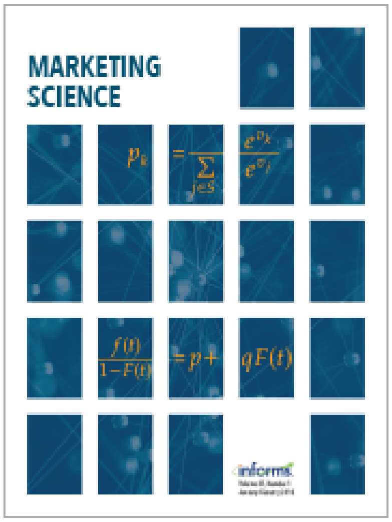


Marketing Science

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

Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights

Mingyung Kim, Eric T. Bradlow, Raghuram Iyengar

informs
https://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 41, No. 4, July–August 2022, pp. 848–866
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Selecting Data Granularity and Model Specification Using the Scaled Power Likelihood with Multiple Weights

Mingyung Kim,[a,*] Eric T. Bradlow,[a] Raghuram Iyengar[a]

[a] Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
*Corresponding author

Contact: mingyung@wharton.upenn.edu, https://orcid.org/0000-0002-3463-4023 (MK); ebradlow@wharton.upenn.edu, https://orcid.org/0000-0002-3405-6083 (ETB); riyengar@wharton.upenn.edu, https://orcid.org/0000-0002-7999-302X (RI)

Received: May 24, 2020
Revised: October 15, 2020; May 26, 2021; September 24, 2021; November 1, 2021
Accepted: November 3, 2021
Published Online in Articles in Advance: February 8, 2022

https://doi.org/10.1287/mksc.2021.1340

Abstract. Firms employ temporal data for predicting sales and making managerial decisions accordingly. To use such data appropriately, managers need to make two major analysis decisions: (a) the temporal granularity (e.g., weekly, monthly) and (b) an accompanying demand model. In most empirical contexts, however, model selection, sales forecasts, and managerial decisions are vulnerable to both of these choices. Whereas extant literature has proposed methods that can select the best-fitted model (e.g., Bayesian information criterion) or provide predictions robust to model misspecification (e.g., weighted likelihood), most methods assume that the granularity is either correctly specified or prespecify it. Our research fills this gap by proposing a method, the scaled power likelihood with multiple weights (SPLM), that not only identifies the best-fitted granularity-model combination jointly, but also conducts doubly (granularity and model) robust prediction against their potentially incorrect selection. An extensive set of simulations shows that SPLM has higher statistical power than extant approaches for selecting the best-fitted granularity-model combination and provides doubly robust prediction in a wide variety of misspecified conditions. We apply our framework to predict sales for a scanner data set and find that, similar to our simulations, SPLM improves sales forecasts due to its ability to select the best-fitted pair via SPLM's dual weights.

History: Olivier Toubia served as the senior editor and Carl Mela served as associate editor for this article.
Supplemental Material: The online appendices and data are available at https://doi.org/10.1287/mksc.2021.1340.

Keywords: data granularity • granularity-model selection • doubly robust prediction

1. Introduction

Firms and researchers routinely employ temporal data for investigating a variety of marketing decisions. In some instances, one may wish to infer the underlying data generation process and derive insights regarding optimal microlevel decisions. Alternatively, a different goal may be to forecast more temporally aggregate demand to plan the marketing mix (e.g., advertising, price), production levels, and sales force allocation. Still another goal, given that marketing actions (e.g., advertising budgets) are routinely decided in "batched and chunked" temporal periods (e.g., quarterly; Mela et al. 1997), is to predict drivers of demand at the granularity of the firm's managerial use (e.g., quarterly). For addressing some of these questions, it is ideal to use the most granular data available, as this "analyzes the data as they lie." For example, in a retail context, one goal can be to assess how customers traverse in real time through a store and what products they purchase (e.g., Hui et al. 2009). In this case, real-time (the most granular) path data likely offer the best way to address the question, as opposed to data aggregated at either a daily or weekly level, which both have significant information (feature) loss. In contrast, for addressing other questions like forecasting (as addressed in this paper), firms and researchers have leeway in choosing (a) a level of temporal granularity at which to analyze their in-sample data and (b) a model given the chosen granularity (i.e., the granularity-model pair). This is the topic addressed in this research.

