make more effective use of

Table 4a
Summary statistics for the customer perceptions and satisfaction tracking survey.

Month	Sample averages												
	Would you recommend this dealer for service and repair work	Satisfied with the service department of the dealer	Technical competence of the staff	Friendliness/ helpfulness of staff	Ability to understand my individual problems	Information about any unscheduled work	Explanation of the work carried out on your car	Explanation of the bill	Getting a service appointment within reasonable time	Ability to keep an agreed schedule	Quality of the work performed	Cleanliness of the car when it was returned	Reliability of cost estimates
1	4.5	4.4	4.5	4.7	4.5	4.3	4.4	4.3	4.6	4.5	4.4	4.4	4.5
2	4.3	4.3	4.4	4.6	4.4	4.3	4.3	4.4	4.3	4.3	4.3	4.3	4.3
3	4.2	4.0	4.4	4.7	4.4	4.1	4.2	4.1	4.2	4.2	4.3	4.2	4.2
4	4.5	4.3	4.5	4.7	4.6	4.3	4.4	4.2	4.4	4.4	4.4	4.2	4.4
5	4.5	4.3	4.5	4.7	4.6	4.3	4.4	4.1	4.4	4.4	4.5	4.4	4.3
6	4.4	4.4	4.5	4.6	4.5	4.2	4.5	4.3	4.5	4.6	4.5	4.5	4.5
7	4.7	4.5	4.7	4.8	4.6	4.5	4.6	4.5	4.5	4.7	4.6	4.5	4.7
8	4.5	4.3	4.5	4.7	4.4	4.4	4.5	4.3	4.4	4.4	4.3	4.4	4.4
9	4.5	4.4	4.6	4.8	4.7	4.2	4.6	4.4	4.7	4.6	4.6	4.6	4.5
10	4.7	4.4	4.6	4.7	4.6	4.4	4.6	4.4	4.6	4.5	4.6	4.5	4.5
11	4.3	4.3	4.4	4.6	4.5	4.2	4.3	4.2	4.3	4.5	4.5	4.5	4.5
12	4.5	4.4	4.5	4.7	4.5	4.5	4.4	4.2	4.6	4.5	4.5	4.7	4.4
13	4.7	4.6	4.7	4.8	4.8	4.7	4.7	4.6	4.6	4.6	4.7	4.6	4.5
14	4.4	4.4	4.6	4.7	4.5	4.3	4.4	4.3	4.5	4.4	4.4	4.4	4.4
15	4.7	4.6	4.7	4.7	4.7	4.6	4.6	4.5	4.5	4.6	4.7	4.7	4.6
16	4.5	4.3	4.5	4.7	4.4	4.3	4.4	4.3	4.3	4.3	4.3	4.5	4.2
17	4.7	4.5	4.7	4.8	4.6	4.4	4.5	4.4	4.6	4.6	4.6	4.6	4.3
18	4.5	4.3	4.5	4.7	4.3	4.3	4.4	4.3	4.3	4.4	4.3	4.6	4.3
19	4.5	4.3	4.4	4.6	4.4	4.3	4.4	4.3	4.3	4.4	4.4	4.6	4.2
20	4.6	4.3	4.4	4.6	4.6	4.4	4.5	4.2	4.4	4.5	4.5	4.6	4.4
21	4.6	4.3	4.5	4.7	4.5	4.3	4.4	4.3	4.4	4.3	4.4	4.5	4.3
22	4.5	4.4	4.6	4.7	4.6	4.4	4.4	4.5	4.6	4.5	4.5	4.6	4.5
23	4.6	4.4	4.6	4.8	4.6	4.5	4.5	4.5	4.5	4.6	4.4	4.7	4.5
24	4.5	4.3	4.5	4.7	4.6	4.4	4.4	4.4	4.5	4.4	4.5	4.4	4.4

Table 4a (*continued*)

