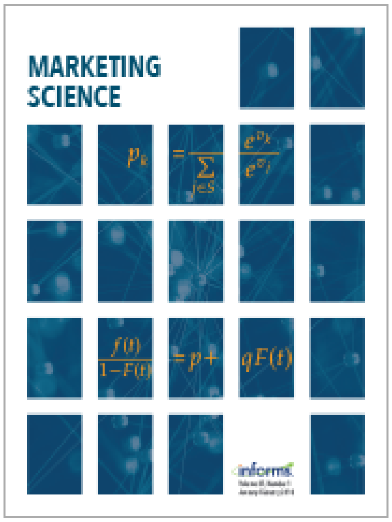

Marketing Science

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

Optimal Data Interval for Estimating Advertising Response

Gerard J. Tellis, Philip Hans Franses,

MARKETING SCIENCE
Vol. 25, No. 3, May–June 2006, pp. 217–229
ISSN 0732-2399 | EISSN 1526-548X | 06 | 2503 | 0217

DOI 10.1287/mksc.1050.0178

Optimal Data Interval for Estimating Advertising Response

Gerard J. Tellis
Marshall School of Business, University of Southern California, Los Angeles, California 40089-0443, tellis@usc.edu

Philip Hans Franses
Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, franses@few.eur.nl

The abundance of highly disaggregate data (e.g., at five-second intervals) raises the question of the optimal data interval to estimate advertising carryover. The literature assumes that (1) the optimal data interval is the interpurchase time, (2) too disaggregate data causes a disaggregation bias, and (3) recovery of true parameters requires assumption of the underlying advertising process. In contrast, we show that (1) the optimal data interval is what we call *unit exposure time*, (2) too disaggregate data does not cause any disaggregation bias, and (3) recovery of true parameters does not require assumption of the advertising process but only data at the unit exposure time. These results hold for any linear dynamic model linking sales with current and past advertising.

Key words: advertising carryover; duration of ad effects; optimal data interval; interpurchase time; interexposure time
History: This paper was received June 14, 2004, and was with the authors 8 months for 2 revisions; processed by Michel Wedel.

1. Introduction

How long do the effects of advertising last? What is the appropriate data interval to estimate these effects? This topic has important implications for marketers who want to optimally schedule advertising, for policy makers who determine whether advertising has any negative long-term effects, and for legal analysts who assess possible negative effects of advertising on consumers.

It is long known in marketing and econometrics that temporally aggregated data biases estimates of the duration of advertising's effect (e.g., Clarke 1976, 1982). It is also known that the bias is due to specification error caused by assuming that the true model at the microdata applies as well to the aggregate data (Bass and Leone 1983, Rao 1986, Weinberg and Weiss 1982). Researchers have proposed several approaches for retrieving the true parameters from aggregate data. Russell (1988) showed that these approaches all concern special cases of assumptions about the underlying advertising process.

We now live in an era of abundant disaggregate data. New technologies now record data at a highly disaggregate level, including by week, day, minute, and even second of advertising exposure and consumer purchase. This abundance raises a new set of questions. Which microdata interval is the most appropriate for econometric analysis: the interpurchase time, the unit exposure time, the week, day, minute, or second? Is there a data disaggregation bias? Moreover, while aggregate data bias the estimated coefficients, disaggregate data (e.g., in seconds or minutes) greatly increases the costs of data storage, processing, and analysis. So, is there an optimal data interval?

In the absence of clear answers to these questions, marketing researchers have resorted to assumptions or speculations. As early as 1976, Clarke (1976) documented the serious bias in estimated advertising carryover from data aggregation: "The weekly results are biased downward because the purchase cycle is probably longer than the data interval" (p. 355). He thus seeded the notion that the purchase interval was the true data interval and that too disaggregate data is also bad. No one since has refuted these suspicions. Rather, one of his suspicions seems to have become part of the axioms in marketing. For example, seven years later, Weiss et al. (1983, p. 279) said, "An emerging convention seems to consider ... [the true microinterval] ... to be the interval between purchases." Finally, in a *Marketing Science* issue on marketing generalizations, Leone reiterated an earlier recommendation of Bass and Leone that "if one should choose the data interval used in the analysis, one should select the interval corresponding to the inter-purchase time for the product category"

(Leone 1995, p. G47). Yet Vanhonacker (1983), among others, pointed out that neither mathematical proof nor any theory supports this premise. The econometrics literature has also not provided any answer about an optimal data interval.

In this paper we prove for econometric estimates of the duration of advertising carryover that (1) the optimal data interval at which the data should be collected is the unit exposure time, because for these data one can estimate the true carryover effects of advertising in the microdata without bias; (2) the optimal data interval is independent of the interpurchase time; and (3) more disaggregate data does not cause any downward bias in estimated duration of advertising's effect, provided that one adapts the model. Importantly, our results do not require knowledge of the advertising process at the microlevel. More microdata, however, leads to unnecessary cost in time, space, and labor in collecting, maintaining, and analyzing such detailed data. We present the assumptions necessary for these conclusions and possible extensions of our argument. These results hold for a variety of models, but for ease of exposition we initially demonstrate our arguments for the commonly used Koyck model. We then extend our argument to other models.

The rest of the paper is organized as follows. Section 2 describes the preliminaries. Section 3 proves our thesis for the case of the Koyck model, with supporting simulations. Section 4 extends our thesis to other models. Section 5 provides an empirical illustration. Section 6 discusses the implications of our thesis.

2. Preliminaries

To avoid ambiguity, this section first defines the terms and premises of the analysis and then clarifies its intuition.

2.1. Definitions

The *data interval* is the temporal level of the records, for example, annual, weekly, daily, by minute. The *interpurchase time* is the smallest calendar time between any two consumer purchases. The *duration interval* is the length of time that some fraction of the effects of advertising lasts. *Calendar time* is the discrete time period such as second, minute, hour, day, week, month, or year. The *exposure time* is the moment a pulse of advertisement first hits a consumer, irrespective of its withdrawal; examples are the 30 seconds for TV ads, the day or hour for print ads, the day for mail ads, or the week for outdoor ads.

The unit exposure time is the largest calendar period in the time frame under study such that advertising exposure occurs *at most* once in that period, and if it occurs, it does so at the same time in that period. Figure 1 illustrates the concept of the unit exposure time, with four hypothetical schedules, each shown for 15 days. Assume, for simplicity, that no exposures occur in the second 15 days of the month and the pattern shown is the same for the remaining months of the year. In Schedule 1, advertising occurs only on the sixth day of the month. In this case, monthly data will have at most one exposure per month, and that at the same day of the month. So, the unit exposure time is the month. Schedule 2 has exposures on the third

Figure 1 **Four Alternate Advertising Schedules**

Schedule	Days of the month															Unit exposure time*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1						↑										Months
2			↑							↑						Weeks
3 At 6:00 P.M.	↑			↑		↑					↑	↑			↑	Days
4	↑		↑ ↑ ↑ ↑		↑				↑	↑						(see below)
	Hours for day 3															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
4 On the hour								↑	↑		↑			↑		Hours

*Corresponding unit exposure time.

and 10th of each month. In this case, weekly data will have at most one exposure per week, and that at the same day of the week. So the unit exposure time is the week. Schedule 3 has exposures on the first, fourth, sixth, 11th, 12th, and 15th of each month at 6:00 P.M. In this case, daily data have at most one exposure per day and at the same time in the day. So the unit exposure time is the day. Schedule 4 has many exposures on the third of the month. Analysis of that day shows exposures at 8:00 A.M., 9:00 A.M., 11:00 A.M., and 2:00 P.M. on the hour. In this case, hourly data contain at most only one exposure per hour and at the same time in the hour. So, the unit exposure time is the hour.