We address this topic from the point of view of in-sample fit. Specifically, we propose in this paper a novel heuristic selection tool (statistic) that builds upon extant research from the statistics and machine learning literature and allows a researcher to *jointly* identify the best-fitted in-sample granularity-model pair that can be used for out-of-sample forecasting. We note that this is in stark contrast (as mentioned earlier) to extant research that assumes the granularity level before model selection.

We employ the following vignette to showcase our granularity-model pair selection tool and contrast it with extant approaches. Suppose that a brand manager of orange juice wants to generate the upcoming year's forecasts based on a demand model relating sales to advertising and price. To this end, she asks her analyst to provide the best-fitted model specification that she can use to forecast the upcoming year's sales. The analyst estimates a demand model at three commonly used granularities—weekly, monthly, and quarterly. He realizes that a single model specification is not appropriate across granularities, as the week-to-week demand variation versus the quarter-to-quarter variation needs to be captured differently. In other words, he needs to choose not just a demand model but a model-granularity pair. Thus, at each granularity, the analyst specifies a set of models. He also discovers that using the most granular data (e.g., weekly) does not necessarily provide the most accurate forecasts. More exploration indicates that model misspecification due to unmodeled short-term dynamics (e.g., stockpiling of juice) causes errors in the week-to-week variation in demand that do not show up in monthly or quarterly data. A large amount of error variance in the week-to-week purchasing patterns gets averaged out in a coarser granularity. On the other hand, extant research (e.g., Christen et al. 1997) indicates that using aggregated data to estimate demand models can also lead to bias in parameter inference depending on the type of model specification. The aforementioned vignette emphasizes the link between the choices of model specification and the level of granularity for demand analysis.

The analyst decides to use extant model selection tools to select the model with the best in-sample fit (e.g., smallest Bayesian information criterion [BIC]) at each level of granularity, and then uses the chosen models to forecast sales for the upcoming year. He finds that the out-of-sample sales forecasts for the upcoming year vary greatly with the choice of granularity, and the accompanying model conditional on the granularity. This result highlights two key notable issues. First, a comparison of model fit across granularity-model pairs is necessary to determine "the best among the best." Traditional model selection tools (e.g., Akaike information criterion [AIC], BIC, deviance information criterion), however, cannot be used for this task, as they are designed to select the best-fitted model *conditional* on the chosen granularity (i.e., for a fixed given data set), as per the panel on the left-hand side of Figure 1. These model selection tools (like BIC) vary with not only the in-sample fit but also other elements of the data set (e.g., the number of observations). Thus, they cannot be used to *jointly* select the best-fitted granularity-model pair, as described in the panel on the right-hand side of in Figure 1. Second, when one selects a granularity-model pair, there is the strong possibility that either of the two choices—granularity choice and model choice conditional on the granularity—is misspecified. Hence, a selection tool should provide in-sample fit that is robust against their erroneous choices—granularity misselection and model misspecification conditional on the granularity. Extant research, however, has addressed single robustness—robustness to model choice conditional on the granularity (e.g., Wang et al. 2017). In this paper, we address both these issues by proposing a selection tool that determines the best-fitted in-sample granularity-model pair and provides "doubly" robust protection.[1]

Our granularity-model selection tool has three desirable properties:

1. *Comparability*: The tool is comparable not only across models conditional on a granularity (as in extant research) but also across granularities. Our tool is not affected by elements of the data set that vary across granularities (e.g., the number of observations).