Month	Standard error of the mean												
	Would you recommend this dealer for service and repair work	Satisfied with the service department of the dealer	Technical competence of the staff	Friendliness/ helpfulness of staff	Ability to understand my individual problems	Information about any unscheduled work	Explanation of the work carried out on your car	Explanation of the bill	Getting a service appointment within reasonable time	Ability to keep an agreed schedule	Quality of the work performed	Cleanliness of the car when it was returned	Reliability of cost estimates
1	0.10	0.10	0.09	0.06	0.09	0.12	0.10	0.11	0.09	0.09	0.09	0.09	0.09
2	0.11	0.09	0.08	0.08	0.08	0.12	0.09	0.09	0.10	0.09	0.10	0.11	0.09
3	0.12	0.11	0.09	0.06	0.10	0.13	0.10	0.11	0.11	0.12	0.11	0.13	0.11
4	0.10	0.09	0.08	0.07	0.08	0.12	0.09	0.11	0.09	0.09	0.10	0.12	0.10
5	0.10	0.09	0.07	0.05	0.07	0.11	0.09	0.12	0.09	0.10	0.09	0.10	0.11
6	0.10	0.08	0.07	0.07	0.08	0.12	0.09	0.10	0.08	0.07	0.09	0.09	0.09
7	0.06	0.08	0.06	0.05	0.07	0.10	0.09	0.09	0.07	0.05	0.07	0.10	0.06
8	0.10	0.10	0.08	0.06	0.08	0.09	0.08	0.10	0.09	0.09	0.11	0.10	0.10
9	0.09	0.10	0.08	0.04	0.06	0.13	0.08	0.10	0.06	0.07	0.08	0.08	0.08
10	0.07	0.09	0.08	0.07	0.07	0.11	0.08	0.11	0.08	0.09	0.08	0.09	0.09
11	0.13	0.11	0.10	0.08	0.09	0.12	0.10	0.11	0.11	0.11	0.10	0.10	0.09
12	0.11	0.10	0.09	0.06	0.09	0.11	0.09	0.11	0.07	0.08	0.08	0.06	0.10
13	0.08	0.08	0.06	0.05	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.08	0.08
14	0.11	0.09	0.08	0.06	0.08	0.13	0.10	0.09	0.09	0.10	0.10	0.10	0.10
15	0.08	0.09	0.07	0.08	0.08	0.09	0.09	0.10	0.09	0.08	0.07	0.06	0.09
16	0.11	0.09	0.09	0.06	0.09	0.10	0.09	0.10	0.09	0.09	0.10	0.08	0.11
17	0.08	0.08	0.08	0.05	0.08	0.10	0.08	0.10	0.08	0.08	0.07	0.09	0.11
18	0.09	0.10	0.08	0.06	0.10	0.10	0.10	0.09	0.08	0.09	0.10	0.08	0.09
19	0.08	0.09	0.10	0.08	0.10	0.10	0.09	0.10	0.10	0.10	0.08	0.08	0.10
20	0.09	0.09	0.08	0.08	0.07	0.11	0.08	0.10	0.09	0.08	0.08	0.07	0.09
21	0.07	0.09	0.09	0.07	0.09	0.11	0.09	0.10	0.09	0.10	0.09	0.10	0.10
22	0.08	0.09	0.07	0.06	0.07	0.10	0.09	0.07	0.07	0.08	0.09	0.09	0.07
23	0.08	0.09	0.08	0.06	0.08	0.08	0.09	0.08	0.08	0.07	0.10	0.06	0.10
24	0.09	0.09	0.08	0.06	0.07	0.11	0.10	0.11	0.08	0.09	0.09	0.10	0.10

Table 4b
Sample sizes for the customer perceptions and satisfaction tracking survey.

Month	Would you recommend this dealer for service and repair work	Satisfied with the service department of the dealer	Technical competence of the staff	Friendliness/ helpfulness of staff	Ability to understand my individual problems	Information about any unscheduled work	Explanation of the work carried out on your car	Explanation of the bill	Getting a service appointment within reasonable time	Ability to keep an agreed schedule	Quality of the work performed	Cleanliness of the car when it was returned	Reliability of cost estimates
1	95	100	96	100	100	79	97	97	98	99	99	98	95
2	93	100	99	100	100	76	98	97	100	100	100	97	96
3	95	100	100	100	100	79	100	95	98	100	98	96	97
4	98	100	99	100	100	70	94	92	98	97	100	93	94
5	96	100	99	100	99	80	98	92	100	100	98	96	90
6	98	100	100	100	99	82	99	95	100	100	100	96	94
7	95	100	100	100	100	88	97	98	99	99	98	97	95
8	91	100	99	99	99	76	97	97	98	99	99	97	97
9	99	100	100	100	100	85	100	97	100	100	99	99	90
10	98	100	99	100	98	77	99	97	97	100	99	97	95
11	98	100	100	100	100	87	98	99	99	98	99	100	100
12	100	100	100	100	100	73	98	97	100	100	100	99	98
13	96	100	98	98	97	76	98	97	97	98	100	100	97
14	100	100	98	100	96	80	98	95	98	99	98	98	97
15	100	100	98	100	98	85	97	92	100	99	99	99	95
16	100	100	94	100	98	83	97	90	98	96	98	98	96
17	96	100	98	100	96	84	98	94	99	98	98	99	96
18	98	100	98	100	98	77	97	97	99	98	100	99	98
19	96	100	99	100	98	76	96	94	98	99	99	97	96
20	100	100	100	100	100	74	99	93	93	100	100	99	92
21	100	100	95	100	99	87	97	95	98	100	99	99	98
22	100	100	99	100	99	84	100	94	98	99	99	98	98
23	97	100	98	100	98	85	99	95	95	98	100	97	98
24	100	100	100	100	99	90	100	96	97	99	98	98	99

Table 5
Average cross-sectional (within month) between-item correlations.

