International Journal of Forecasting 33 (2017) 90–101

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Modeling intra-seasonal heterogeneity in hourly advertising-response models: Do forecasts improve?

Meltem Kiygi-Calli[a,*], Marcel Weverbergh[b,1], Philip Hans Franses[c,2]

[a] Faculty of Business and Administrative Sciences, Business Administration, Okan University, Tuzla Campus, 34959, Istanbul, Turkey
[b] Faculty of Applied Economics, Marketing Department, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium
[c] Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Advertising effectiveness
Advertising response
Aggregation
Normative and predictive validity

ABSTRACT

We examine the situation in which hourly data are available for designing advertising-response models, whereas managerial decision-making can concern hourly, daily or weekly intervals. A key notion is that models for higher frequency data require the intra-seasonal heterogeneity to be addressed, while models for lower frequency data are much simpler. We use three large, actual real-life datasets to analyze the relevance of these additional efforts for managerial interpretation and for the out-of-sample forecast accuracy at various frequencies.

1. Introduction

Ever since time series data on advertising and sales have become available, there have been discussions about the appropriate level of aggregation to use when estimating advertising responses, and when forecasts at different aggregation levels are required. A classic study is that by Clarke (1976), who shows that the longitudinal impact of advertising will be overestimated grossly if the analyst considers the same type of model for different levels of aggregation; see also Russell (1988). This notion is illustrated again by Tellis and Franses (2006), who show that the familiar Koyck model for higher frequency data becomes another, and more involved, time series regression model for aggregated data. In brief, one by-product of aggregation is that the model must change too. Of course, the reverse situation also holds true: high frequency data may reveal intra-seasonal heterogeneity, like hour-of-the-day effects, and hence, disaggregated data often require a more complex model.

A similar result appears for data transformations; see De Bruin and Franses (1999). When the higher frequency data are, say, log-transformed to mitigate the impact of extreme observations and to dampen the variance, a move towards aggregated data may remove the need for such transformations, as these extreme observations will be "aggregated away".

This paper considers the temporal aggregation of advertising response models, ranging from micro models at the hourly level to models for data aggregated to the weekly level. We find that different aggregation levels require appropriate data transformations. Our motivation for this study arises from the concern as to whether models at the hourly level can be expected to perform well as forecast tools for lower frequency data. It is well understood that, under perfect aggregation, disaggregated models will outperform models at a lower frequency, because of the loss of information incurred by aggregation.

* Corresponding author. Fax: +90 216 677 16 67.
E-mail addresses: meltem.kiygicalli@okan.edu.tr (M. Kiygi-Calli), marcel.weverbergh@uantwerpen.be (M. Weverbergh), franses@ese.eur.nl (P.H. Franses).
[1] Fax: +32 3 265 47 99.
[2] Fax: +31 10 408 91 62.

http://dx.doi.org/10.1016/j.ijforecast.2016.06.005
0169-2070/

However, this loss might be limited to a couple of percentage points for medium-sized samples. Thus, in real life situations, simpler models for aggregate data might conceivably prove to be more robust and give more reliable aggregate forecasts. A second question is the extent to which the implications of aggregate models are consistent with micro models. For example, is the effect of advertising obtained from a daily model similar to that from a micro model for hourly data? These questions are inspired by management practice, which usually involves the construction of a media plan based on a micro model, but where the performance assessment is based on aggregate data. We study these issues through a detailed analysis of three large databases of hourly advertising and sales data.

The remainder of the paper is organized as follows. Section 2 briefly discusses the literature related to our topic. Section 3 addresses the data used for the empirical application. Section 4 discusses the model specification for the hourly data. As the data show strong intra-day and intra-week seasonality, a two-level model is proposed, with the hour in the week as the observation unit. Section 5 presents single-level models for hourly, daily data and weekly data. Sections 6 and 7 summarize the forecasting and inference results. Finally, Section 8 contains the main conclusions and limitations.

2. Literature

The advertising literature contains various studies that address the degree of aggregation used to measure advertising effectiveness. A typical workhorse model is the familiar Koyck model, which correlates current sales with current advertising and past sales, and includes an error term with first order dynamics. Alternative models involve variants of the autoregressive distributed lag model (ADL) without moving average terms (MA). For both types of models, it holds that the parameter estimates can be used to infer the long-run (or cumulative) effects of advertising on sales, the immediate effect on sales, and the shape of the decay function, which gives the speed at which the effects of advertising impulses fade out to zero.

A key aspect of the models used in this body of literature is the fact that the parameter estimates, and their derivative functions, can depend on the aggregation level of the data. For example, if one is analyzing monthly data but the underlying process works at the weekly level because advertising impulses are given at the weekly level, then one may make estimation errors. This insight goes back to the work of Clarke (1976), as well as various subsequent studies, such as those of Bass and Leone (1983, 1986), Leone (1995), Windal and Weiss (1980) and Tellis and Weiss (1995).

There are various possible responses to this phenomenon. The first is to acknowledge the aggregation effects and modify the estimation routine accordingly; see Weiss, Weinberg, and Windal (1983), who propose a nonlinear GLS estimation technique that takes into account the effects of aggregation on both the estimates and the error terms. They use simulated monthly data and aggregate these into half-yearly and annual data, then look at how the estimates of the autoregressive term and the autocorrelation functions change under aggregation. They report that the level of aggregation does not cause an upward bias in the parameter estimates, but note that the probability of overestimating the lagged depended variable does increase because the aggregation is applied to small samples. Also, they attribute the upward bias in the parameter estimates to a misspecification of the model.

Second, it is recognized that aggregation can change the model. For example, when an autoregression of order 1 (AR(1)) is adequate for describing weekly data, the model becomes an autoregressive moving average model of order 1,1 (ARMA(1,1)) after aggregation. Russell (1988) argues that temporal aggregation of the data does not change the underlying advertising–sales relationship, but typically the model is misspecified at the aggregated level. To retrieve the micro-frequency parameters, one needs to know (or assume) what the micro-frequency model looks like. Recently, Tellis and Franses (2006) used this result to determine the optimal level of aggregation, in order to ensure that this retrieval is still possible.

A third response to aggregation issues, which has become possible due to the recent increase in the availability of high frequency data, is to simply rely on models for the highest frequency. The seminal articles on high frequency models are those by Chandy, Tellis, MacInnis, and Thaivanich (2001) and Tellis, Chandy, and Thaivanich (2000). These studies indicate that models for high frequency data can be used as the basis for decision support systems in media planning regarding the optimal choice of channels, time slots, and spot lengths, and insights about the relative effectiveness of appeals and wear-in wear-out effects can also be obtained. Tellis and Franses (2006) use such high frequency data to show that the optimal data interval or aggregation level is the unit exposure time, which is the time interval in which consumers would typically be exposed to a single advertising spot. These studies advocate the use of high frequency data, even if decisions have to be made at a more aggregated level. This paper will examine the latter issue in more detail.