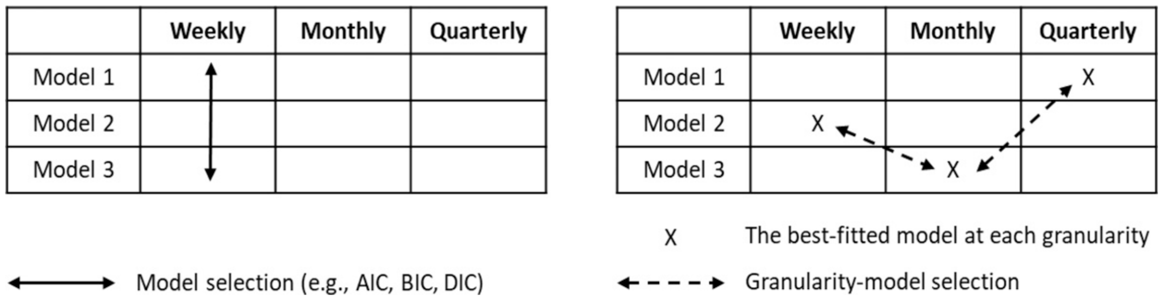
2. *Doubly fit*: The tool measures in-sample fit of a granularity-model pair, which can be used to identify the granularity-model pair that provides the best in-sample fit ("the best among the best").

3. *Doubly robust*: The tool provides in-sample fit that is robust to not only model selection conditional on the granularity (i.e., classic robustness) but also granularity choice. Specifically, the tool "corrects" for in-sample misfit due to not only outlying observations conditional on the data at the chosen granularity (i.e., classic robustness) but also the entire data set if it does not fit the chosen model.

Whereas these three properties are about in-sample fit, we suggest that our proposed method can also be more effective in improving out-of-sample prediction than existing (in-sample) model selection tools (e.g., BIC) and estimation methods (e.g., likelihood) for two reasons. First, since the first two properties (i.e., *comparability*, *doubly fit*) ensure the selection of the best-fitted in-sample granularity-model pair, if the out-of-sample period is stationary with the in-sample period, more accurate out-of-sample predictions will ensue. Thus, the manager in our vignette can improve the out-of-sample sales forecast by using the best-fitted pair identified by our approach. Second, since the third property (i.e., *doubly robust*) enables our method to correct for the in-sample misfit due to the choice of granularity and the accompanying model, it can reduce out-of-sample prediction errors (which we call *doubly robust prediction* hereafter). Thus, our method can reduce out-of-sample prediction errors more so than extant estimation methods (e.g., likelihood [L], weighted likelihood [WL]), which are either not robust or singly robust.

To discuss how our proposed approach relates to, nests, as well as extends extant (in-sample) estimation methods and satisfies the desired properties, we provide a "modeling tree" in Figure 2. As shown in the first level of the tree, unlike the standard likelihood method, which assigns the same weight to all observations, our

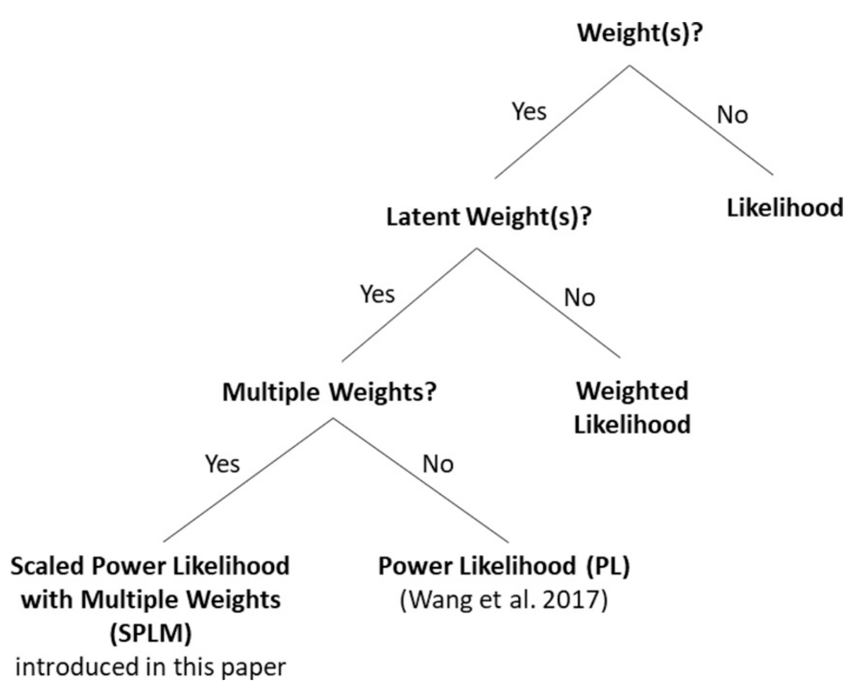
Figure 1. Comparison Between Model Selection and Granularity-Model Selection