	Would recommend	Satisfaction	Tech competence	Friendliness	Understand your needs	Information about new work	Explained work done	Explained the bill	Getting appointment when needed	Keeps agreed schedule	Quality of work done	Cleanliness of the car	Reliable cost estimates
Would recommend	1.00	0.70	0.61	0.49	0.59	0.58	0.61	0.45	0.58	0.54	0.67	0.46	0.60
Satisfaction	0.70	1.00	0.68	0.49	0.64	0.68	0.65	0.55	0.65	0.57	0.74	0.48	0.63
Tech competence	0.61	0.68	1.00	0.56	0.66	0.65	0.62	0.53	0.57	0.55	0.64	0.41	0.63
Friendliness	0.49	0.49	0.56	1.00	0.58	0.66	0.59	0.49	0.51	0.45	0.46	0.43	0.48
Understand your needs	0.59	0.64	0.66	0.58	1.00	0.67	0.71	0.52	0.58	0.51	0.58	0.37	0.60
Information about new work	0.58	0.68	0.65	0.66	0.67	1.00	0.74	0.56	0.59	0.55	0.60	0.53	0.63
Explained work done	0.61	0.65	0.62	0.59	0.71	0.74	1.00	0.60	0.58	0.56	0.65	0.47	0.62
Explained the bill	0.45	0.55	0.53	0.49	0.52	0.56	0.60	1.00	0.50	0.51	0.56	0.40	0.64
Getting appointment when needed	0.58	0.65	0.57	0.51	0.58	0.59	0.58	0.50	1.00	0.68	0.60	0.41	0.51
Keeps agreed schedule	0.54	0.57	0.55	0.45	0.51	0.55	0.56	0.51	0.68	1.00	0.59	0.44	0.49
Quality of work done	0.67	0.74	0.64	0.46	0.58	0.60	0.65	0.56	0.60	0.59	1.00	0.49	0.63
Cleanliness of the car	0.46	0.48	0.41	0.43	0.37	0.53	0.47	0.40	0.41	0.44	0.49	1.00	0.44
Reliable cost estimates	0.60	0.63	0.63	0.48	0.60	0.63	0.62	0.64	0.51	0.49	0.63	0.44	1.00

repeated cross-sectional data, which could either improve accuracy or reduce data-gathering costs.

The state-space model we proposed and tested in this study is an attempt to help researchers to better track population means based on repeated cross-sectional surveys, taking full advantage of the respondent-level information. Our proposed model simultaneously taps into three features that are commonly present in tracking data. First, it treats data from independent random samples from different periods as inter-dependent, reflecting the fact that population means are temporally dependent and thus information can be borrowed from other time periods in inferring the population means at any given point in time. Second, our model is multivariate and imposes a parsimonious factor structure to take advantage of the fact that most tracking studies include multiple correlated measures, thereby borrowing information from all the other measures in inferring the population means of any given measure. Third, our model is applied to respondent-level data, thus controlling for cross-sectional co-variations across variables, which helps distinguish temporal co-movements in population means from the cross-sectional covariances between individual responses.

One way to see our methodological contribution is to think of our proposed model as a fusion between the Secondary DFA model proposed by Du and Kamakura (2012) and the primary model proposed by Lind (2005). The Secondary DFA model can leverage the first and second features mentioned above, while Lind's model can leverage the first and third features. However, neither Du and Kamakura (2012) nor Lind (2005) tested their model's statistical performances in recovering population means from repeated cross-sectional data. The first of our two