Even though much of the literature is dedicated to the issue of aggregation and its effects on (functions of) parameter estimates, the focus is rarely on out-of-sample forecasting. Since the main use of econometric advertising models is for budgeting and strategy planning, most managers require accurate forecasts for planning and decision making. Note that the planning and decision horizon does not necessarily match the available data frequency. For example, managers may need to make budgetary plans for the next year, and therefore require forecasts of the next year's sales level. If a modeler has access to weekly data, then it is an open question as to whether a model using such weekly data is useful for forecasting the next year's total sales. Aggregation would entail a loss of information, and hence the parameter estimators would be less efficient. At the same time, aggregation should not lead to inconsistency if the model is modified and specified appropriately. However, the main issue is that higher-frequency data require more complex models to take into account the increased information content. More data may also mean more data patterns, such as intra-seasonal patterns, and these should be modeled too.

The aim of the present study is to shed light on this managerially relevant issue. We analyze a case that is actually

Table 1
Characteristics of the three actual data sets.

Feature	Datasets		
	1 Flanders region	2 Walloon region	3 Spain
Sample	01-10-2004 22-06-2007	01-01-2005 22-06-2007	01-01-2004 30-06-2006
Number of hours: total	23,880	21,672	9251
Estimation sample	16,056	14,448	6646
Forecast sample	7824	7224	2605
Number of days	995	903	608
Number of weeks	143	130	88
Number of weeks with commercials	48	43	37
Number of radio spots	6144	7432	4827
Number of TV spots	252	330	0
Call center operating hours per week	168	168	70

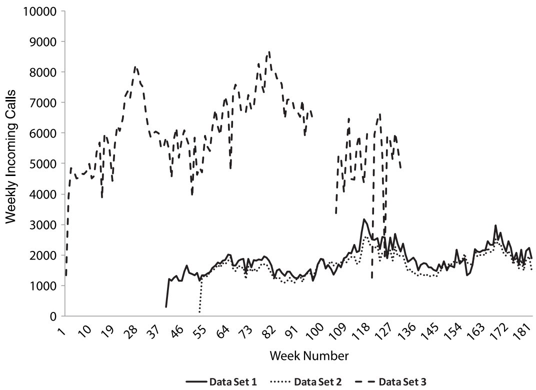

Fig. 1. Weekly incoming calls.

very relevant for the company at hand, who allowed us to use their data. The data and the high frequency model for one of the three datasets used below were discussed by Kiygi Calli, Weverbergh, and Franses (2012), and many of the details on the specification of models with intra-seasonal features can be found there too. As high frequency (e.g., hourly) data can show large variations due to potential outliers, we also show that aggregation impacts data transformation. Indeed, it is common to use transformations such as the natural log transformation to dampen the variation in, say, hourly data; however, we find this transformation to be less relevant for aggregated weekly or monthly data.

3. Real-life data

This section analyzes data relating to a car repair service provided by a multinational. We consider three different areas where data are collected, namely Flanders (the Dutch-speaking part of Belgium), Wallonia (the French-speaking part of Belgium), and Spain. All three samples are relatively large (23,880, 21,672 and 9251 hourly observations respectively). For the evaluation, each datasets is split into an estimation sample (2/3 of the observations) and an out-of-sample forecast part (1/3). National call centers, for Belgium (an almost 24/7 service) and Spain separately, collect all requests for information and service appointments from consumers. The two markets run on similar principles, but with different operating hours. Of course, the three areas' media plans and communication channels are all quite different. The dependent variable is the number of incoming calls received by the call center. Table 1 provides an overview of the three datasets, where the high frequency data are hourly data.

In all cases, the company relies mainly on radio advertising, and to a limited extent on television advertising in Belgium. There is no comparable service advertised at the national level, and therefore the data are very well suited for an analysis of advertising effects. There is little or no variation in the advertising themes, but the core message is framed in different formats or spot lengths.

For the Flanders region (data set 1), the calls data cover the period from October 2004 to June 2007. Our analysis is performed on data at hourly intervals (23,880), which include a total of 6144 radio spots and 252 television spots. Television advertising occurred in 2006 and 2007 only.

For the Walloon region (data set 2), the calls data cover the period from January 2005 to June 2007, giving 21,672 data points, a total of 7432 radio spots and 330 television spots. The television advertising occurred in 2006 and 2007 only. For radio, 20-s spots are used predominantly, with other spot lengths being used much less. For television, there is a relatively large number of five-second spots, which are the so-called 'sponsoring spots' associated with the weather forecasts.

For Spain (data set 3), the available data cover the period from January 1, 2004, to June 30, 2006, with a two-week period missing in the middle, as Fig. 1 shows. For this market, we have data for 9251 h. During the observation period, 4827 radio commercials were broadcasted on twenty-five radio stations, with most being 60-s commercials. Spots at 8 AM were used frequently. The distribution of broadcasts is more or less equal over the weekdays. In the dataset, the commercials start at 6 AM. The call centre operates from 8 AM to 8 PM on weekdays, has shorter service hours on Saturdays, and is closed on Sundays. Thus, the call centre operates 70 h per week.

The advertising strategy of the company makes use of 'pulsing'. This means that all advertising is scheduled in a subset of weeks, which are alternated with non-active weeks. Advertising reach of a commercial is measured by means of Gross Rating Points (GRPs) (Tellis, 2004). Table 1 shows that this subset consists of approximately one third

Table 2
Average radio GRPs per hour of the day.

Dataset 1	Hour	6	7	8	9	10	11	12	
	GRPs	6	14	7	6	6	6	11	
	Spots	45	201	154	141	153	130	169	
	Hour	13	14	15	16	17	18	19	20
	GRPs	3	2	4	6	3	4	2	1
	Spots	33	17	34	51	94	133	85	9
Dataset 2	Hour	6	7	8	9	10	11	12	
	GRPs	5	15	16	11	13	10	7	
	Spots	54	181	161	75	113	109	98	
	Hour	13	14	15	16	17	18	19	20
	GRPs	10	12	2	4	3	3	7	0
	Spots	126	2	13	70	82	64	39	0
Dataset 3	Hour	6	7	8	9	10	11	12	
	GRPs	9	9	36	6	2	2	2	
	Spots	87	127	177	68	37	25	19	
	Hour	13	14	15	16	17	18	19	20
	GRPs	4	3	1	2	1	3	1	1
	Spots	20	88	1	17	5	12	3	6

$$
\begin{aligned}
Y_{h,t} = {} & \theta_h + \overbrace{\lambda_{1,h}Y_{h-1,t} + \lambda_{2,h}Y_{h-2,t} + \lambda_{3,h}Y_{h-3,t} + \lambda_{24,h}Y_{h-24,t} + \lambda_{168,h}Y_{h,t-1}}^{\text{Autoregressive Terms}} \\
& + \overbrace{\phi_h^0 R_{h,t} + \phi_h^1 R_{h-1,t} + \phi_h^2 R_{h-2,t}}^{\text{Radio advertising}} + \overbrace{\gamma_h^0 TV_{h,t} + \gamma_h^d TVD_{h,t}}^{\text{TV advertising}} \\
& + \overbrace{\delta_h^S \sin\frac{2\pi t}{52} + \delta_h^C \cos\frac{2\pi t}{52}}^{\text{seasonality}} + \overbrace{\beta_1 Tr_{h,t}}^{\text{trend}} + \overbrace{\beta_2 B_{h,t} + \beta_3 B_{h-24,t} + \beta_4 B_{h+24,t}}^{\text{holiday}} + \varepsilon_{h,t},
\end{aligned}
\tag{2}
$$

Box I.

of the weeks, with average GRP levels equal to 237 for data set 1, 312 for data set 2, and 249 for data set 3.