approach builds on extant robust estimation methods (e.g., weighted likelihood) and down-weights observations that are discrepant from an assumed model. Next, as shown in the second level of the tree, unlike the weighted likelihood, which determines weights *prior* to estimation (e.g., weights set as a function of the inverse of the absolute residuals), our approach builds on the power likelihood (PL, hereafter), which allows for *latent* weights that are estimated jointly with the model parameters (e.g., Wang et al. 2017). Specifically, the PL raises each observation-specific likelihood term to a latent weight, and the posterior distribution of the weights is inferred from data. The trade-off between the prior on the weights and the likelihood automatically down-weights outlying observations, which are highly discrepant from an assumed model (i.e., those having extremely low likelihoods due to the in-sample misfit between themselves and the assumed model).

Note that PL still assumes that the granularity is prespecified and thus does not satisfy the three properties noted earlier. We extend the PL framework to the context where both granularity and model specification can vary. We briefly discuss the intuition behind our extension here, and the next section contains the details. As shown in the third level of the tree in Figure 2, we include two types of latent weights. One set of weights identifies the degree of in-sample fit between an entire data set and a model (i.e., the in-sample fit of a granularity-model pair), satisfying the *doubly fit* property. The other set of weights measures the degree of in-sample fit between an individual observation and the model conditional on the entire data set (i.e., conditional on the granularity). Our method automatically down-weights not only the latter set of weights for individual outlying observations conditional on the granularity (as PL does) but also

Figure 2. Comparison Between the Scaled Power Likelihood with Multiple Weights (SPLM) and Extant Approaches



the former set of weights if the entire data set does not fit the chosen model. Hence, the use of these two types of weights enables our method to correct for in-sample misfit due to not only individual outlying observations conditional on the chosen granularity (one aspect of robustness) but also the entire data set if it does not fit the chosen model (thus, the *doubly robust* property). Finally, to make the in-sample fit comparable across granularity-model pairs, we scale the likelihood (thus, the *comparability* property). To describe its multiple components exactly, we denote our method as the scaled power likelihood with multiple weights (SPLM, hereafter).

To assess the properties, performance, and limitations of SPLM, we conduct a large-scale simulation study and provide three key findings. First, SPLM performs better in identifying the best-fitted granularity-model pair and in providing doubly robust prediction than extant approaches do. It is due to the fact that SPLM scales the likelihood and relaxes the weight constraints of the other extant methods, as described in Figure 2. Second, we show that conducting in-sample model estimation at the SPLM-chosen granularity (even if it is coarser than the most granular data available) improves out-of-sample prediction accuracy. We find this result particularly important because it cautions against the aforementioned common practice of "tautologically" conducting in-sample estimation at the finest granularity. Lastly, even when the managerial goal is to predict sales at a certain fixed granularity, conducting in-sample estimation at the SPLM-chosen granularity (even if it differs from the managerially desired one) improves out-of-sample prediction accuracy. This result also indicates that SPLM can be applied even when researchers at a firm have conflicting prediction goals and hence need sales forecasts at different granularities. Specifically, those researchers can improve sales forecasts at each of their desired granularities by performing in-sample estimation at the SPLM-chosen granularity.

Following the simulation study, we apply SPLM as well as extant approaches to estimate a demand model and to perform out-of-sample sales forecasting using a Nielsen scanner panel data set that contains store-level sales and marketing actions (i.e., price, feature advertising). We show with this secondary data set that SPLM improves sales forecasts due to its ability (1) to select the best-fitted granularity-model pair and (2) to conduct doubly robust prediction.