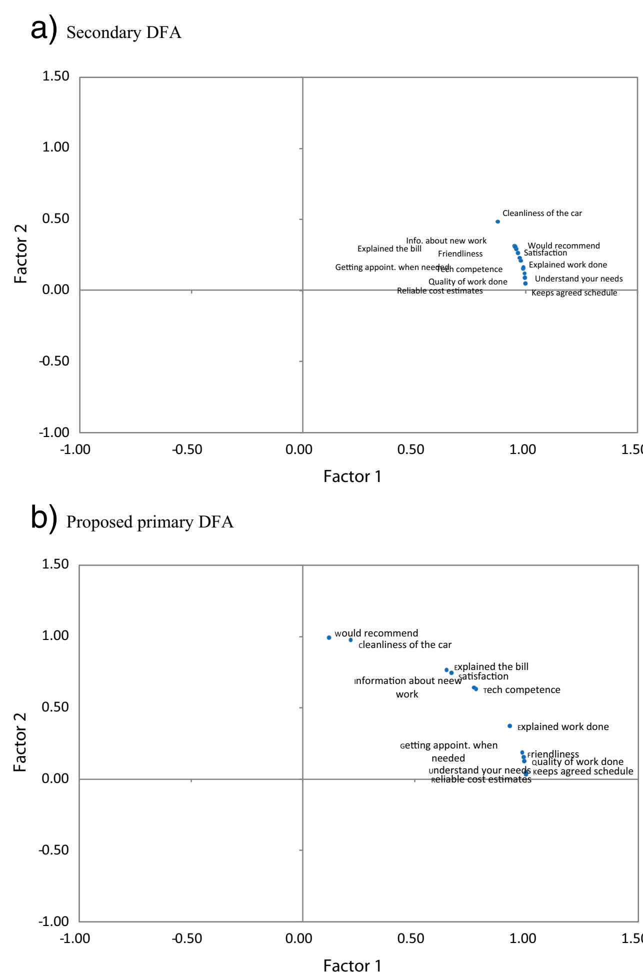


Fig. 3. Factor loadings for the Secondary DFA and the proposed Primary DFA.

Table 6
Estimated error-covariance matrices (Σ) for the Secondary DFA and proposed model (Primary DFA).

	Would recommend	Satisfaction	Tech competence	Friendliness	Understand your needs	Information about new work	Explained work done	Explained the bill	Getting appointment when needed	Keeps agreed schedule	Quality of work done	Cleanliness of the car	Reliable cost estimates
a) Secondary DFA													
Would recommend	0.00383	0	0	0	0	0	0	0	0	0	0	0	0
Satisfaction	0	0.00080	0	0	0	0	0	0	0	0	0	0	0
Tech competence	0	0	0.00098	0	0	0	0	0	0	0	0	0	0
Friendliness	0	0	0	0.00224	0	0	0	0	0	0	0	0	0
Understand your needs	0	0	0	0	0.00005	0	0	0	0	0	0	0	0
Information about new work	0	0	0	0	0	0.00284	0	0	0	0	0	0	0
Explained work done	0	0	0	0	0	0	0.00069	0	0	0	0	0	0
Explained the bill	0	0	0	0	0	0	0	0.00518	0	0	0	0	0
Getting appointment when needed	0	0	0	0	0	0	0	0	0.00293	0	0	0	0
Keeps agreed schedule	0	0	0	0	0	0	0	0	0	0.00294	0	0	0
Quality of work done	0	0	0	0	0	0	0	0	0	0	0.00162	0	0
Cleanliness of the car	0	0	0	0	0	0	0	0	0	0	0	0.00616	0
Reliable cost estimates	0	0	0	0	0	0	0	0	0	0	0	0	0.00553
b) Proposed model (Primary DFA)													
Would recommend	0.84	0.68	0.59	0.48	0.55	0.56	0.56	0.44	0.52	0.50	0.63	0.41	0.55
Satisfaction	0.57	0.84	0.71	0.54	0.65	0.67	0.64	0.51	0.60	0.57	0.72	0.43	0.59
Tech competence	0.45	0.54	0.67	0.61	0.66	0.64	0.63	0.51	0.54	0.52	0.64	0.41	0.56
Friendliness	0.30	0.34	0.34	0.42	0.61	0.57	0.59	0.49	0.50	0.47	0.48	0.39	0.51
Understand your needs	0.40	0.47	0.44	0.33	0.64	0.64	0.69	0.52	0.55	0.53	0.61	0.38	0.58
Information about new work	0.48	0.57	0.50	0.35	0.47	0.86	0.72	0.56	0.54	0.50	0.60	0.44	0.59
Explained work done	0.46	0.52	0.47	0.35	0.48	0.60	0.78	0.63	0.52	0.51	0.63	0.44	0.64
Explained the bill	0.39	0.45	0.40	0.31	0.39	0.49	0.54	0.94	0.49	0.47	0.50	0.40	0.64
Getting appointment when needed	0.41	0.47	0.39	0.29	0.38	0.43	0.39	0.40	0.73	0.67	0.54	0.37	0.50
Keeps agreed schedule	0.41	0.46	0.38	0.28	0.37	0.42	0.39	0.39	0.50	0.77	0.57	0.40	0.47
Quality of work done	0.52	0.60	0.48	0.29	0.44	0.51	0.50	0.43	0.42	0.45	0.82	0.46	0.57
Cleanliness of the car	0.32	0.34	0.29	0.23	0.26	0.35	0.33	0.32	0.27	0.30	0.35	0.80	0.42
Reliable cost estimates	0.46	0.49	0.41	0.31	0.42	0.48	0.51	0.56	0.39	0.37	0.46	0.34	0.86

Note: The upper diagonal cells contain correlations, while all the other cells contain variance–covariances.

applications is designed to fill in this gap, where we know the true population means over time and are able to draw independent random samples from the population in each period, establishing a realistic testing ground for performance evaluation and model comparison.