Note, we neither require nor assume equally spaced exposure times, as all the schedules in Figure 1 demonstrate. Rather, the analyst needs disaggregate data such that he or she can find an interval in which ad exposure occurs no more than once, and always at the same microperiod within that interval. This identification should be easy with the increasing availability of highly disaggregate data. In the final section we discuss how to handle the practical problem of bursts of tightly packed ad exposures separated by long intervals.

Note also that in the case of TV advertising, because consumers might vary in their exposure to ads and these exposures may be less frequent than the insertion of ads, the analyst could as well substitute ad insertions for ad exposures. However, this substitution might not be valid for monthly magazine or billboard advertising, where exposures might occur more often than (say, the monthly) insertion of ads.

Furthermore, the carryover effect of advertising is the total impact that any current advertising exposure has on consumer purchases in current and all future time intervals. So the duration of advertising's effects is the length of time during which any single advertising exposure continues to affect purchases. The p% duration interval is the length of time that accounts for p% of the advertising effect. The current effect of advertising is that portion of the total advertising effect that occurs in the same time interval as the exposure. The decay function is the mathematical expression for the duration effect of advertising. Finally, the duration interval bias is the difference between the carryover effect estimated at the true interval and that estimated on aggregate data.

2.2. Intuition

The principle underlying the analysis is that while aggregate data bias estimates, too disaggregated data greatly increase the costs of data collection, storage, and analysis. We suggest, then, that the largest data interval that provides unbiased estimates is the optimum level of data aggregation. The intuition behind the analysis is the following. Aggregate data assigns all exposures within the data interval to one arbitrary point in the interval and attributes the observed decay (primarily of the last exposure) as the average effect of all the exposures within that interval. This inappropriate attribution exaggerates the duration interval of all the exposures.

A data interval coincident with the unit exposure time turns out to be free from such misattribution. A data interval smaller than the unit exposure time does not interfere with the attribution of effects of successive exposures, and so does not bias the advertising's estimated duration interval, though it adds to the costs of data collection storage and analysis. Similarly, whatever the interpurchase time, it does not interfere with the observation of successive exposures, and so does not bias advertising's estimated duration interval.

2.3. Premises

Consistent with past research in this tradition, to simplify the analysis we assume a single supplier varying only advertising intensity to influence sales. We do not model the real but complex problem of cross-sectional aggregation. So, our analysis applies for sales arising from individual consumers or a group of homogenous consumers. In principle, we could allow any type of advertising response function, but for ease of notation we first discuss a Koyck-type advertising decay, and later we discuss other functions.

A basic premise is that there are two time series, one on sales and one on advertising. The empirical question concerns estimating the current effect, the carryover effect, and the duration interval of advertising's effect on sales.

Suppose the analyst has aggregate data. This aggregation really amounts to two processes:

1. Temporal aggregation of the microlevel data
2. Systematic sampling of only one of the averages in each interval of the aggregate data

For example, suppose the microdata are days and the aggregate data are weeks, measured on Sunday. Then what the researcher really has for each week are data from the prior Monday to Sunday, but not the prior Tuesday to Monday, or prior Wednesday to Tuesday, and so on. As we shall see later in the analysis, this systematic sampling is the core problem of aggregate data.

The second premise is that there is a true microfrequency at which the relationship between sales and advertising could be estimated without bias or model transformation, although such microdata might not be available to the researcher. We are interested in estimating the current effect, the carryover effect, and the duration interval for that particular frequency with minimal cost. At the present level of data collection, technology enables the researchers to obtain data at the highly disaggregate level, for example at minutes and seconds for TV advertising. So the research question we address is, "What is the most aggregate level at which to collect data without resulting in bias?"

A third important premise is that we can measure current and carryover effects only if we postulate a model for sales. All empirical outcomes are based on this model for sales. Another model gives another set of estimates. As is well known in the marketing literature, a model for aggregated data is usually not the same model as the one for disaggregated data. Examples are

1. An autoregression (AR) of order 1 for monthly data becomes an autoregressive-moving average (ARMA) model of orders $(1, 1)$ for bimonthly data.
2. The geometric distributed lag (also called Koyck) model becomes a model with lagged sales, current and lagged advertising, and with a moving average of order 1 on aggregation (see § 3.2.2).

This well-known result implies that fitting the same model to data at various levels of aggregation must lead to incorrect inferences.

3. The Koyck Model

In this section, we illustrate our main results for the familiar Koyck model. Later we will address other dynamic models.

3.1. Setup

Denote s_t and a_t as the sales and advertising variables at the true microdata interval, respectively. Assume that the true link between these two variables is given by the geometric distributed lag model, that is,

$$s_t = \mu + \beta a_t + \beta\lambda a_{t-1} + \beta\lambda^2 a_{t-2} + \cdots + \varepsilon_t, \qquad (1)$$

where μ is the intercept, and where ε_t is an uncorrelated error variable with variance σ_ε^2. The intercept μ does not play a role in what follows, and so we set it equal to 0 for convenience of notation. The current effect of advertising is β, while the carryover effect is equal to $\beta/(1-\lambda)$. The duration interval depends on λ. Under the assumption that $0 < \lambda < 1$, the so-called Koyck (1954) transformation (that is, multiply both sides by $1-\lambda L$, where L is the familiar lag operator defined by $L^k y_t = y_{t-k}$) reduces (1) to

$$s_t = \lambda s_{t-1} + \beta a_t + \varepsilon_t - \lambda\varepsilon_{t-1}. \qquad (2)$$

Assume that t amounts to the microinterval, so this model holds for this interval.

3.2. Relating Aggregate Model to True Micromodel

In order to determine the optimal data interval, we need to relate the aggregate model to the true micromodel. If one assumes the aggregate model to be the same Equation (1) that is true at the true microlevel, then one can derive the implied bias through the following three stages: (1) define the terms of the aggregate data, (2) derive the expression for the true aggregate form of the micromodel based on correctly aggregating and estimating the microdata, and (3) compare the true aggregate model with the estimated aggregate model.