The remainder of this paper proceeds as follows. In Section 2, we formally define SPLM and discuss its use as a doubly robust granularity-model selection tool. In Section 3, simulation studies show that SPLM is more effective (statistically more powerful) as a tool for granularity-model selection and doubly robust prediction than other extant approaches. In Section 4, we report an empirical study that explores the extent to which SPLM affects out-of-sample sales forecasting performance. Finally, Section 5 concludes with limitations and future research directions.

2. Scaled Power Likelihood with Multiple Weights (SPLM)

Our proposed granularity-model pair selection tool builds on a singly robust method recently introduced in the statistics and machine learning literature known as the *power likelihood* (PL) (e.g., Wang et al. 2017, Miller and Dunson 2019). In this section, we begin with the technical definition of the PL, discuss its benefits and limitations, as well as our generalization. To help with exposition, we use the context of an analyst trying to forecast store-level sales for the next year using a store-level panel data set for the current year. This context matches the vignette described in the introduction and our empirical application.

Consider a repeated measures (panel) data set of $n_\theta = \sum_{i=1}^{I} T_{\theta,i}$ observations of sales ($sales_{\theta,it}$) for store $i = 1, 2, \ldots, I$ at time $t = 1, 2, \ldots, T_{\theta,i}$ at granularity θ (e.g., month). Suppose also that $sales_{\theta,it}$ follows density $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_{\theta,i}})$, where $\boldsymbol{\beta_{\theta,i}}$ is a heterogeneous parameter vector at granularity θ that follows density $p(\boldsymbol{\beta_{\theta,i}} \mid \boldsymbol{\beta_\theta})$. Note that the density $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_{\theta,i}})$ depends on granularity θ, indicating that different models may be used at different granularities, as described in the panel on the right-hand side of Figure 1. The term, PL_θ, in Equation (1) is defined by raising each marginal likelihood term, $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta}) = \int p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_{\theta,i}})\ p(\boldsymbol{\beta_{\theta,i}} \mid \boldsymbol{\beta_\theta}) d\boldsymbol{\beta_{\theta,i}}$, to an observation-specific *latent* weight $w_{\theta,it} > 0$:

$$PL_\theta = \prod_{i=1}^{I} \prod_{t=1}^{T_{\theta,i}} p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})^{w_{\theta,it}} = \prod_{it=1}^{n_\theta} p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})^{w_{\theta,it}} = \prod_{it=1}^{n_\theta} \left[\int p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_{\theta,i}}) p(\boldsymbol{\beta_{\theta,i}} \mid \boldsymbol{\beta_\theta}) d\boldsymbol{\beta_{\theta,i}} \right]^{w_{\theta,it}}. \quad (1)$$

As described later, we set the prior mode of $w_{\theta,it}$ at 1 so that the standard likelihood is the default.

To extend the PL in Equation (1) to a fully Bayesian framework, Equation (2) provides the joint distribution of data ($\boldsymbol{sales_\theta}$), model parameters ($\boldsymbol{\beta_\theta}$), and latent weights ($\boldsymbol{w_\theta}$):

$$\log p_\theta(\boldsymbol{sales_\theta}, \boldsymbol{\beta_\theta}, \boldsymbol{w_\theta}) \propto \log p(\boldsymbol{\beta_\theta}) + \log p(\boldsymbol{w_\theta}) + \log PL_\theta = \log p(\boldsymbol{\beta_\theta}) + \sum_{it=1}^{n_\theta} \log p(w_{\theta,it}) + \sum_{it=1}^{n_\theta} w_{\theta,it} \cdot \log p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta}). \quad (2)$$

For model identification, we impose a prior on $w_{\theta,it}$, $p(w_{\theta,it})$, that is an increasing function of $w_{\theta,it}$ (e.g., beta, Dirichlet). Thus, for an observation that has a negative log likelihood, by lowering $w_{\theta,it}$ from 1 (the prior mode) toward 0, we obtain a loss in the second term, $\log p(w_{\theta,it})$, but a gain in the third term, $w_{\theta,it} \cdot \log p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})$. Since the gain exceeds the loss for observations that have a very negative log likelihood (i.e., do not fit a chosen model conditional on the granularity θ), Bayesian inference using PL automatically down-weights (i.e., lowers $w_{\theta,it}$ for) those outlying observations (e.g., Wang et al. 2017).[2] Therefore, PL allows for more flexibility in estimation than the weighted likelihood, which determines the weights prior to estimation (e.g., McCarthy and Jensen 2016, Wang et al. 2017).