While useful for model validation (because we have access to the entire population), using shopping basket data from a retailer's loyalty program may have been too conservative in highlighting one of our model's most valuable features: its ability to isolate longitudinal trends from sampling errors in the repeated cross-sections. This happened because household expenditures on the various product categories sold by a supermarket tend to be seasonal and directly affected by promotions by the retailer and its competition, which were not available to us.

Nevertheless, we were able to clearly demonstrate via a Monte Carlo experiment (15 variables, 36 periods and 100 replications across three different sample sizes) that our proposed model can improve the accuracy of repeated cross-sectional tracking studies by double digits, without incurring any additional data-gathering costs (or equivalently, reducing the data-gathering costs by double digits while maintaining the desired accuracy level). These gains are larger when sample sizes are smaller and when compared with using simple sample averages or the secondary analysis. To the best of our knowledge, this is the first study that has shown empirical evidence based on real-life data from a large population, establishing the superior statistical performance of the primary analysis (previous studies focused only on computational efficiency). Even when compared with other fairly-sophisticated alternatives for the primary analysis, the performance of our proposed model has proven to be either on par or superior.

Our second application, using monthly customer satisfaction survey data gathered by a luxury automobile distributor, exemplifies how our proposed model can be useful in uncovering the underlying population trends in each survey indicator, along with the latent common trends hidden behind multiple indicators. The estimates for each indicator, obtained as $\hat{\mu}_t = \hat{\Lambda} z_{t|T} + \hat{B} x_t$, isolate the underlying longitudinal trends from random sampling errors, providing the manager with a more interpretable tracking indicator of the shifts in customer perceptions and satisfaction. The latent scores $z_{t|T}$ can be interpreted (via the loading matrix Λ) as the location of the firm in the state space of customer perceptions/satisfaction at month t.

While our proposed model applies to continuous (interval or ratio scaled) measures, it can be extended to discrete measures. For that, we suggest the formulation proposed by Tanizaki (1993), which we leave for future research. Finally, another advantage of our state-space model for the primary analysis lies in its capability in handling missing individual responses. With the secondary analysis, missing individual responses must be ignored or imputed when calculating sample averages in the aggregation step. By contrast, with our proposed model, missing individual responses can be readily bypassed in the E step (Eq. (A4)).

Appendix A. Expectation-Maximization (EM) estimator for the proposed model

A.1. Kalman filter (forward pass in the E-step)

Step 1 — initialize $z_{0|0}$ and $\mathrm{var}(z_{0|0})$, i.e., prior of the initial state at period 0

$z_{0|0} = \mu_0$, and $\mathrm{var}(z_{0|0}) = \Omega_0$

Repeat Steps 2 and 3 for t = 1, …, T

Step 2 — calculate $z_{t|t-1}$ and $\mathrm{var}(z_{t|t-1})$ i.e., expectation and uncertainty about the state in t given data observed up until t − 1

$$z_{t|t-1} = z_{t-1|t-1} \tag{A1}$$

$$\mathrm{var}\left(z_{t|t-1}\right) = \mathrm{var}\left(z_{t-1|t-1}\right) + \Omega \tag{A2}$$

Step 3 — calculate $z_{t|t}$ and $\mathrm{var}(z_{t|t})$ i.e., expectation and uncertainty about the state in t given data observed up until t

$$\mathrm{var}\left(z_{t|t}\right) = \left[\mathrm{var}\left(z_{t|t-1}\right)^{-1} + \Lambda' N_t \Sigma^{-1} \Lambda\right]^{-1} \tag{A3}$$

$$z_{t|t} = z_{t|t-1} + \mathrm{var}\left(z_{t|t}\right)\Lambda' N_t \Sigma^{-1}\left[\bar{y}_t - \Lambda z_{t|t-1} - B x_t\right], \text{ where } \bar{y}_t \equiv \frac{1}{N_t}\sum_{i(t)=1}^{N_t} y_{i(t)}. \tag{A4}$$

A.2. Kalman smoother (backward pass in the E-step)

Repeat Step 4 for t = T, T − 1, …, 1

Step 4 — calculate $z_{t-1|T}$ and $\mathrm{var}(z_{t-1|T})$, i.e., expectation and uncertainty about the state in period t − 1, given all the data available through T

$$L_{t-1} = \mathrm{var}\left(z_{t-1|t-1}\right)\left[\mathrm{var}\left(z_{t|t-1}\right)\right]^{-1} \tag{A5}$$

$$z_{t-1|T} = z_{t-1|t-1} + L_{t-1}\left[z_{t|T} - z_{t|t-1}\right] \tag{A6}$$

$$\mathrm{var}\left(z_{t-1|T}\right) = \mathrm{var}\left(z_{t-1|t-1}\right) + L_{t-1}\left[\mathrm{var}\left(z_{t|T}\right) - \mathrm{var}\left(z_{t|t-1}\right)\right]L'_{t-1} \tag{A7}$$