3.2.1. Defining the Terms of the Aggregate Data. Denote S_T as the aggregate sales series that results from (1) aggregating sales in the K microperiods from the current to the $K-1$th prior period, and (2) sampling at the current period. Thus,

$$S_T = s_t + s_{t-1} + s_{t-2} + \cdots + s_{t-(K-1)}. \qquad (3)$$

To simplify notation, it is convenient to use the lag operator L. Then,

$$S_T = (1 + L + L^2 + \cdots + L^{K-1})s_t. \qquad (4)$$

Analogously, for aggregate advertising, aggregate error ε, and aggregate lagged sales we have, respectively,

$$A_T = (1 + L + L^2 + \cdots + L^{K-1})a_t, \qquad (5)$$

$$\varepsilon_T = (1 + L + L^2 + \cdots + L^{K-1})\varepsilon_t, \qquad (6)$$

and

$$S_{T-1} = (1 + L + L^2 + \cdots + L^{K-1})s_{t-K}. \qquad (7)$$

3.2.2. Deriving the True Aggregate Form of the Micromodel. We now derive the expression for the true aggregate form of the micromodel. To do so, first express Equation (1) (while setting $\mu = 0$) in terms of K microperiods of (1) taken together. That is,

$$\begin{aligned}(1+L+\cdots+L^{K-1})s_t &= \beta(1+L+\cdots+L^{K-1})a_t + \beta\lambda(1+L+\cdots+L^{K-1})a_{t-1}\\ &\quad + \beta\lambda^2(1+L+\cdots+L^{K-1})a_{t-2} + \cdots + (1+L+\cdots+L^{K-1})\varepsilon_t.\end{aligned} \qquad (8)$$

Appendix A.1 shows that application of the Koyck transformation to Equation (8) results in

$$\begin{aligned}(1 + L + \cdots + L^{K-1})s_t &= \lambda^K(1+L+\cdots+L^{K-1})s_{t-K} + \beta(1+L+\cdots+L^{K-1})a_t\\ &\quad + \beta\lambda(1 + \lambda L + \lambda^2 L^2 + \cdots + \lambda^{K-1}L^{K-1})\\ &\quad \cdot (1 + L + \cdots + L^{K-1})a_{t-1} + (1 + L + \cdots + L^{K-1})\varepsilon_t\\ &\quad - \lambda^K(1 + L + \cdots + L^{K-1})\varepsilon_{t-K}.\end{aligned} \qquad (9)$$

Now, from the definitions of the aggregate data in Equations (5) to (7), Equation (9) reduces to

$$S_T = \lambda^K S_{T-1} + \beta A_T + \beta\lambda(1+\lambda L+\lambda^2L^2+\cdots+\lambda^{K-1}L^{K-1}) \cdot (1+L+\cdots+L^{K-1})a_{t-1} + \varepsilon_T - \lambda^K\varepsilon_{T-1}. \tag{10}$$

Equation (9) is the true aggregate form of the micromodel in Equation (1). In other words, to estimate Equation (1) on aggregate data, the researcher can correctly use the aggregate forms of current sales, lagged sales, and current advertising. However, although the addition of aggregate lagged advertising A_{T-1} helps in approximating (10), it is still inadequate because

$$A_{T-1} \neq (1+\lambda L+\lambda^2L^2+\cdots+\lambda^{K-1}L^{K-1}) \cdot (1+L+\cdots+L^{K-1})a_{t-1}. \tag{11}$$

Also, there is no way for the researcher to accurately recover the information that is lost in aggregation of microadvertising data to arrive at A_{T-1}.

Past researchers have used various assumptions of the microadvertising process and estimation techniques to approximate the inequality in (11), if the true data interval were known. Russell (1988) suggested that all of them are special cases of a general model based on assuming a microstructure for a_{t-1}.

In contrast, we posit that we can correctly estimate the true parameters of the model, if we use sufficiently disaggregate data—what we call the *optimal data interval*.

3.3. Main Results

Proposition 1. *If* (a) *an advertising exposure occurs at most once in each Kth period*, and (b) *within the period, the pulse occurs at the same ith position in terms of microperiods, then the current effects of advertising, the carryover effects, and the duration interval at the true microfrequency can be estimated without bias* (*for the Koyck model*).

Proof. We prove our thesis for the Koyck model, because the model is a simple one and extensively used in marketing. Equation (11) represents an average of K successive microadvertising pulses, appropriately lagged and raised to a power of λ. Let advertising occur only once in each aggregate period at a microtime, say i. Then all K except i of the micropulses of advertising, within each moving average, equals 0. Thus (11) simplifies to $K-1$ expressions, in each of which there is at most only one advertising pulse at microperiod i. It depends on the location of i within K whether the pulse is in A_T or in A_{T-1}. To see this, define a_t^* such that it equals

$$a_t^* = \beta A_T + \beta\lambda(1+\lambda L+\lambda^2L^2+\cdots+\lambda^{K-1}L^{K-1}) \cdot (1+L+\cdots+L^{K-1})a_{t-1}, \tag{12}$$

and hence (10) becomes

$$S_T = \lambda^K S_{T-1} + a_t^* + \varepsilon_T - \lambda^K\varepsilon_{T-1}. \tag{13}$$

Appendix A.2 derives the properties of a_t^* and shows that if advertising occurs only once in each aggregate period, at microtime i, then Equation (12) becomes

$$a_t^* = \beta(1+\lambda+\cdots+\lambda^{K-i})A_T + \beta(\lambda^{K-i+1}+\cdots+\lambda^{K-1})A_{T-1}, \tag{14}$$

because the advertising at the other microperiods within T are zero. In sum, the model for the *Kth* period aggregated data becomes

$$S_T = \lambda^K S_{T-1} + \beta_1 A_T + \beta_2 A_{T-1} + \varepsilon_T - \lambda^K\varepsilon_{T-1}, \tag{15}$$

with

$$\beta_1 = \beta(1+\lambda+\cdots+\lambda^{K-i}), \tag{16}$$

and

$$\beta_2 = \beta(\lambda^{K-i+1}+\cdots+\lambda^{K-1}), \tag{17}$$

where for (15) $\beta_2 = 0$ if $i = 1$.

The first result that we arrive at is that the carryover effect of advertising is equal to

$$\frac{\beta_1+\beta_2}{1-\lambda^K} = \frac{\beta(1+\lambda+\cdots+\lambda^{K-i}) + \beta(\lambda^{K-i+1}+\cdots+\lambda^{K-1})}{1-\lambda^K}. \tag{18}$$

So, the carryover effect for the aggregate data is $(\beta_1+\beta_2)/(1-\lambda^K)$. This expression is equal to $\beta/(1-\lambda)$ if one samples the data at the optimal interval. Hence, we can simply use the extended model in Equation (13) and estimate it at the optimal data interval to get the carryover effects at the true microinterval!

Next, the true duration interval for the microinterval can be estimated without bias by taking the K-root of $\hat{\lambda}^K$. Finally, combining the estimated carryover effects and the value of λ allows one to estimate without bias the value of β, that is, the current effect of advertising.