Fig. 1 shows the numbers of incoming calls ("sales") in the three areas. The numbers of daily incoming calls for data sets 1 and 2 are approximately equal, while the market size of data set 3 is substantially larger.

The three panels in Fig. 2 show the average levels of incoming calls for the three call centers for hours with and without advertising. The intraday patterns are highly similar. The peak in calls occurs between 8 AM and 9 AM, with a smaller peak in the early afternoon, at around 2 PM.

Table 2 provides a summary of the average radio GRPs per hour of the day. The radio commercials are most often broadcasted between 6 AM and 8 PM, and the average GRPs vary between 0.70 and 11.52. For data set 1, the GRPs are highest at 7 AM. For data sets 2 and 3, the maximum values are at 8 AM.

Next, we will analyze these hourly data using multi-level models, after which we will aggregate to the daily and weekly levels, and fit models and produce forecasts accordingly.

4. Two-level models for hourly data

We resort to a two-level regression model in order to capture the substantial variation in the data at the hourly level properly. In short, we treat the hour within the day as the observation unit, so we deal with sales denoted by $Y_{h,t}$, where the subscript h denotes the hour and t denotes the week. As such, we have a panel of time series, where the panel consists of 168 units and the time frame covered is 143, 130 or 88 weeks, respectively (see Table 1). For this panel, we choose to consider the linear mixed model (LMM; see for example Verbeke & Molenberghs, 2000), where the model reads as

$$
\begin{aligned}
& Y_{h,t} = Z_{h,t} b_h + \varepsilon_{h,t}, \\
& \text{with} \\
& b_h = W_{h,t} b + \varepsilon_h .
\end{aligned}
\tag{1}
$$

In words, the first level contains parameters that can possibly vary with the hour of the week (h), and the second level correlates those parameters to the characteristics of the particular hours that are incorporated in $W_{h,t}$. As such, this model allows the substantial variation in the data that is present at this highly disaggregated level to be captured.

After some experimentation (along lines similar to those outlined by Kiygi Calli et al., 2012), we fix the first level of the three models as given in Box I. where the variables in this first-level model are defined as follows:

$Y_{h,t} = \log(Calls_{h,t} + 1)$, where h runs from 1 to 168 and t runs from 1 to T,

$R_{h,t} = \log(RadioGRP_{h,t} + 1)$,

$TV_{h,t} = \log(TVGRP_{h,t} + 1)$, and

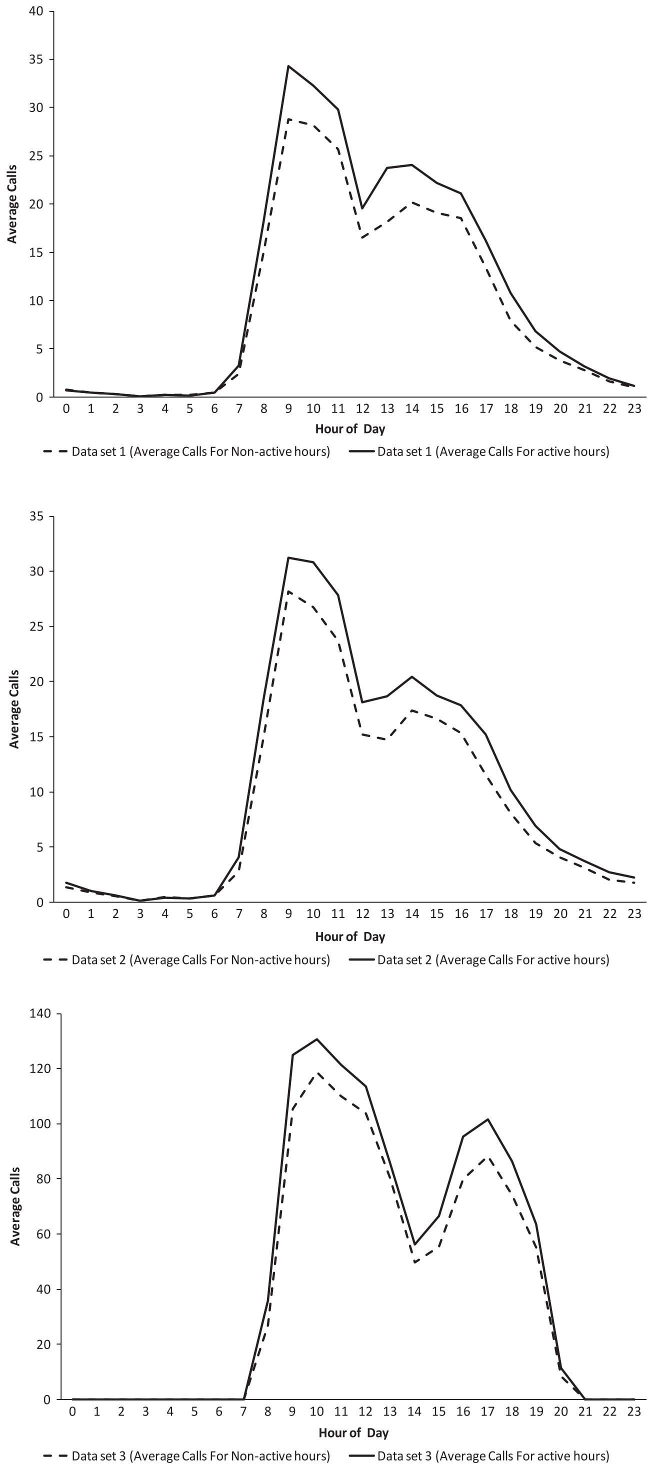

Fig. 2. Average numbers of calls in weeks with and without commercial activity.

$\sin\frac{2\pi t}{52}$, $\cos\frac{2\pi t}{52}$ are harmonic or goniometric regressors that capture the intra-year seasonality in the data;

and where $TVD_{h,t}$ denotes the log of the total amount of TV GRPs during the previous day. Note that this last variable takes the same value for 24 h in a row. $B_{h,t}$ is a dummy variable for bank holidays, and $Tr_{h,t}$ is a trend defined as $t/52$, where t is 1,2,…, total week number.

The second-level equations allow for distinct specifications across the three data sets, and include time variation in the coefficients across the hour of the week or the day. For example, the autocorrelation is found to vary substantially all three samples, with a short cycle of one day. The results show that the autocorrelation is much higher at peak hours, which leads in part to a higher effectiveness of advertising during peak hours. The time variability for the advertising coefficients is less pronounced. The characteristics of advertising spots, length and channel are included in order to detect differences in advertising effectiveness.