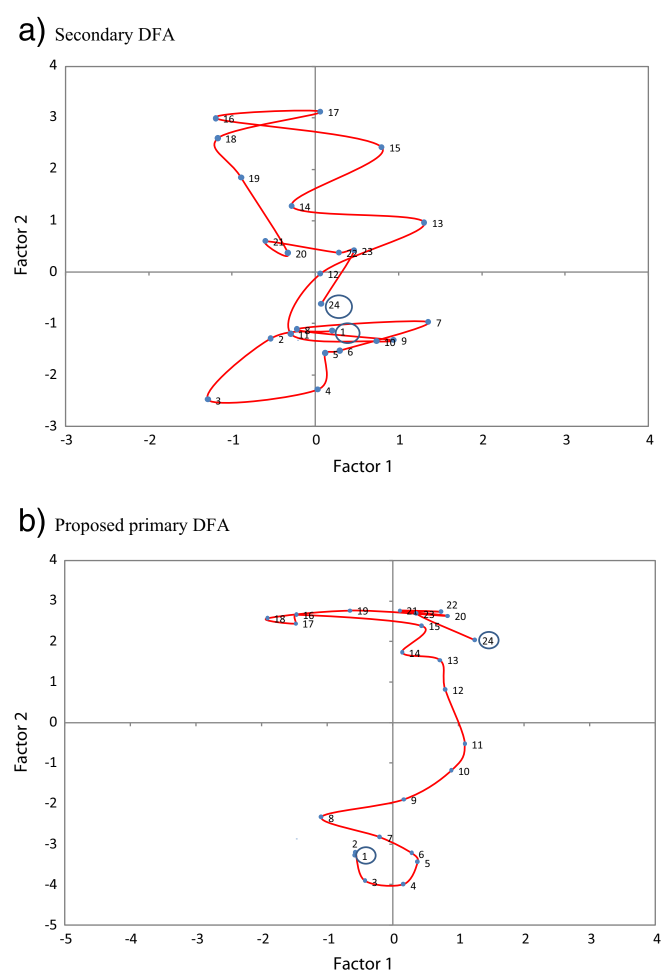


Fig. 4. Population paths, according to the Secondary DFA and the proposed Primary DFA.