These results all hold for any value of K, the level of aggregation, and any value of i, the position of ad pulse within K. In fact, it is not even necessary to know the value of i. However, when this value is known, one can also retrieve the value of β from the equality $\beta_1 = \beta(1+\lambda+\cdots+\lambda^{K-i})$. However, all these conclusions are true only when we choose a sufficiently disaggregate data interval such that advertising occurs only once and at the same point in that interval. □

Proposition 2. *Data intervals smaller than the unit exposure time* (*the optimal period*) *give unbiased estimates, provided that the model equations are properly adjusted.*

Proof. If the data are collected at the true microinterval, then the advertising effects can be estimated without bias. If the data are collected at a more disaggregate level, one needs to modify the model to take care of the appropriate lag structure. For example, suppose the microfrequency is weeks, and that the geometric distributed lag model

$$s_t = \beta a_t + \beta\lambda a_{t-1} + \beta\lambda^2 a_{t-2} + \cdots + \varepsilon_t \tag{19}$$

holds at this weekly frequency. Using the same techniques as before, but now in reverse order, we derive that for more disaggregate data such as days, this model effectively amounts to

$$\begin{aligned}(1+L+\cdots+L^6)s_{t^*} &= \beta(1+L+\cdots+L^6)a_{t^*} + \beta\lambda(1+L+\cdots+L^6)a_{t^*-7}\\ &\quad + \beta\lambda^2(1+L+\cdots+L^6)a_{t^*-14} + \cdots + \varepsilon_{t^*},\end{aligned} \tag{20}$$

where s_{t^*} and a_{t^*} denote measurements of sales and advertising at the daily level here, and with ε_{t^*} is $\varepsilon_t/(1+L+\cdots+L^6)$. Dividing both sides of this equation by $(1+L+\cdots+L^6)$ gives the geometric distributed lag model

$$s_{t^*} = \beta a_{t^*} + \beta\lambda a_{t^*-7} + \beta\lambda^2 a_{t^*-14} + \cdots + \varepsilon_{t^*}, \tag{21}$$

which upon using the Koyck transformation becomes

$$s_{t^*} = \lambda s_{t^*-7} + \beta a_{t^*} + \varepsilon_{t^*} - \lambda\varepsilon_{t^*-7}. \tag{22}$$

It can be seen that one can estimate the current and carryover effects without bias.

Note that the bias in estimates of the duration of advertising's effect is independent of the interpurchase time. It depends only on the unit exposure time. □

3.4. Simulations

3.4.1. Design of Simulations. This section reports the outcomes of a limited simulation experiment to show that data at the unit exposure time does not bias estimated coefficients. Because the values of the model's parameters define the values of both the current effect and the carryover effect, we focus on the latter two effects.

The simulation consists of 100 replications. We found that the general picture does not change if we increase the number of replications. The true model or data generating process (DGP) is a Koyck model. Data on advertising and sales are created in days, and then aggregated to the weekly and monthly level. We generate 7,000 days of data, which leads to 1,000 weeks and about 250 months of data, so our results are not likely to be due to small sample effects.

We choose the following values for the DGP1: σ_ε is 0.1, i is 4, β is 1, and λ is 0.9. The i value would imply an advertising pulse occurring, say, on a Thursday; the data are aggregated on Sundays. The value of λ implies a long-run effect $\beta/(1-\lambda)$ equal to 10. Finally, the advertising impulse at the daily level is generated as a downward rounded draw from a uniform distribution on the interval $[0, 1]$. Any other DGP for advertising would result in similar simulation results, because the model does not matter—only the pulse frequency matters.

Because the DGP is at the daily level and has little variance, we expect to estimate the original Koyck model at the daily level with great precision at this level.

For the aggregated weekly data we estimate a Koyck model with an additional lagged advertising term. Because advertising occurs once per week, Proposition 1 indicates that our weekly estimates will be unbiased estimates of the true current and carryover effects at the daily level.

For data aggregated to the monthly level, we again estimate a Koyck model with an additional advertising term, but now we expect the estimates to be biased, because aggregation entails another model. We do not consider further aggregation, because the results for the monthly data will already show bias, as we shall see.

3.4.2. Results of Simulation. Tables 1 to 4 give the results of the simulation. The cells in the tables are the minimum, mean, and maximum values of deviations, or bias if any, of the estimated parameters from the true parameters at the daily level. So, the smaller the value, the less the bias.

For the DGP1, Table 1 shows that the estimation precision at the daily and weekly level is about equal, with slightly larger minimum and maximum values for the weekly data, due to a smaller number of

Table 1 Simulation Results for DGP1

Parameter	Days[1]	Weeks[2]	Months[3]
Current effect (β)			
Minimum	−0.0174	−0.0461	−0.4573
Mean	0.0004	0.0015	1.6096
Maximum	0.0122	0.0437	9.2013
Carryover effect ($\beta/(1-\lambda)$)			
Minimum	−0.1848	−0.2188	−0.3157
Mean	−0.0065	−0.0003	0.0020
Maximum	0.1433	0.1688	0.2852

Note. λ is 0.9, β is 1, σ_ε is 0.1, and i is 4. The numbers in the cells measure the bias.

[1]The model estimated for the hourly frequency matches with the data-generating process. Sample size is 7,000.

[2]The model for the weekly data is the extended Koyck model, and the method to retrieve the key parameters is described in the text. The number of observations is 1,000.

[3]The model for the monthly data is also an extended Koyck model, and it is estimated for 250 observations.

Table 2 Simulation Results for DGP2

Parameter	Days[1]	Weeks[2]	Months[3]
Current effect (β)			
Minimum	−0.0142	−0.0320	−0.0973
Mean	−0.0002	0.0034	0.3455
Maximum	0.0142	0.0854	1.0168
Carryover effect ($\beta/(1-\lambda)$)			
Minimum	−0.1757	−0.2220	−0.2742
Mean	−0.0065	−0.0046	0.0067
Maximum	0.2142	0.2226	0.4600

Note. λ is 0.9, β is 0.1, σ_ε is 0.1, and i is 4. The numbers in the cells measure the bias.

[1]The model estimated for the daily frequency matches with the data-generating process. Sample size is 7,000.

[2]The model for the weekly data is the extended Koyck model, and the method to retrieve the key parameters is described in the text. The number of observations is 1,000.

[3]The model for the monthly data is also an extended Koyck model, and it is estimated for 250 observations.

Table 4 Simulation Results for DGP4

Parameter	Days[1]	Weeks[2]	Months[3]
Current effect (β)			
Minimum	−0.0013	−0.0049	−0.2695
Mean	0.0001	−0.0002	1.1322
Maximum	0.0012	0.0042	9.1271
Carryover effect ($\beta/(1-\lambda)$)			
Minimum	−0.0155	−0.0178	−0.1558
Mean	−0.0002	−0.0001	−0.0051
Maximum	0.0132	0.0178	0.1271

Note. λ is 0.9, β is 1, σ_ε is 0.01, and i is 4. The numbers in the cells measure the bias.

[1]The model estimated for the daily frequency matches with the data-generating process. Sample size is 7,000.

[2]The model for the weekly data is the extended Koyck model, and the method to retrieve the key parameters is described in the text. The number of observations is 1,000.

[3]The model for the monthly data is also an extended Koyck model, and it is estimated for 250 observations.

observations and an expected increase in the error variance. However, estimation with monthly data gives huge biases, especially for the current effect. This result shows that aggregation overestimates the current effect. Additionally, the distribution of the parameter estimates has become skewed. Recall that advertising occurs once a week. These results confirm Proposition 1, that data at the unit exposure time does not bias estimates.