The second-level equations start from the same specification for all three data sets. To achieve parsimony, the final model is obtained by deleting non-significant variables with t-values less than one.

The intercept term in each dataset is given by

$$\theta_h = \theta_0 + \theta_1 \sin \frac{2\pi hd}{24} + \theta_2 \cos \frac{2\pi hd}{24} + \sum_{d=1}^{6} \theta_{3,d} D_h^d, \quad (3)$$

where hd is the hour of the day $(1, 2, \ldots, 24)$, and D_h^d is a dummy variable which takes a value of one for day d and zero otherwise, with $d = 1$ meaning a Monday.

For the first-order autoregressive parameters, we specify

$$\lambda_{1,h} = \lambda_{1,0} + \sum_{d=1}^{7} \lambda_{1,d,s} D_h^d \sin \frac{2\pi hd}{24} + \sum_{d=1}^{7} \lambda_{1,d,c} D_h^d \cos \frac{2\pi hd}{24} + \varepsilon_{1,h}. \quad (4)$$

The second-order and third-order autoregressive terms are specified as constants. For the second-order autoregressive term, we specify

$$\lambda_{2,h} = \lambda_{2,0}, \quad (5)$$

whereas for the third-order autoregressive term we have

$$\lambda_{3,h} = \lambda_{3,0}. \quad (6)$$

In all three samples, the two further autoregressive terms are modeled as

$$\lambda_{24,h} = \lambda_{24,0} + \varepsilon_{2,h} \quad (7)$$

and

$$\lambda_{168,h} = \lambda_{168,0} + \varepsilon_{3,h}. \quad (8)$$

For the regressors concerning the commercials on radio and television, we have the following second-level equations for all three data sets:

$$\begin{aligned} \phi_h^0 &= \phi_0^0 + \sum_{j=1}^{Channels} \phi_j^0 FC_{h,j} + \varepsilon_{4,h} \\ \phi_h^1 &= \phi_0^1 + \sum_{j=1}^{Channels} \phi_j^1 FC_{h-1,j} \\ \phi_h^2 &= \phi_0^2 + \sum_{j=1}^{Channels} \phi_j^2 FC_{h-2,j} + \sum_{i=1}^{Spot\ Lengths} \phi_i^2 FL_{h-2,i}, \end{aligned} \quad (9)$$

where $FC_{h,j}$ is the fraction of radio GRPs per channel j, and $FL_{h,i}$ is the fraction of radio GRPs per spot length i. For the first television variable, we specify

$$\gamma_h^0 = \gamma_0^0 + \sum_{j=1}^{Channels} \gamma_j^0 FTV_{h,j} + \sum_{hd=1}^{23} \gamma_{hd}^0 D_{h,hd}, \quad (10)$$

where $FTV_{h,j}$ is the fraction of television GRPs per channel j, and $D_{h,hd}$ is a zero–one dummy variable indicating the hour of the day within a day.

For the next day television variable, we specify

$$\gamma_h^d = \gamma_0^d + \sum_{j=1}^{Channels} \gamma_j^d TVDC_{h,j} + \sum_{i=1}^{Spot\ Lengths} \gamma_i^d TVDL_{h,i} + \sum_{d=1}^{6} \gamma_h^d D_h^d, \quad (11)$$

where $TVDC_{h,j}$ is the fraction of television GRPs per channel j during the previous day, and $TVDL_{h,i}$ is the fraction of television GRPs per spot length.

Finally, for the annual seasonality, we specify

$$\delta_h^s = \delta_0^s + \delta_1^s \sin\left(\frac{2\pi hd}{12}\right) + \delta_2^s \cos\left(\frac{2\pi hd}{12}\right) + \varepsilon_{5,h} \quad (12)$$

$$\delta_h^c = \delta_0^c + \delta_1^c \sin\left(\frac{2\pi hd}{12}\right) + \delta_2^c \cos\left(\frac{2\pi hd}{12}\right) + \varepsilon_{6,h}. \quad (13)$$

The independent variables relate to the media schedule (time of broadcast, channel and length, or equivalently, the theme of a spot). In principle, all equations could be specified with a random error; however, this makes the model too complex, causing estimation problems. For the second-level equations, some trial and error has been applied. We tried several different forms for these equations, but the current ones turned out to be best in terms of the in-sample fit. We are aware that the best in-sample fit does not guarantee the best out-of-sample forecast accuracy, but the complexity of the models is such that we see no other way of selecting our final models. The first-order, daily and weekly autoregressive terms, current radio effect and annual seasonality, are all specified with a random error.

5. Alternative models

In addition to the model in the previous section, we also consider more traditional autoregressive distributed lag (ADL) models for the data at three aggregation levels, namely hourly, daily and weekly. However, the preferred model for the hourly data is that presented in the previous section. This model allows for heterogeneity across hours within a week, and, as such, is very useful for capturing variation in the data.

The two-level (hourly) model in the previous section used a logarithmic specification. However, the transformation to logarithms is not necessarily appropriate for all levels of aggregation. We rely on the Box–Cox transformation to capture the non-stable variance in the data. As was discussed by Box and Cox (1964), this transformation is given by

$$y^{(\lambda)} = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0. \end{cases} \quad (14)$$

For a single-level model for hourly data Y (in levels), where h now runs from 1 to the total number of hours, as indicated in Table 1, we specify

$$
\begin{aligned}
Y_h^{(\lambda)} = {} & \mu + \overbrace{\lambda_1 Y_{h-1}^{(\lambda)} + \lambda_2 Y_{h-2}^{(\lambda)} + \lambda_3 Y_{h-3}^{(\lambda)} + \lambda_{24} Y_{h-24}^{(\lambda)} + \lambda_{168} Y_{h-168}^{(\lambda)}}^{\text{Autoregressive Terms}} \\
& + \overbrace{\phi^0 R_h^{(\lambda)} + \phi^1 R_{h-1}^{(\lambda)} + \phi^2 R_{h-2}^{(\lambda)}}^{\text{Radio advertising}} + \overbrace{\gamma^0 TV_h^{(\lambda)} + \gamma^d TVD_h^{(\lambda)}}^{\text{TV advertising}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \sum_{l=0}^{2} \varphi_{1,j,l} RC_{j,h-1}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \sum_{l=0}^{2} \varphi_{2,i,l} RL_{i,h-1}^{(\lambda)}}^{\text{Radio channel and length effects}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \varphi_{3,j} TVC_{j,h}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \varphi_{4,j} TVL_{i,h}^{(\lambda)}}^{\text{TV channel and length effects}} \\
& + \overbrace{\sum_{hw=1}^{167} \varphi_{5,hw} D_{h,hw} R_{h,hw}^{(\lambda)}}^{\text{Radio heterogeneous effects}} + \overbrace{\sum_{hw=1}^{167} \varphi_{6,hw} D_{h,hw} TV_{h,hw}^{(\lambda)}}^{\text{TV heterogeneuos effects}} \\
& + \overbrace{\delta^s \sin \frac{2\pi h}{168x52} + \delta^c \cos \frac{2\pi h}{168x52}}^{\text{seasonality}} + \overbrace{\beta_1 Tr_{h,t}}^{\text{trend}} \\
& + \overbrace{\beta_2 B_{h,t} + \beta_3 B_{h-24,t} + \beta_4 B_{h+24,t}}^{\text{holiday}} + \overbrace{\sum_{k=1}^{167} \varphi_{9,k} H_{h,k}}^{\text{hour of week}} + \varepsilon_h, \qquad (15)
\end{aligned}
$$