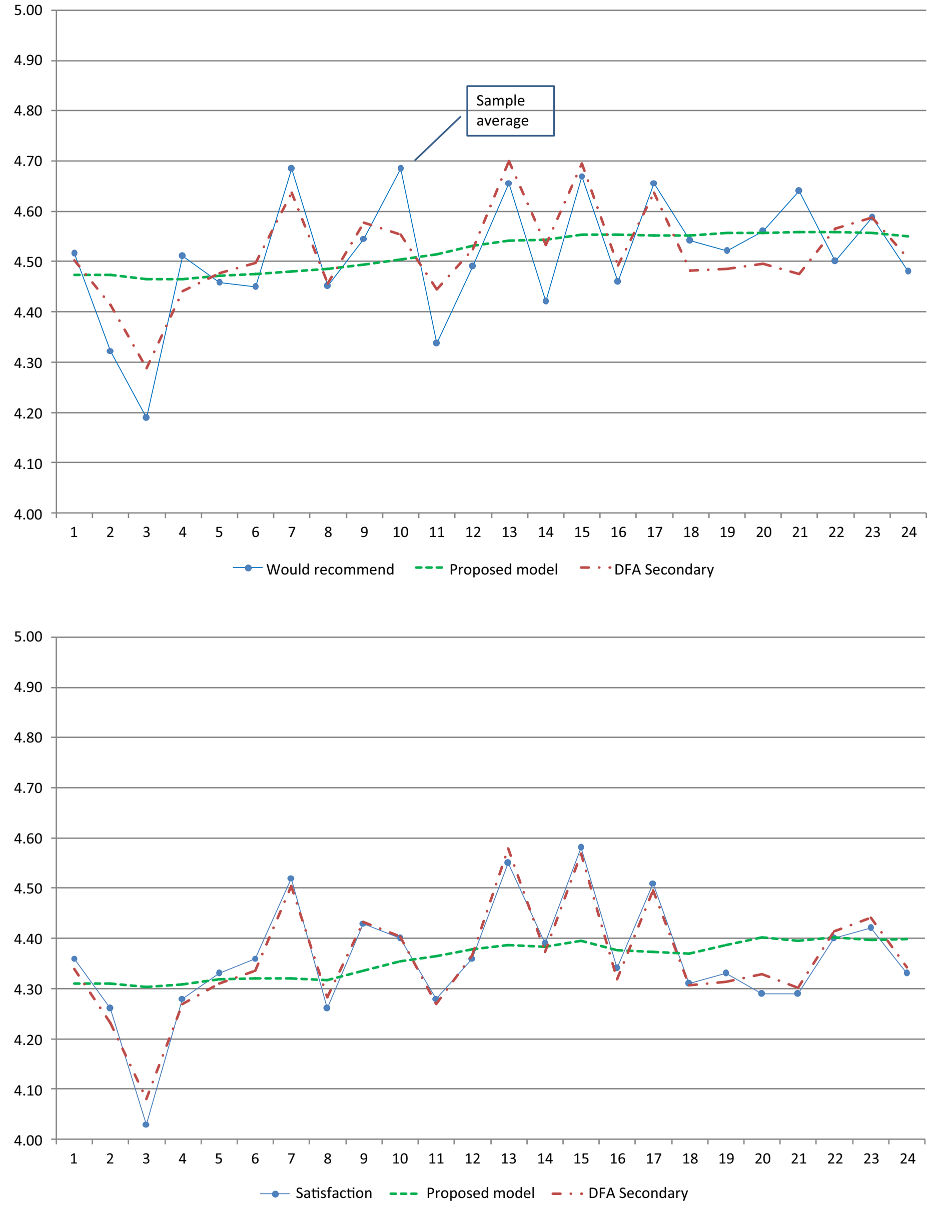


Fig. 5. Population estimates for two indicators of customer satisfaction.

Step 5 — initialize $\mathrm{cov}(z_{T|T}, z_{T-1|T})$, i.e., covariance of uncertainties about states at T and T − 1

$$\mathrm{cov}\left(z_{T|T}, z_{T-1|T}\right) = \left[I - \mathrm{var}\left(z_{T|T}\right)\Lambda' N_t \Sigma^{-1}\Lambda\right]\mathrm{var}\left(z_{T-1|T-1}\right) \quad \text{(A8)}$$

Repeat Step 6 for $t = T, T-1, \ldots, 2$

Step 6 — calculate $\mathrm{cov}(z_{t-1|T}, z_{t-2|T})$, i.e., covariance of uncertainties about states at t − 1 and t − 2, given all the data available through T

$$\mathrm{cov}\left(z_{t-1|T}, z_{t-2|T}\right) = \mathrm{var}\left(z_{t-1|t-1}\right)L'_{t-2} + L_{t-1}\left[\mathrm{cov}\left(z_{t|T}, z_{t-1|T}\right) - \mathrm{var}\left(z_{t-1|t-1}\right)\right]L'_{t-2}. \quad \text{(A9)}$$

For identification purposes, $z_{t|T}$ is mean-centered at the end of each E-step.

A.3. M-step

The maximization-step is where estimates of the hyper-parameters of the model are obtained, taking the estimates of the latent states as given. Let Θ^j collect all the estimates of the hyper-parameters at the *j*-th iteration of the Expectation–Maximization process ($\Theta^j = \{\Lambda, B, \Sigma, \Omega, \mu_0, \text{and } \Omega_0\}$). Then, to compute the expected likelihood conditional on the observed data (Y and X) and hyper-parameters from iteration $j-1$ (Θ^{j-1}), the following statistics obtained from the E-step are sufficient: $z_{t|T}$, $var(z_{t|T})$, and $cov(z_{t|T}, z_{t-1|T})$. The likelihood function to be maximized in iteration j of the EM algorithm is given by:

$$\begin{aligned} E_{Z|\Theta^{j-1}}\left\{-2\ln L\left(Y,X,Z;\Theta^{j}\right)\right\} &\propto \ln|\Omega_0| + \mathrm{tr}\left\{\Omega_0^{-1}\left[\left(z_{0|T}-a_0\right)\left(z_{0|T}-a_0\right)' + var\left(z_{0|T}\right)\right]\right\} \\ &+ T\ln|\Sigma_\Omega| + \mathrm{tr}\left\{\sum\nolimits_\Omega^{-1}\sum_{t=1}^{T}\left[\begin{array}{c}\left(z_{t|T}-z_{t-1|T}\right)\left(z_{t|T}-z_{t-1|T}\right)' + var\left(z_{t|T}\right) + var\left(z_{t-1|T}\right) \\ -cov\left(z_{t|T}, z_{t-1|T}\right) - cov\left(z_{t|T}, z_{t-1|T}\right)'\end{array}\right]\right\} \\ &+ \sum_{t=1}^{T} N_t \ln|\Sigma| \\ &+ \mathrm{tr}\left\{\Sigma^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i(t)=1}^{N_t}\left[\left(y_{i(t)}-\Lambda z_{t|T}-Bx_t\right)\left(y_{i(t)}-\Lambda z_{t|T}-Bx_t\right)' + \Lambda var\left(z_{t|T}\right)\Lambda'\right]\right\}. \end{aligned} \tag{A10}$$