We now change the value of β to a value that might be more often seen in advertising studies; that is, we change the value from 1 to 0.1. Call these data DGP2. The results are in Table 2. This table shows again that the current effects are heavily biased for the monthly data. However, the estimates of the weekly data do not differ from those for the daily data.

Table 3 Simulation Results for DGP3

Parameter	Days[1]	Weeks[2]	Months[3]
Current effect (β)			
Minimum	−0.0150	−0.0261	−0.4071
Mean	0.0003	0.0007	1.1390
Maximum	0.0128	0.0362	9.1766
Carryover effect ($\beta/(1-\lambda)$)			
Minimum	−0.1746	−0.1815	−0.6601
Mean	−0.0162	−0.0222	0.0269
Maximum	0.1493	0.1832	0.3854

Note. λ is 0.9, β is 1, σ_ε is 0.1, and i is 7. The numbers in the cells measure the bias.

[1]The model estimated for the daily frequency matches with the data-generating process. Sample size is 7,000.

[2]The model for the weekly data is the extended Koyck model, and the method to retrieve the key parameters is described in the text. The number of observations is 1,000.

[3]The model for the monthly data is also an extended Koyck model, and it is estimated for 250 observations.

We next set β back to 1, and change the timing of the advertising pulse. It was Thursday (say) and we set it at Sunday (the day of the aggregation and the same day as the observation). That is, we change i from four to seven. Call these data DGP3. The results are in Table 3. Compare the results in Table 1 with those in Table 3. Note that there are not many differences between DGP1 and DGP3 for estimates on the daily and weekly data. However, for monthly data we see bigger biases for estimates on DGP3 than for DGP1, especially for the carryover effects.

Next, we change the error variance in the DGP to 0.01. We then expect less bias for the data at the daily and the weekly level, because the variance is smaller. It remains to be seen if this also holds for the misspecified model at the monthly level. The results are in Table 4. We observe from Table 4 no problems for daily and weekly data, but the misspecified model for the monthly data delivers poor results, even though the variance is smaller.

Finally, we change the λ parameter to 0.7. In that case, the 95% duration interval is about one week. Because this interval comes closer to the aggregation interval, one might expect that it becomes more difficult to estimate the current effects from the weekly aggregated data, because these effects are now captured by single observations. The carryover effects should not be affected at least at the weekly level using our method. The results are in Table 5 and confirm our expectations.

In sum, these simulation results confirm Proposition 1 that data at the unit exposure interval do not bias estimates of the advertising response model.

4. Alternative Models

This section shows that the carryover effect can be retrieved from data aggregated to the unit exposure

Table 5 Simulation Results for DGP5

Parameter	Days[1]	Weeks[2]	Months[3]
Current effect (β)			
Minimum	−0.0145	−0.5794	−0.9971
Mean	−0.0011	0.4108	0.7532
Maximum	0.0119	2.3604	2.5468
Carryover effect ($\beta/(1-\lambda)$)			
Minimum	−0.0842	−0.1392	−0.2479
Mean	−0.0006	−0.0051	0.0771
Maximum	0.0696	0.1389	2.3386

Note. λ is 0.7, β is 1, σ_ε is 0.1, and i is 4. The numbers in the cells measure the bias.

[1]The model estimated for the daily frequency matches with the data-generating process. Sample size is 7,000.

[2]The model for the weekly data is the extended Koyck model, and the method to retrieve the key parameters is described in the text. The number of observations is 1,000.

[3]The model for the monthly data is also an extended Koyck model, and it is estimated for 250 observations.

interval also holds more generally for autoregressive distributed lag (ADL) models. We first discuss the case of stationary variables, and next briefly discuss nonstationary variables. Such ADL models for stationary data allow for a nonmonotonic decay in the response to advertising (sometimes called the Pascal lag). To illustrate, instead of the Koyck model at the microinterval, consider the ADL(1, 1) model

$$s_t = \lambda s_{t-1} + \beta_0 a_t + \beta_1 a_{t-1} + \varepsilon_t, \tag{23}$$

and suppose $K = 2$. Then we can write

$$(1+L)s_t = \lambda^2(1+L)s_{t-2} + (1+\lambda L)(1+L) \cdot (\beta_0 a_t + \beta_1 a_{t-1} + \varepsilon_t). \tag{24}$$

Using the same arguments as before, we can derive that the model, at the $K = 2$ frequency T with $i = 1$, is

$$S_T = \lambda^2 S_{T-1} + (\beta_0 + \beta_0\lambda + \beta_1)A_T + \beta_1\lambda A_{T-1} + U_T, \tag{25}$$

where U_T is a moving average process of order 1, MA(1) process. Evidently,

$$\frac{\beta_0 + \beta_0\lambda + \beta_1 + \beta_1\lambda}{1-\lambda^2} = \frac{\beta_0 + \beta_1}{1-\lambda}, \tag{26}$$

which is the carryover effect at the microlevel.

This model for the aggregated data has three unknown parameters (for the variables S_{T-1}, A_T, and A_{T-1}). The model at the microlevel also has three unknown parameters. Therefore, one can even retrieve the parameters β_0 and β_1.

Such a situation does not always occur, as can be understood from the ADL(1, 2) model with $K = 3$:

$$s_t = \lambda s_{t-1} + \beta_0 a_t + \beta_1 a_{t-1} + \beta_2 a_{t-2} + \varepsilon_t. \tag{27}$$

In that case, the model for the aggregate data would also have the regressors S_{T-1}, A_T, and A_{T-1}, but now there are four unknown parameters at the true microlevel. Hence, for this situation, one can retrieve the carryover effects, but not the true microlevel current effects.

So, what shall one do in practice? First, one needs to make sure that one collects data at the unit exposure time. Next, one fits an ADL-type model, or an extended Koyck model, for S_T. Using the parameters in this model, one can estimate the carryover effects at the true microlevel for any model that could lead to the aggregate model. If one additionally wants to estimate the current effects at the microlevel, one better assumes a microlevel model for s_t that allows for the identification of the microlevel current effects. For example, if one finds that

$$S_T = \rho S_{T-1} + \alpha_0 A_T + \alpha_1 A_{T-1} + U_T, \tag{28}$$

with U_T as an MA(1) process, is the best model for aggregate data, one better assumes that the true microfrequency model for s_t is an ADL(1, 1) model, because only in that case can one retrieve the current microeffects.

The above results also hold for multiple-equation models such as vector autoregressions. If the data are collected at the optimal data interval, one can use the impulse-response function to estimate the carryover effects by examining the ultimate value of this function. However, the shape of this function cannot easily be retrieved, because this depends on the model parameters. Of course, the impulse-response function of a model for data larger than the unit exposure time is also biased. The result follows directly from the direct dependence of the impulse-response function and the model parameters (see Rossana and Seater 1995, Swanson and Granger 1997, Marcellino 1999). So, impulse-response functions are not immune to temporal aggregation.

Finally, we briefly consider the case where the data on advertising and sales are not stationary, for example because they are upward or downward trending although not according to a fixed path. In time series language, one then says that these variables have a *unit root*; see Dekimpe and Hanssens (1999), among others.