where

$$
\begin{aligned}
Y_h &= Calls_h + 1, \\
R_h &= RadioGRP_h + 1, \\
TV_h &= TVGRP_{h,t} + 1,
\end{aligned}
$$

and $RC_{j,h}$ are the radio GRPs per channel j, $RL_{i,h}$ are the radio GRPs per spot length i, and the equivalent variables for television are $TVC_{j,h}$ and $TVL_{i,h}$. $D_{h,hw}$ is a dummy for hour hw of a week, and $H_{h,k}$ is a dummy for hour h in the week.

This model is similar to that of Tellis et al. (2000), who also use hourly data, but they consider referrals to health care services. Eq. (15) allows a substantial degree of heterogeneity both through the hourly dummies affecting the intercept and in the effects of the explanatory variables.

The model for daily data is

$$
\begin{aligned}
Y_T^{(\lambda)} = {} & \theta + \overbrace{\lambda_1 Y_{T-1}^{(\lambda)} + \lambda_2 Y_{T-7}^{(\lambda)}}^{\text{Autoregressive Terms}} \\
& + \overbrace{\phi^0 R_T^{(\lambda)}}^{\text{Radio advertising}} + \overbrace{\gamma^0 TV_T^{(\lambda)} + \gamma^1 TV_{T-1}^{(\lambda)}}^{\text{TV advertising}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \varphi_{1,j} RC_{j,T}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \varphi_{2,i} RL_{i,T}^{(\lambda)}}^{\text{Radio channel and length effects}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \varphi_{3,j} TVC_{j,T}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \varphi_{4,i} TVL_{i,T}^{(\lambda)}}^{\text{TV channel and length effects}} \\
& + \overbrace{\sum_{d=1}^{6} \varphi_{5,d} D_{d,T} R_T^{(\lambda)}}^{\text{Radio heterogeneous effects}} + \overbrace{\sum_{d=1}^{6} \varphi_{6,d} D_{d,T} TV_T^{(\lambda)}}^{\text{TV heterogeneous effects}} \\
& + \overbrace{\delta^S \sin \frac{2\pi T}{52 \times 7} + \delta^C \cos \frac{2\pi T}{52 \times 7}}^{\text{seasonality}} + \overbrace{\beta_1 Tr_{d,T}}^{\text{trend}} \\
& + \overbrace{\beta_2 B_{d,T} + \beta_3 B_{d-1,T} + \beta_4 B_{d+1,T}}^{\text{holiday}} \\
& + \overbrace{\sum_{d=1}^{6} \varphi_0 D_{d,T}}^{\text{day of week}} + \varepsilon_T, \qquad (16)
\end{aligned}
$$

where $D_{d,T}$ is the day of the week. Finally, the model for weekly data is

$$
\begin{aligned}
Y_W^{(\lambda)} = {} & \theta + \overbrace{\lambda_1 Y_{W-1}^{(\lambda)}}^{\text{Autoregressive Term}} + \overbrace{\phi^0 R_W^{(\lambda)}}^{\text{Radio advertising}} + \overbrace{\gamma^0 TV_W^{(\lambda)}}^{\text{TV advertising}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \varphi_{1,j} RC_{j,W}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \varphi_{2,i} RL_{i,W}^{(\lambda)}}^{\text{Radio channel and length effects}} \\
& + \overbrace{\sum_{j=1}^{\text{Channels}} \varphi_{3,j} TVC_{j,W}^{(\lambda)} + \sum_{i=1}^{\text{Lengths}} \varphi_{4,i} TVL_{i,W}^{(\lambda)}}^{\text{TV channel and length effects}} \\
& + \overbrace{\delta^S \sin \frac{2\pi W}{52} + \delta^C \cos \frac{2\pi W}{52}}^{\text{seasonality}} \\
& + \overbrace{\beta_1 Tr_W}^{\text{trend}} + \overbrace{\beta_2 B_W}^{\text{holiday}} + \varepsilon_W. \qquad (17)
\end{aligned}
$$

We also consider the two-level model in the previous section, where we apply the Box–Cox transformation. In the first level of the Box–Cox transformed LMM model, Eq. (2) becomes the equation in Box II.

Finally, we also consider mixed data sampling (MIDAS) models, which allow the regressand and regressors to be measured with different frequencies (Clements & Galvao, 2009), with the regressors appearing at a higher frequency than the regressand (Ghysels, Sinko, & Valkanov, 2007). In our case, the MIDAS specification is obtained by aggregating the dependent variable, but keeping the micro detail on the right hand side, at least for the variables of interest, a procedure that Ghysels et al. (2007) claim protects against aggregation bias.

We analyze daily and weekly MIDAS models. The heterogeneity of the hour, channel and length effects of radio and TV ads are maintained by specifying such effects at a high frequency in the MIDAS models. Daily and weekly MIDAS model specifications, respectively, are given as

$$
\begin{aligned}
Y_T = {} & \theta + \overbrace{\lambda_1 Y_{T-1} + \lambda_2 Y_{T-7}}^{\text{Autoregressive Terms}} + \overbrace{\phi^0 R_T}^{\text{Radio advertising}} \\
& + \overbrace{\gamma^0 TV_T + \gamma^1 TV_{T-1}}^{\text{TV advertising}}
\end{aligned}
$$

$$Y_{h,t}^{(\lambda)} = \theta_h + \overbrace{\lambda_{1,h}Y_{h-1,t}^{(\lambda)} + \lambda_{2,h}Y_{h-2,t}^{(\lambda)} + \lambda_{3,h}Y_{h-3,t}^{(\lambda)} + \lambda_{24,h}Y_{h-24,t}^{(\lambda)} + \lambda_{168,h}Y_{h,t-1}^{(\lambda)}}^{\text{Autoregressive Terms}}$$
$$+ \overbrace{\phi_h^0 R_{h,t}^{(\lambda)} + \phi_h^1 R_{h-1,t}^{(\lambda)} + \phi_h^2 R_{h-2,t}^{(\lambda)}}^{\text{Radio advertising}} + \overbrace{\gamma_h^0 TV_{h,t}^{(\lambda)} + \gamma_h^d TVD_{h,t}^{(\lambda)}}^{\text{TV advertising}}$$
$$+ \overbrace{\delta_h^S \sin\frac{2\pi t}{52} + \delta_h^C \cos\frac{2\pi t}{52}}^{\text{seasonality}} + \overbrace{\beta_1 Tr_{h,t}}^{\text{trend}} + \overbrace{\beta_2 B_{h,t} + \beta_3 B_{h-24,t} + \beta_4 B_{h+24,t}}^{\text{holiday}} + \varepsilon_{h,t}. \quad (18)$$