We build our approach on the PL but make the following two modifications to satisfy the three properties described earlier (i.e., *comparability*, *doubly fit*, and *doubly robust*). As the first modification, we make $\log PL_\theta$ in Equation (2) satisfy the *comparability* property—that is, be comparable across granularities—by the following two steps. Since $\log PL_\theta$ in Equation (2) sums up n_θ log likelihood terms, we divide $\log PL_\theta$ by n_θ as the first step, which is equivalent to the log of the fractional likelihood that is commonly used to assess the goodness of fit per observation (e.g., Ruiz et al. 2020). However, the fractional likelihood still depends on the level of granularity θ, because each likelihood term at granularity θ, which we denote, as earlier, by $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})$, varies with θ. Thus, as the second step, we scale each likelihood term at granularity θ to the finest granularity while ensuring that the scaling does not create additional in-sample misfit. Specifically, we (a) scale the dependent variable (not covariates) at granularity θ (here, $sales_{\theta,it}$) to the finest granularity and then (b) replace the dependent variable in each likelihood term at granularity θ with this scaled one. Note that, to perform (b) without creating additional in-sample misfit, the scaling in (a) should preserve the functional form of the distribution of the dependent variable.

To make this second step more tangible, let us revisit our motivating example and assume that the analyst applies a monthly log-log demand model (where the dependent variable, $sales_{\mathrm{m},it}$, with the subscript "m" for month, follows a log-normal distribution) as in Sections 3 and 4. The analyst in this case can scale each monthly likelihood term to the finest granularity (here, weekly) without creating additional in-sample misfit by (a) simply dividing monthly sales by the number of weeks in a given month and then (b) replacing monthly sales in each likelihood term with the scaled one (i.e., $\frac{sales_{\mathrm{m},it}}{\#weeks_{\mathrm{m},t}}$). It is appropriate to (b) replace monthly sales in each likelihood term with the scaled one, since the scaling in (a) preserves the functional form of the distribution (i.e., the scaled one follows a log-normal distribution). Whereas we have assumed a log-log demand model, the proposed scaling can be used for other types of models if it preserves the functional form of the distribution of the dependent variable. Two commonly used models that satisfy this condition include a linear model (where a dependent variable follows a normal distribution) and a log-linear model (where a dependent variable follows a log-normal distribution). In contrast, this simple divisible scaling cannot be used for a Poisson (count) model. The scaled dependent variable will likely have noninteger values and thus cannot be assumed to follow a Poisson distribution. We discuss situations where the proposed scaling step of SPLM cannot be simply accomplished as an area for future research in our concluding section.

Formally, we make $\log PL_\theta$ in Equation (2) satisfy the *comparability* property by dividing it with n_θ and then by replacing the likelihood term $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})$ for each observation ("*it*" subscript) with the scaled term $p_{1(\theta)}(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})$ (where subscript "1" indicates the finest granularity) that transforms $p_\theta(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})$ to the finest granularity. Thus, the log of the scaled power likelihood, $\log SPL_{1(\theta)}$, is

$$\log SPL_{1(\theta)} = \frac{1}{n_\theta} \cdot \sum_{it=1}^{n_\theta} w_{\theta,it} \cdot \log p_{1(\theta)}(sales_{\theta,it} \mid \boldsymbol{\beta_\theta}). \tag{3}$$