There are two possible models for this situation. The first is where the aggregated data can be described by

$$S_T - S_{T-1} = \alpha_0(A_T - A_{T-1}) + (\rho - 1) \cdot \left(S_{T-1} - \frac{\alpha_0 + \alpha_1}{1-\rho} A_{T-1}\right) + U_T. \tag{29}$$

In this case, the two variables are cointegrated. The long-run effect is again equal to $(\alpha_0 + \alpha_1)/(1-\rho)$. The current effect is still equal to α_0, but the duration interval is now equal to infinity. This is due to the fact that impulses to a nonstationary variable last forever. Hence, in this case, the duration interval is irrelevant.

The second situation is the case where the data can be described by

$$S_T - S_{T-1} = \alpha_0(A_T - A_{T-1}) + U_T, \tag{30}$$

which is the case without cointegration. In this situation the total effect is infinite, because there is no tendency for the series to stabilize around an equilibrium value. The duration interval is infinitely long, as well. The only interesting parameter here is the current effect, which is again equal to α_0. Finally, when $\alpha_1 = -\alpha_0$, then A_T has a unit root and S_T not, and in that case the current effect is α_0 and the carryover is zero.

These results for nonstationary variables show that the estimation of the duration interval is irrelevant. However, the same results hold for the carryover effect as for the Koyck model.

5. Empirical Illustration

In this section we illustrate the above results for a real-life data set, taken from Tellis et al. (2000). We first describe the data, and next we show how our propositions carry through to real data.

5.1. The Data

For the empirical analysis, we use the data in Tellis et al. (2000), for the following reasons. First, advertising is the only marketing variable affecting consumer response. Second, data are at a highly disaggregate level of hours. Third, a model estimated on these data is already published in a leading marketing journal. Our goal is not to question the findings of that model. Our goal concerns how data aggregation or disaggregation affects the estimates of advertising response as predicted by our theoretical analysis. To aid the reader, we briefly describe the context of those data. Interested readers may pursue the original paper for details.

5.2. Study Context

The advertiser in Tellis et al. (2000) is a medical referral service. The firm advertises a toll-free number that customers can call to get the phone number and address of medical service providers. Consumers know the service by its advertised brand name, which represents the toll-free number (e.g., 1-800-BUILDER, 1-800-DENTIST, 1-800-LAWYER). They seek the service because they have recently moved, are unhappy with their current service provider, or have an immediate need for the service. When a customer calls the number, a representative of the firm answers the call. The representative queries the customer and then recommends a suitable service provider based on location, preferences, and specific type of service needed. Typically, the representative tries to connect the customer to the service provider directly by phone to minimize delay, miscommunication, and loss of customers. Any resulting contact between a customer and the service provider is called a referral. Customers do not pay a fee for the referral, but service providers, who are the firm's clients, pay a fixed monthly fee for a specific minimum number of referrals a month. The firm screens service providers before including them as clients. The firm began operations in March 1986 in the Los Angeles market with 18 service providers and a $30,000 monthly advertising budget. It currently advertises in over 62 major markets in the United States, with a multimillion dollar advertising budget that includes over 3,500 TV ad exposures per month.

The key dependent variable of interest to the firm is referrals. A referral is a call by a customer to the firm's answering service, which is successfully connected to a service provider by a firm's representative. The primary marketing variable that affects referrals is advertising. There are no distribution effects, product visibility effects, or store effects. The firm does not use sales promotions such as games, premiums, coupons, and so on. The primary means of advertising is television, though in some cities the firm advertises through radio, billboards, and yellow pages. In addition to advertising, experience and word of mouth can also affect referrals. For the time period being studied, the firm did not have any serious competition.

5.3. Illustrative Estimation Results

We illustrate the contents of our propositions for the Miami market. For this purpose we have 10,776 hourly data; see Tellis et al. (2000) for details on the data.

Given the nature of the advertising data, it is safe to assume that the microinterval here is 30 seconds. We do not have data at this interval. We do have data at the hourly interval. We assume that the unit exposure time is one hour. In our notation, this means that K is equal to 120, because there are 120 periods of 30 seconds within an hour. Next, we assume that the advertising pulse usually occurs right after the entire hour, so that i is close to or equal to K.

For the hourly data, we can estimate a model where R (referrals) depends on $R(-1)$, A, $A(-1)$ and a moving average term of order 1. We correct for the time that the service is available (open). This model is the extended Koyck model. We also include 24 dummies to capture intraday hourly seasonality. As Russell (1988) also shows, when an extended Koyck model can be fitted to aggregated data, the basic Koyck model may hold for the true microdata (30 seconds), which we do not have.

We use our methods and retrieve the λ and β parameters for the microinterval using the extended model for the K period aggregated interval (hourly data), we get that λ is 0.998 (with standard error 0.000267) and $\beta/(1-\lambda)$ is 2.834 (with standard error 0.124). This implies that β itself is estimated as 0.0070. With these parameters, we compute the 95% duration interval at the microfrequency, which turns out to be about 1,215 times 30 seconds, which is about 10.1 hours.

If $K = 120$ really matches with the unit exposure time, then an analysis of the model for hourly data would result in similar results. Indeed, our claim is that it is optimal to estimate the model for data collected at the unit exposure time. If we do that, we find for the basic Koyck model the duration interval is 9.39 hours. Notice that these numbers are very close to what we would find with data at 30-second intervals (suppose we would have these).

Now, we start aggregating the data. We first look at the 449 daily data. Again, we fit two models, the extended Koyck and the basic Koyck model. For the first model we get a 95% duration interval of about 220 (!) days, while for the second we get 6.8 days, which is of course way beyond the true value of 10 hours. The total carryover effect of advertising is estimated to equal 5.170 and 2.343 for the two, respectively. These numbers are reasonably close to the findings for hourly data, which confirms the theoretical results that temporal aggregation does not affect the total effect. Of course, smaller samples may lead to estimation biases and, hence, the potential reason for having differing estimated long-run effects.

Finally, we look at 65 observations of weekly data. We again fit the extended Koyck and basic Koyck model. For the first model we get a 95% duration interval of about 11.4 weeks, and for the second we get 5.4 weeks. Both again are way beyond the true interval of 10 hours. The total carryover effect of advertising for the two models is estimated to equal 4.195 and 3.781, respectively. Again, these numbers are reasonably close to the findings for hourly data.

In sum, the optimal level to analyze the Miami data is the hourly level. If the data are aggregated too much (such as in days and weeks), and the model is not changed, the estimated 95% duration intervals get way too high!

Further checks of our finding can be done as follows. Because we are able to estimate an extended Koyck model for the weekly data, we could see what would happen if we would have assumed that the optimal data interval is days. The weekly estimation results would then imply a λ of 0.9571 and a 95% duration interval (in terms of days) of 68.3 days. Notice that this is about 10 times as much as the 6.8 days we found above! This suggests that the weekly interval is also not the appropriate data interval.

Finally, we can see if perhaps the daily interval would be the optimal data interval, where we still assume that the 30 seconds is the true microinterval. The extended Koyck model for the daily data implies an estimate of λ for the hourly data of 0.99943, which in turn implies a 95% duration interval of 5,282 hours, which is about 220 days, whereas the Koyck model for the hourly data implies a duration interval of just 10.1 hours, as we saw earlier. This result suggests that the daily interval certainly is also not the appropriate interval. In sum, we conclude that the hourly data are the optimal data interval.