Box II.

$$+ \overbrace{\sum_{h=1}^{23} \varphi_{1,h} D_h R_{h,T}}^{\text{High frequency radio heterogeneous hour effects}}$$
$$+ \overbrace{\sum_{h=1}^{23} \varphi_{2,h} D_h TV_{h,T}}^{\text{High frequency TV heterogeneous hour effects}}$$
$$+ \overbrace{\sum_{h=1}^{23} \varphi_{3,h} D_h R_{h-24,T}}^{\text{Radio high frequency previous day effect}}$$
$$+ \overbrace{\sum_{h=1}^{23} \varphi_{4,h} D_h TV_{h-24,T}}^{\text{TV high frequency previous day effect}}$$
$$+ \overbrace{\overbrace{\sum_{j=1} \varphi_{5,j} RC_{j,T}}^{\text{Channels}} + \overbrace{\sum_{i=1} \varphi_{6,i} RL_{i,T}}^{\text{Lengths}}}^{\text{Radio channel and length effects}}$$
$$+ \overbrace{\overbrace{\sum_{j=1} \varphi_{7,j} TVC_{j,T}}^{\text{Channels}} + \overbrace{\sum_{i=1} \varphi_{8,i} TVL_{i,T}}^{\text{Lengths}}}^{\text{TV channel and length effects}}$$
$$+ \overbrace{\delta^S \sin\left(\frac{2\pi T}{365}\right) + \delta^C \cos\left(\frac{2\pi T}{365}\right)}^{\text{seasonality}} + \overbrace{\beta_1 Tr_T}^{\text{trend}}$$
$$+ \overbrace{\beta_2 B_{d,T} + \beta_3 B_{d-1,T} + \beta_4 B_{d+1,T}}^{\text{holiday}}$$
$$+ \overbrace{\sum_{d=1}^{6} \varphi_{9,d} D_{d,T}}^{\text{day of week}} + \varepsilon_T \quad (19)$$

and

$$Y_W = \theta + \overbrace{\lambda_1 Y_{W-1}}^{\text{Auroregressive Term}} + \overbrace{\phi^0 R_W}^{\text{Radio advertising}} + \overbrace{\gamma^0 TV_W}^{\text{TV advertising}}$$
$$+ \overbrace{\sum_{d=1}^{6} \varphi_{1,d} D_d R_{d,W}}^{\text{High frequency radio heterogeneous day effects}}$$
$$+ \overbrace{\sum_{d=1}^{6} \varphi_{2,d} D_d TV_{d,W}}^{\text{High frequency TV heterogeneous day effects}}$$
$$+ \overbrace{\sum_{d=1}^{6} \varphi_{3,d} D_d R_{d-7,W}}^{\text{Radio high frequency previous day effect}}$$
$$+ \overbrace{\sum_{d=1}^{6} \varphi_{4,d} D_d TV_{d-7,W}}^{\text{TV high frequency previous day effect}}$$
$$+ \overbrace{\overbrace{\sum_{j=1} \varphi_{5,j} RC_{j,d,W}}^{\text{Channels}} + \overbrace{\sum_{i=1} \varphi_{6,i} RL_{i,d,W}}^{\text{Lengths}}}^{\text{Radio channel and length effects}}$$
$$+ \overbrace{\overbrace{\sum_{j=1} \varphi_{7,j} TVC_{j,d,W}}^{\text{Channels}} + \overbrace{\sum_{i=1} \varphi_{7,i} TVL_{i,d,W}}^{\text{Lengths}}}^{\text{TV channel and length effects}}$$
$$+ \overbrace{\delta^S \sin\frac{2\pi W}{52} + \delta^C \cos\frac{2\pi W}{52}}^{\text{seasonality}}$$
$$+ \overbrace{\beta_1 Tr_W}^{\text{trend}} + \overbrace{\beta_2 B_W}^{\text{holiday}} + \varepsilon_W. \quad (20)$$

In addition, we also estimate linear, logarithmic and Box–Cox transformed MIDAS models and evaluate their predictive accuracies. All of the model parameters are estimated using maximum likelihoods.

6. Estimation results

We compare the various models of interest by considering their forecast accuracies.

We begin with the two-level model for hourly data in Section 4. The (5% significant) parameter estimates for data set 1 are displayed in Table 3. The estimation results for the other two data sets can be obtained from the authors, but are not given here to save space. Clearly, these estimation results show the relevance of the two-level model. The parameters show a strong variation across the hours, especially for the first order autoregressive parameters.

When we estimate the models in Eqs. (15)–(17), the values of the estimated Box–Cox parameters are as

Table 3
Estimation results for the LMM model for data set 1 (the estimation results for the other data sets can be obtained from the authors).

Second level	Parameter	Estimate	Standard error	Second level	Parameter	Estimate	Standard error
Intercept	θ_0	0.559	0.020	Current radio GRP	ϕ_0^0	NS	
	θ_1	−0.363	0.018		ϕ_2^0	0.055	0.018
	θ_2	−0.459	0.019		ϕ_4^0	0.043	0.024
First-order lag	$\lambda_{1,0}$	0.194	0.013		ϕ_5^0	0.044	0.030
	$\lambda_{1,1,s}$	−0.118	0.041	One-hour lag	ϕ_0^1	0.026	0.008
	$\lambda_{1,1,c}$	−0.308	0.038		ϕ_1^1	−0.063	0.027
	$\lambda_{1,2,s}$	−0.081	0.045		ϕ_4^1	−0.070	0.039
	$\lambda_{1,2,c}$	−0.233	0.042				
	$\lambda_{1,4,s}$	−0.040	0.043	Two-hour lag	ϕ_0^2	NS	
	$\lambda_{1,4,c}$	−0.228	0.040		ϕ_1^2	0.073	0.024
	$\lambda_{1,5,s}$	−0.089	0.043	Current TV GRP	γ_0^0	0.240	0.059
	$\lambda_{1,5,c}$	−0.200	0.040		γ_1^0	−0.068	0.068
	$\lambda_{1,3,s} = \lambda_{1,6,s} = \lambda_{1,7,s}$	−0.020	0.025	Hour of the day	γ_2^0	−1.251	0.555
	$\lambda_{1,3,c} = \lambda_{1,6,c} = \lambda_{1,7,c}$	−0.071	0.024	One-day lag	γ_0^d	0.012	0.006
Second-order lag	$\lambda_{2,0}$	0.079	0.008		γ_2^d	−0.028	0.012
Third-order lag	$\lambda_{3,0}$	0.057	0.008	Seasonality	δ_0^s	0.035	0.017
One-day lag	$\lambda_{24,0}$	0.068	0.010		δ_1^s	0.009	0.007
One-week lag	$\lambda_{168,0}$	0.071	0.012		δ_2^s	−0.001	0.001
Trend	β_1	0.070	0.007		δ_0^c	0.068	0.014
Bank holiday	β_2	−0.413	0.026		δ_1^c	−0.034	0.006
	β_3	0.154	0.025		δ_2^c	0.003	0.001
	β_4	−0.059	0.023				

Note: NS means "not significant".