Given the expected likelihood function above, Θ^j can be estimated by solving the first-order conditions, which leads to the following estimates:

$$\hat{a}_0 = z_{0|T} \tag{A11}$$

$$\hat{\Omega}_0 = \mathrm{diag}\left[var\left(z_{0|T}\right)\right] \tag{A12}$$

$$\left[\hat{\Lambda}\quad \hat{B}\right] = \left[\sum\nolimits_{t=1}^{T}\sum\nolimits_{i(t)=1}^{N_t} y_{i(t)}\left(z_{t|T}\right)' \quad y_{i(t)}x_t'\right]\left[\sum\nolimits_{t=1}^{T} N_t\left[\begin{array}{cc} z_{t|T}\left(z_{t|T}\right)' + var\left(z_{t|T}\right) & z_{t|T}x_t' \\ x_t\left(z_{t|T}\right)' & x_t x_t'\end{array}\right]\right]^{-1} \tag{A13}$$

$$\hat{\Sigma} = \frac{1}{\sum_{t=1}^{T} N_t}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i(t)=1}^{N_t}\left[\left(y_{i(t)}-\hat{\Lambda}z_{t|T}-\hat{B}x_t\right)\left(y_{i(t)}-\hat{\Lambda}z_{t|T}-\hat{B}x_t\right)' + \hat{\Lambda}var\left(z_{t|T}\right)\hat{\Lambda}'\right] \tag{A14}$$

$$\hat{\Omega} = \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathrm{diag}\left[\begin{array}{c}\left(z_{t|T}-z_{t-1|T}\right)\left(z_{t|T}-z_{gt-1|T}\right)' + var\left(z_{t|T}\right) + var\left(z_{t-1|T}\right) \\ -cov\left(z_{t|T}, z_{t-1|T}\right) - cov\left(z_{t|T}, z_{t-1|T}\right)'\end{array}\right]. \tag{A15}$$

References

Du, Rex, & Kamakura, Wagner A. (2012). Quantitative trendspotting. *Journal of Marketing Research*, *49*(4), 514–536.

Feder, Moshe (2001). Time series analysis of repeated surveys: the state-space approach. *Statistica Nederlandica*, *55*(2), 182–199.

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge, UK: Cambridge University Press, 449–456.

Harvey, A.C., & Chung, C.H. (2000). Estimating the underlying change in unemployment in the U.K. *Journal of the Royal Statistical Society, Series A*, *163*, 303–339.

Hsiao, Cheng (2007, March). *Panel data analysis — advantages and challenges*. Sociedad de Estadística e Investigación Operativa.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers D*, *82*, 35–45.

Lind, Jo Thori (2005). Repeated surveys and the Kalman filter. *Econometrics Journal*, *8*, 418–427.

Norris, Bruce I., Peters, Kay, & Naik, Prasad A. (2012). Discovering how advertising grows sales and build brands. *Journal of Marketing Research*, *49*, 793–806.

Pauwels, Koen, & Hanssens, Dominique M. (2007). Performance regimes and marketing policy shifts. *Marketing Science*, *26*(3), 293–311.

Pfefferman, Danny (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, *9*(2), 163–175.

Scott, A.J., & Smith, T.M.F. (1974). Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association*, *69*(347), 674–678.

Scott, A.J., Smith, T.M.F., & Jones, R.G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical ReviewRevue Internationale de Statistique*, *45*(1), 13–28.

Shumway, Robert H., & Stoffer, David S. (2000). *Time series analysis and its applications*. New York: Springer-Verlag.

Srinivasan, Shuba, Vanhuele, Marc, & Pauwels, Koen (2010, August). Mindset metrics in market response models: an integrative approach. *Journal of Marketing Research*, *47*, 672–684.

Tam, S.M. (1987). Analysis of repeated surveys using a dynamic linear model. *International Statistical Review*, *55*(1), 63–73.

Tanizaki, Hisashi (1993). Kalman filters model with qualitative dependent variables. *The Review of Economics and Statistics*, *75*(4), 747–752.

Tourangeau, Roger (2003). *Recurring surveys: issues and opportunities*. National Science Foundation.

Xie, Jinhong, Song, Michael, Sirbu, Marvin, & Wang, Qiong (1997, August). Kalman filter estimation of new product diffusion models. *Journal of Marketing Research*, *34*, 378–393.

Zuur, Alain F., Fryer, R.J., Jolliffe, I.T., Dekker, R., & Beukema, J.J. (2003). Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*, *14*(7), 665–685.