As the second modification, in order to make our method satisfy the *doubly fit* and *doubly robust* properties, we decompose the single latent weight $w_{\theta,it}$ in Equation (3) into two types of latent weights—namely, a cross-granularity weight (i.e., $G_\theta = \frac{\sum_{it=1}^{n_\theta} w_{\theta,it}}{n_\theta}$) and within-granularity weight (i.e., $f_{\theta,it} = \frac{w_{\theta,it}}{\sum_{it=1}^{n_\theta} w_{\theta,it}}$). Thus, the log of the scaled power likelihood with multiple weights, $\log SPLM_{1(\theta)}$, is expressed as

$$\begin{aligned} \log SPLM_{1(\theta)} &= \frac{1}{n_\theta} \cdot \sum_{it=1}^{n_\theta} \mathrm{n}_\theta \cdot \frac{\sum_{it=1}^{n_\theta} w_{\theta,it}}{n_\theta} \cdot \frac{w_{\theta,it}}{\sum_{it=1}^{n_\theta} w_{\theta,it}} \\ &\quad \cdot \log p_{1(\theta)}(sales_{\theta,it} \mid \boldsymbol{\beta_\theta}) \\ &= \sum_{it=1}^{n_\theta} G_\theta \cdot f_{\theta,it} \cdot \log p_{1(\theta)}(sales_{\theta,it} \mid \boldsymbol{\beta_\theta}). \end{aligned} \tag{4}$$

We extend SPLM in Equation (4) to a fully Bayesian framework in Section A1 of Online Appendix A.

Our method, because of the use of G_θ (cross granularity, column 1 in Table 1) and $f_{\theta,it}$ (within granularity, column 2 in Table 1), allows for more flexibility than extant methods. Note that G_θ and $f_{\theta,it}$ are not allowed to be fully flexible—that is, $0 < G_\theta \le 1$ and $\sum_{it=1}^{n_\theta} f_{\theta,it} = 1$—by definition. We set G_θ and $f_{\theta,it}$ to have a prior mode at 1 and $1/n_\theta$, respectively (in order to set the standard likelihood approach, which assumes that both granularity and model are correctly specified, as the default). Also, in order to lower G_θ and $f_{\theta,it}$ from the default values of 1 and $1/n_\theta$ only when the assumed model does not fit the entire data

Table 1. Comparison Between Our Method (SPLM) and the Extant Approaches

	Cross-granularity weight (G_θ)	Within-granularity weight ($f_{\theta,it}$)
Likelihood (L)	1	$1/n_\theta$
Weighted likelihood (WL)	1	A function of the inverse of the absolute residuals
Power likelihood (PL)	1	Latent
Our method (SPLM)	Latent	Latent

set and individual observations, respectively, we set the priors on G_θ and $f_{\theta,it}$ to be an increasing function of G_θ and $f_{\theta,it}$ (similar to the PL). By choosing priors that meet the aforementioned conditions, our method can identify two sets of weights[3] and satisfy the *doubly fit* and *doubly robust* properties, which we explain next. Whereas there are many candidates for the prior distributions, we assume that G_θ has a spike-and-slab beta prior (with a spike at 1 as mentioned) and the vector of $f_{\theta,it}$ has a symmetric Dirichlet prior (centered at $1/n_\theta$). Our prior choices are explained in detail in Section A1 of Online Appendix A.

Our method satisfies the *doubly fit* property by specifying G_θ in addition to $f_{\theta,it}$. That is, one can identify the best-fitted granularity-model pair by using G_θ as a granularity-model selection criterion. Here is why. As demonstrated in Section A1 of Online Appendix A (especially Equation (A7)), there is a trade-off between the prior on G_θ, which is an increasing function of G_θ, and the product of G_θ and the "scaled" in-sample fit of a granularity-model pair—that is, $\frac{\sum_{it=1}^{n_\theta} \log p_{1(\theta)}(sales_{\theta,it} \mid \boldsymbol{\beta_\theta})}{n_\theta}$. This trade-off allows our method to lower G_θ from 1 (the prior mode) toward 0 when the scaled in-sample fit of a granularity-model pair is poor, as demonstrated in Section A1 of Online Appendix A (especially Equations (A8)–(A9)). Since G_θ represents the scaled in-sample fit, which is comparable across granularity-model pairs,[4] researchers can identify the best-fitted granularity-model pair (among a set of granularity-model candidates) by selecting the one that maximizes G_θ. Furthermore, it is easy to interpret G_θ; G_θ will be (almost) 1 if the analysis is conducted at the perfectly fitted granularity-model pair and smaller than 1 if it is not. Therefore, G_θ is the main heuristic selection tool and contribution of this research. We demonstrate its properties as a granularity-model selection criterion in Sections 3 and 4.