6. Discussion

Our study made several important conclusions. In contrast to Russell (1988), we show that the researcher does not need to assume the underlying advertising process in order to assess the true carryover of advertising. Rather, the researcher has only to pick the optimal data interval. We show that the optimal data interval is the unit exposure time, not the interpurchase time, as the literature states. If disaggregate data are not available, then one can still recover the true carryover of advertising using formulas that we provide. Most importantly, we show that using disaggregate data does not cause a disaggregation bias. To assess the usefulness of our proof, we next consider the role of our assumptions and the practicality of our recommendations. We then conclude with implications.

6.1. Role of Our Assumptions

We consider relaxing three assumptions of our model regarding other independent variables, heterogeneity, and consumer choices.

6.1.1. Other Independent Variables. Our results hold equally well for other independent variables besides advertising, such as price, promotion, and features. If any of these has its own carryover effect on sales, then the appropriate data interval would be the smallest one necessary to estimate the carryover of any of the independent variables. The reason is that data too aggregate for any independent variable would bias the effects of that variable and thus of all other independent variables correlated with it. But too disaggregate data does not bias coefficients of any variable.

6.1.2. Heterogeneity. Heterogeneity in the purchase pattern or the advertising response or decay function merely requires that we estimate the empirical model over homogenous segments, an approach facilitated by single-source advertising-purchase data (Tellis 1988).

6.1.3. Consumer Choices. With the availability of scanner data, a number of researchers have used choice data to analyze the effects of advertising (e.g., Tellis 1988, Tellis and Weiss 1995). Would our results extend to models of consumer choice? We believe that they would. To see this, consider a reformulation of the model for consumer choice. McFadden (1986) showed that with the assumption that consumers choose the brand that has the highest utility, consumer choice can be reformulated as one that maximizes utilities that are a linear function of brand attributes and Gumbel-distributed error terms. Advertising, price, and so on would then be incorporated as brand attributes analogous to our independent variables. Choice (0 or 1) would be incorporated as utility, analogous to our sales. Although a formal derivation of the appropriate data interval would need to be shown, we suspect our basic intuition may carry over. The level of the independent variable (advertising) would likely determine the level of the appropriate data, not the level of the dependent variable (choice).

6.2. Practicality of Our Recommendations

We consider several practical issues in implementing our recommendations: other measures of advertising, multiple media with varying times, channels with overlapping exposures, and advertising bursts.

6.2.1. Other Measures of Advertising. Our proof was developed for advertising measured as exposures. The reader might ask: Will the proof hold for advertising expenditures or gross ratings points (GRPs), which are commonly available measures for advertising data? We argue that it will. To see this, consider that advertising dollars or GRPs are merely an aggregate expression for advertising exposures. For example, x exposures total y GRPs for z dollars. Thus, what is true about the optimal data interval for exposures must be true for GRPs and dollars.

6.2.2. Multiple Media with Varying Times. When media have varying unit exposure times, we should use the medium with the smallest unit exposure time. That ensures that that there is no bias in estimating response to that particular medium. At the same time, the media with the larger unit exposure time will suffer no (disaggregation) bias from using a smaller interval.

6.2.3. Channels with Overlapping Exposures. Consider two media or channels with overlapping exposures. For example, consider CNBC news and ABC news, each with overlapping 30-second spots for the same brand. If a substantial number of consumers are routinely exposed to overlapping ads because they switch between channels, then one could consider the ad to last the duration of the sum of the two ads, while the unit exposure time would be the smaller of that for each medium. If ads on the two channels occur at varying times, then treat all ads as exposures from the same advertiser and find the optimal advertising interval from this series of exposures.

6.2.4. Advertising Bursts. Assume a highly clustered advertising schedule, in which each cluster has n densely packed exposures, with long periods of no advertising between clusters. We can treat each of these clusters of advertising as a single burst of advertising of strength n, where n is the number of exposures in the cluster. We can then apply our rule of the optimal data interval at the level of bursts of advertising rather than individual exposures. In so doing, we achieve savings in costs of data processing (gathering, storing, and analyzing) while perhaps introducing some small aggregation bias, relative to working solely at the level of exposures.

For example, consider multiple exposures of an ad that appears in a one-hour TV program that airs at the same time once a week. Suppose further that the unit exposure time is five minutes. At five-minute intervals our data task is enormous. However, by treating each hour of advertising as a burst of strength n, we can analyze data at the weekly level. As such, we may have allowed some small aggregation bias but may have saved substantially in reducing data processing costs.

6.3. Implications

Our paper has four important implications about past findings and future research. First, the most important implication of our study concerns the derived duration interval in reviews of the literature. By assuming that the true data interval is the interpurchase time, Clarke (1976, p. 355) concludes that "the duration of cumulative advertising effect on sales is between 3 and 15 months." Similarly, by assuming that the interpurchase time is the true data interval, Leone's (1995, p. G142) arrives at Generalization 2 that the duration interval of the effect of advertising averages between "six to nine months." However, for many product categories in the studies covered by Clarke (1976) and Leone (1995), the unit exposure time might be substantially below the interpurchase time. Thus, both the above estimates of the duration of the cumulative advertising effect might be too high.

Second, a major implication of our study is that researchers should attempt to carefully compute the unit exposure time in the category they study. They can then collect and analyze data at that interval or adjust for the bias if they use more aggregate data. For example, single-source data combine purchases recorded by checkout scanners with TV viewing recorded by TV-meters. These combined data make advertising data available at a highly disaggregate level. Researchers have analyzed these data at the weekly level because store promotions change by the week and consumers often shop on a weekly basis. However, for many product categories, television and radio advertising have unit exposure times in days. Thus, using even weekly data may positively bias advertising's estimated duration interval as established above.

Third, studies that have used highly disaggregate data are probably estimating true carryover effects rather than suffering any *disaggregation bias*. For example, using daily data, McDonald (1971) found that the most effective exposures were those seen not more than four days prior to a purchase. In other words, advertising's duration interval was unlikely to be more than four days long. The shorter duration intervals from the use of shorter data intervals are quite consistent with this and prior analyses (e.g., Clarke 1976) that aggregate data upwardly bias estimates of advertising's effects.

Fourth, the most important research direction would be to reanalyze the studies covered by Clarke (1976) and Leone (1995) to determine the duration of advertising's cumulative effect using the unit exposure time, as can be best estimated from the studies themselves or by additional historical research.

Fifth, dynamic effects are pervasive in marketing and the study of these important phenomena are quite common (e.g., Akcura et al. 2004, Besanko et al. 2005, Dubé and Manchanda 2005, Pauwels 2004, Seetharaman 2004, Van den Bulte and Stremersch 2004, Van Heerde et al. 2004). Researchers may want to assess to what extent the use of aggregate data or the inappropriate assumption of an aggregate data interval as the correct one has compromised findings from such studies.