Table 4
Estimated Box–Cox parameters (with standard errors) for Eqs. (15)–(17).

Frequency	Dataset		
	1	2	3
Hours	0.192	0.225	0.620
	(0.005)	(0.006)	(0.009)
Days	0.211	0.469	0.601
	(0.031)	(0.053)	(0.030)
Weeks	1.561	1.518	1.327
	(0.009)	(0.000)	(0.000)

presented in Table 4. As expected, the estimated values get closer to zero for the higher frequency data and closer to or over one for the aggregated data, like weekly data. This pattern is quite consistent across the three data sets. These results show that the optimal model changes from logarithmic to linear as the aggregation level changes. This largely solves the problem of aggregate models producing unrealistic estimates of the advertising effectiveness.

Table 5 shows (part of the) estimation results of the model for hourly data in Eq. (15), for the logarithmic model, which is equivalent to a Box–Cox parameter of zero. The estimation results in this table can be compared to those in Table 3. When we compare the parameters for the TVGRP variables, we do see quite considerable differences across tables.

To highlight some of the crucial differences across the estimated models in Eqs. (15)–(17), consider the parameter estimated for the weekly lag in Table 6, which corresponds to lag 168 for the model for hourly data

Table 5
Selected estimation results of Eq. (15) for hourly data for data set 1.

Variable	Parameter estimate	(standard error)
Intercept	0.343	(0.035)
AR term, 1-hour lag	0.186	(0.008)
AR term, 2-hour lag	0.110	(0.008)
AR term, 3-hour lag	0.065	(0.008)
AR term, 24-hour lag	0.089	(0.008)
AR term, 168-hour lag	0.096	(0.007)
Radio GRP, current hour	−0.051	(0.081)
Radio GRP, 1-hour lag	0.005	(0.011)
Radio GRP, 2-hour lag	0.032	(0.009)
TV GRP, current hour	0.189	(0.064)
TV GRP, 24-hour lag	0.008	(0.006)
Yearly sine function	0.063	(0.006)
Yearly cosine function	−0.015	(0.005)

and lag 7 for the model for daily data. Note that this parameter is crucial for computing the rate of decay of the advertising impact. As expected, given the seminal work of Clarke (1976) and others, this parameter increases as the frequency decreases; that is, upon aggregation. To illustrate, for data set 1 and no transformation, consider the parameter 0.096 for the hourly data, compared to 0.233 for the weekly data. Hence, the effect of advertising seems to last longer when weekly data are considered.

7. Forecasting accuracy

In order to compare different models' out-of-sample forecasting performances, we back-transform Box–Cox

Table 6
Estimates of the parameter (and associated standard errors) for the weekly lag (168 for hourly data, seven for daily data and one for weekly data), based on Eqs. (15)–(17).

Transformation	Frequency	Dataset		
		1	2	3
None	Hourly	0.096 (0.007)	0.079 (0.008)	0.238 (0.008)
	Daily	0.088 (0.021)	0.130 (0.032)	0.011 (0.034)
	Weekly	0.233 (0.112)	0.100 (0.042)	0.330 (0.080)
Box–Cox	Hourly	0.101 (0.007)	0.084 (0.007)	0.212 (0.008)
	Daily	0.109 (0.021)	0.143 (0.031)	0.149 (0.032)
	Weekly	0.657 (0.023)	0.494 (0.074)	0.647 (0.000)

transformed models to the original scale and compare their predictive accuracies with those of the other models. In other words, we estimate forecasts in the transformed form for the Box–Cox transformed models (Eqs. (15)–(18)). Hence, back-transformation induces a bias in the forecasts (De Bruin & Franses, 1999). For one-step-ahead forecasts, we compute de-biasing factors for the back-transformations as given by Guerrero (1993). This de-biasing factor involves the standard deviations of individual observations, and therefore cannot be used for multi-step forecasts, given the autoregressive nature of the models. We then compute the average empirical bias from the estimation sample according to the number of steps ahead, and correct the forecasts for this bias (Guerrero, 1993). The results indicate that de-biasing factors have no positive impact on the forecasting accuracy, and sometimes even have a negative impact. The bias appears to improve the fit in the estimation sample, but may cause a deterioration in forecast accuracy in the forecast sample.

Table 7 reports the predictive accuracy of the models for hourly, daily and weekly data, where the aim is to forecast the data as hours, days and weeks. We investigate both one-step-ahead and multi-step-ahead forecasts, using recursive formulas for the latter forecasts.

Table 7 also shows the Diebold–Mariano test results, which examine whether there is a statistical difference between the accuracies of two competing forecasts (Diebold & Mariano, 1995). In the table, the numbers in boldface have the highest forecast accuracies. For each panel (forecast horizon) and column, RMSPE values that are underlined and in boldface are not significantly different at the 5% level from that with the highest forecast accuracy, according to the Diebold–Mariano test.[3] The first panel of Table 7 shows that the Box–Cox transformed ADL and LMM models in Eqs. (15) and (18) perform best for forecasting hourly data. Interestingly, the second panel shows that the same models also perform best when forecasting days ahead. They do better across the models for hourly data, but they also outperform the other models for daily data. Likewise, approximately similar results are obtained when it comes to forecasting weeks ahead.

The outcomes of the Diebold–Mariano tests indicate that, for hour-ahead forecasts, Box–Cox transformed ADL and LMM models provide better forecast accuracies than the other models. For day-ahead forecasts, hourly ADL linear and Box–Cox transformed ADL and LMM models give better accuracy results than the others. For week-ahead forecasts, hourly Box–Cox transformed LMM and ADL models give the highest forecast accuracies, while weekly Box–Cox transformed and linear ADL models also give higher forecast accuracies than daily models.

Table 8 presents the predictive accuracies of the MIDAS models that are given in Eqs. (19) and (20), and shows that Box–Cox transformed and linear MIDAS models achieve the highest forecast accuracies. We also conduct F-tests with the null hypothesis of homogenous time, channel and length effects, and conclude that these effects are not significantly different from zero. The F-test results can be obtained from the authors, but are not given here to save space. We also conduct the F-test with a null hypothesis that there is no significant difference between the hourly GRP effects. For the daily MIDAS estimation, we fail to reject the null hypothesis, with an F-value of 0.53 (p-value: 0.90). Therefore, we conclude that the daily MIDAS model given in Eq. (19) with non-significant hourly GRP effects boils down to a daily ADL model.

8. Conclusion and discussion

This paper presents detailed estimation and forecasting results for three large datasets that contain high frequency information on advertising and responses. We cannot claim to have discovered a generalizing principle for these three cases, but we do find that properly-specified models for high frequency data can achieve appropriate forecast accuracies, and compare well with models for temporally aggregated data.