Acknowledgments
The authors thank Byung-Do Kim, Koen Pauwels, Chuck Whiteman, and Doyle Weiss for help and comments on early stages of research, Marielle Non for excellent research assistance, and Christiaan Heij for very useful feedback. Finally, the authors thank the editor and four reviewers for their detailed suggestions that helped to improve the paper.

Appendix
In this appendix we show how to arrive at Equations (9) and (14).

A.1 Derivation of Equation (9)
We begin with Equation (8) and write it concisely as

$$\begin{aligned} S(L)s_t &= \beta S(L)a_t + \beta\lambda S(L)a_{t-1} \\ &\quad + \beta\lambda^2 S(L)a_{t-2} + \cdots + S(L)\varepsilon_t, \end{aligned} \tag{31}$$

where $S(L)$ denotes $(1 + L + \cdots + L^{K-1})$. Multiplying both sides of this equation with $\lambda^K L^K$ gives

$$\begin{aligned} \lambda^K S(L)s_{t-K} &= \beta\lambda^K S(L)a_{t-K} + \beta\lambda\lambda^K S(L)a_{t-K-1} \\ &\quad + \beta\lambda^2\lambda^K S(L)a_{t-K-2} + \cdots + \lambda^K S(L)\varepsilon_{t-K}. \end{aligned} \tag{32}$$

Subtracting (32) from (31) gives

$$\begin{aligned} S(L)s_t &= \lambda^K S(L)s_{t-K} + \beta S(L)a_t \\ &\quad + \beta\lambda\big[S(L)a_{t-1} + \lambda L S(L)a_{t-1} + \cdots + \lambda^{K-1}L^{K-1}S(L)a_{t-1}\big] \\ &\quad + S(L)\varepsilon_t - \lambda^K S(L)\varepsilon_{t-K}. \end{aligned} \tag{33}$$

When changing $S(L)$ back into $(1 + L + \cdots + L^{K-1})$, (33) can be written as (9).

A.2 Derivation of Equation (14)
The next equation we establish in this appendix is (14). Because A_T is $S(L)a_t$, and $a_{t-1} = La_t$, we can write Equation (12) as

$$a_t^* = \beta S(L)a_t + \beta\lambda L(1 + \lambda L + \lambda^2 L^2 + \cdots + \lambda^{K-1}L^{K-1})S(L)a_t. \tag{34}$$

This, in turn, can be written as

$$a_t^* = \beta(1 + \lambda L + \lambda^2 L^2 + \lambda^3 L^3 + \cdots + \lambda^{K-1}L^{K-1})S(L)a_t. \tag{35}$$

If advertising occurs only once in each aggregate period, at microtime i,

$$\beta(1 + \lambda L + \lambda^2 L^2 + \lambda^3 L^3 + \cdots + \lambda^K L^K) \tag{36}$$

can be decomposed into

$$\beta(1 + \lambda + \cdots + \lambda^{K-i}), \tag{37}$$

which gets assigned to A_T and

$$\beta(\lambda^{K-i+1} + \cdots + \lambda^{K-1}), \tag{38}$$

which gets assigned to A_{T-1} and then (35) becomes (14).

References

Akcura, M. T., F. F. Gönül, E. Petrova. 2004. Consumer learning and brand valuation: An application on over-the-counter drugs. *Marketing Sci.* **23**(1) 156–169.

Bass, F. M., R. Leone. 1983. Temporal aggregation, the data interval bias, and empirical estimation of bimonthly relations from annual data. *Management Sci.* **29**(January) 1–11.

Besanko, D., J.-P. Dubé, S. Gupta. 2005. Own-brand and cross-brand retail pass-through. *Marketing Sci.* **24**(1) 123–137.

Clarke, D. G. 1976. Econometric measurement of the duration of advertising effect on sales. *J. Marketing Res.* **13**(November) 345–357.

Clarke, D. G. 1982. A reply to Weinberg and Weiss. *J. Marketing Res.* **19**(November) 592–594. *J. Marketing Res.* **23**(August) 298–304.

Dekimpe, M. G., D. M. Hanssens. 1999. Sustained spending and persistent response: A new look at long-term marketing profitability. *J. Marketing Res.* **36**(November) 397–412.

Dubé, J.-P., P. Manchanda. 2005. Differences in dynamic brand competition across markets: An empirical analysis. *Marketing Sci.* **24**(1) 81–95.

Koyck, L. M. 1954. *Distributed Lags and Investment Analysis*. North-Holland Publishing Company, Amsterdam, The Netherlands.

Leone, R. P. 1995. Generalizing what is known about temporal aggregation and advertising carryover. *Marketing Sci.* **14** G141–G150.

Marcellino, M. 1999. Some consequences of temporal aggregation in empirical analysis. *J. Bus. Econom. Statist.* **17**(January) 129–136.

McDonald, C. 1971. What is the short-term effect of advertising? David Corkindale, Sherri Kennedy, eds. *Measuring the Effect of Advertising*. Saxon House Studies, Farnborough, UK, 463–487.

McFadden, D. 1986. The choice theory approach to market research. *Marketing Sci.* **5** 275–297.

Pauwels, K. 2004. How dynamic consumer response, competitor response, company support, and company inertia shape long-term marketing effectiveness. *Marketing Sci.* **23**(4) 596–610.

Rao, R. C. 1986. Estimating continuous time advertising-sales models. *Marketing Sci.* **5**(Spring) 125–142.

Rossana, R. J., J. J. Seater. 1995. Temporal aggregation and economic time—series. *J. Bus. Econom. Statist.* **13**(October) 441–451.

Russell, G. J. 1988. Recovering measures of advertising carryover from aggregate data: The role of the firm's decision behavior. *Marketing Sci.* **7**(3) 252–270.

Seetharaman, P. B. 2004. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Sci.* **23**(2) 263–271.

Swanson, N. R., C. W. J. Granger. 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *J. Amer. Statist. Assoc.* **92**(March) 357–367.

Tellis, G. J. 1988. Advertising exposure, loyalty and brand purchase: A two-stage model of choice. *J. Marketing Res.* **25**(2) 134–144.

Tellis, G. J., D. Weiss. 1995. Does TV advertising really affect sales? The role of measures, models and data aggregation. *J. Advertising* **24**(3) 1–12.

Tellis, G. J., R. Chandy, P. Thaivanich. 2000. Which ad works, when, where, and how often? Modeling the effects of direct television advertising. *J. Marketing Res.* **37**(February) 32–46.

Van den Bulte, C., S. Stremersch. 2004. Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Sci.* **23**(4) 530–544.

Van Heerde, H. J., P. S. H. Leeflang, D. R. Wittink. 2004. Decomposing the sales promotion bump with store data. *Marketing Sci.* **23**(2) 317–334.

Vanhonacker, W. R. 1983. Carryover effects and temporal aggregation in a partial adjustment model framework. *Marketing Sci.* **2**(Summer) 297–317.

Weinberg, C. B., D. L. Weiss. 1982. On the econometric measurement of the duration of advertising effects on sales. *J. Marketing Res.* **19**(November) 585–591.

Weiss, D. L., C. B. Weinberg, P. H. Windal. 1983. The effects of serial correlation and data aggregation on advertising measurement. *J. Marketing Res.* **20**(August) 268–279.