Therefore, the main argument in this paper is that models for higher frequency data probably do not have constant parameters, and that a forecaster should take care of that. This intra-seasonal heterogeneity can be captured in a two-level regression model (LMM). In our case, a model for hourly data requires the variation across those hours to be modeled, and we advocate using the LMM for this purpose. Aggregating the hourly data to weekly data reduces this variation. Using three real-life cases, we have shown that, when done properly, models for higher frequency data are more useful than forecasts from models for aggregated data in terms of managerial use, and equally useful in terms of forecasting such aggregated data.

Hence, rather than aggregating high frequency data, designing models for the aggregated data and computing forecasts, we recommend modeling the high frequency data directly. It should also be noted that the specification changes, as the aggregation level increases, from logarithmic at the hourly level, to a square root transformation at the daily level, to approximately linear at the weekly level. This finding is particularly relevant for normative models. We observed that a simple aggregation that did not take

[3] We provide the Diebold–Mariano test results in the supplementary document (Appendix A).

Table 7
Forecast performances, in terms of root mean squared prediction errors.

Frequency	Model	Dataset 1	Dataset 2	Dataset 3
		Forecast horizon: hours ahead		
Hours (one-step-ahead forecast)	LMM logs	5.10	4.83	23.51
	LMM Box–Cox	4.82	**4.16**	**16.14**
	ADL logs	4.72	4.45	35.80
	ADL linear	4.76	4.52	23.76
	ADL Box–Cox	**4.66**	4.35	23.74
		Forecast horizon: days ahead		
Hours (multi-step-ahead forecast)	LMM logs	68.8	87.2	332
	LMM Box–Cox	59.4	56.9	**196**
	ADL logs	**54.9**	52.8	385
	ADL linear	55.2	**36.9**	234
	ADL Box–Cox	**57.2**	**46.5**	305
Days (one-step-ahead forecast)	LMM logs	70.4	62.0	399
	LMM Box–Cox	61.6	51.1	448
	ADL logs	63.5	60.5	423
	ADL linear	58.1	44.7	256
	ADL Box–Cox	57.8	47.0	402
		Forecast horizon: weeks ahead		
Hours (multi-step-ahead forecast)	LMM logs	305	512	1382
	LMM Box–Cox	209	215	**922**
	ADL logs	**203**	237	2348
	ADL linear	232	226	**1354**
	ADL Box–Cox	222	238	1422
Days (multi-step-ahead forecast)	LMM logs	428	398	2156
	LMM Box–Cox	412	342	2082
	ADL logs	389	424	2758
	ADL linear	274	258	1901
	ADL Box–Cox	334	292	1807
Weeks (one-step-ahead forecast)	ADL logs	276	439	1619
	ADL linear	211	**190**	1263
	ADL Box–Cox	228	**187**	2134

Notes: the boldface and underlined boldface numbers are the smallest RMSEs per panel and column.

Table 8
Forecast performances, in terms of root mean squared prediction errors of MIDAS specifications.

Frequency	Model	Dataset 1	Dataset 2	Dataset 3
		Forecast horizon: days ahead		
Days (one-step-ahead forecast)	MIDAS log	71.1	63.7	**401**
	MIDAS linear	**60.5**	**47.0**	**412**
	MIDAS Box–Cox	**60.7**	50.2	450
		Forecast horizon: weeks ahead		
Days (multi-step-ahead forecast)	MIDAS log	459	448	2611
	MIDAS linear	282	278	**2290**
	MIDAS Box–Cox	356	343	**1892**
Weeks (one-step-ahead forecast)	MIDAS log	**216**	381	3327
	MIDAS linear	**232**	**164**	3555
	MIDAS Box–Cox	**208**	**273**	3424

Notes: the boldface and underlined boldface numbers are the smallest RMSEs per panel and column.

into account this change in functional form led to unrealistic normative implications, which may differ substantially from those of the corresponding micro models. In particular, we observed considerably overestimated advertising effects in some cases.

In general, the results in our paper support the notion that high frequency data deserve to be analyzed. These high frequency data models include parameters that provide the proper interpretation in terms of decay rates and the short-run effects of advertising, and, as we have shown, also provide accurate forecasts for any policy horizon of interest. In general, our study supports the modeling exercises of Chandy et al. (2001), Tellis et al. (2000) and Tellis and Franses (2006), which generally recommend analysis at the most detailed level possible, as aggregation removes useful information and does not necessarily produce better forecasts.

The natural limitation of our study concerns the empirical data at hand. More actual data sets could be analyzed, and additional further work analyzing other elements of the marketing mix, like pricing and promotions, would also be relevant. At present, most marketing-response models rely on (at most) weekly data, and perhaps this is another situation in which one could benefit from the analysis of less aggregated data.

Acknowledgments

The authors thank two anonymous referees and an associate editor for very helpful comments, and acknowledge the financial support of the Flanders Research Foundation (FWO) (Project number: G.0361.07).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2016.06.005.

References

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

Bass, F. M., & Leone, R. P. (1986). Estimating micro relationships from macro data: a comparative study of two approximations of the brand loyal model under temporal aggregation. *Journal of Marketing Research*, *23*(3), 291–297.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *26*, 211–243.

Chandy, R., Tellis, G. J., MacInnis, D., & Thaivanich, P. (2001). What to say when: advertising appeals in evolving markets. *Journal of Marketing Research*, *38*(4), 399–414.

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

Clements, M. P., & Galvao, A. B. (2009). Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics*, *24*, 1187–1206.

De Bruin, P., & Franses, P. H. (1999). Forecasting power-transformed time series. *Journal of Applied Statistics*, *26*(7), 807–815.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–265.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: further results and new directions. *Econometric Reviews*, *26*, 53–90.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, *12*, 37–48.

Kiygi Calli, M., Weverbergh, M., & Franses, P. H. (2012). The effectiveness of high-frequency direct-response commercials. *International Journal of Research in Marketing*, *29*, 98–109.

Leone, R. P. (1995). Generalizing what is known about temporal aggregation and advertising carryover. *Marketing Science*, *4*(3), G141–G150.

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

Tellis, G. J. (2004). *Effective advertising: Understanding when, how, and why advertising works*. California: Sage Publications, Inc.

Tellis, G. J., Chandy, R., & Thaivanich, P. (2000). Which ad works, when, where, and how often? Modeling the effects of direct television advertising. *Journal of Marketing Research*, *37*(1), 32–46.

Tellis, G. J., & Franses, P. H. (2006). Optimal data interval for estimating advertising response. *Marketing Science*, *25*(3), 217–229.

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

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Berlin: Springer Science + Business Media, LLC.

Weiss, D. L., Weinberg, C. B., & Windal, P. M. (1983). The effects of serial correlation and data aggregation on advertising measurement. *Journal of Marketing Research*, *20*(3), 268–279.

Windal, P. M., & Weiss, D. L. (1980). An iterative GLS procedure for estimating the parameters of models with autocorrelated errors using data aggregated over time. *Journal of Business*, *53*, 415–424.