Our method also satisfies the *doubly robust* property—that is, it is robust to granularity selection and model selection conditional on the granularity. Cross-granularity robustness and within-granularity robustness are enabled by G_θ and $f_{\theta,it}$, respectively. First, as seen in Equation (4), SPLM raises the standard likelihood of the entire data set to G_θ on average. In other words, it "flattens the likelihood" of the entire data set by the amount of in-sample misfit (between data at that granularity and the accompanying model), providing in-sample robustness *across* granularities. Second, if we fix G_θ, then our method goes back to PL; thus, $f_{\theta,it}$ works in the same way as $w_{\theta,it}$ in the PL, providing in-sample robustness across models *conditional* on the granularity.

We also propose that our method will offer better out-of-sample forecasts than extant methods if the out-of-sample period is stationary with respect to the in-sample period. First, note that our proposed measure (G_θ) represents the scaled in-sample fit, which is comparable across granularity-model pairs. Thus, we can improve the out-of-sample forecasts by using the pair that maximizes G_θ (indicating that the pair has the best in-sample fit). Second, in-sample double robustness, which corrects for in-sample misfit due to both granularity and model selections, leads to "doubly robust prediction," as explained in the introduction. Thus, our method will be more effective in correcting for out-of-sample prediction errors than extant (in-sample) estimation methods, which are either not robust (e.g., likelihood) or singly robust (e.g., weighted likelihood, PL). We will next assess the dual benefits of SPLM in improving out-of-sample prediction—(1) granularity-model selection and (2) doubly robust prediction—through simulation and an empirical exercise with real data.

3. Simulation

3.1. Simulation Objectives

We designed a large-scale simulation study to assess the viability of our proposed method (SPLM) along three major dimensions: (1) to demonstrate that conducting a demand analysis at the SPLM-chosen granularity (even if the SPLM chosen one is coarser than the most granular data available) can provide better in-sample fit and out-of-sample predictions, suggesting that it is not necessary to always use the most granular data in practice; (2) relatedly, to assess the performance of SPLM in selecting the best-fitted granularity-model pair in contrast to the aforementioned extant methods (L, WL, PL); and (3) to evaluate SPLM's ability to improve predictive accuracy and yield doubly robust predictions (predictions that are robust against granularity-model pair misselection) in comparison with the same three extant methods. Thus, we hope to demonstrate that SPLM is valuable to researchers for granularity-model selection, as well as for predictive accuracy and robustness.

3.2. Simulation Design

We begin with an overview of the design and then explain the details in the next few paragraphs. To provide support for our objectives of the simulation, we created different granularity-model pairs and conducted (in-sample) model estimation and (out-of-sample) prediction using SPLM and the three extant methods. For ease of explanation, we continue with the example of a brand manager who is using store-level sales data for the current year to predict demand for the upcoming year.

Figure 3 lays out the data generation and modeling steps that we pursued in the simulation study. We simulated monthly demand data from a chosen data generating process (DGP) using a 2 (sample size) × 2 (SNR) factorial design, as in Box 1 of Figure 3. For each of the four cells, we generated 100 monthly level data sets, yielding a total of 400 data sets. We then created quarterly data and weekly data, via aggregating and disaggregating the monthly data, respectively, as in Box 2 of Figure 3. We did so in a way that analyzing monthly data will provide better in-sample fit than quarterly and weekly data (first objective). Since we hope to demonstrate that SPLM identifies the best-fitted granularity (rather than to uncover new and interesting underlying mechanisms that determine the best-fitted granularity), we selected aggregation and disaggregation processes that are likely to be observed in empirical data and utilized by researchers. We explain the data generation process in Sections 3.2.1 and 3.2.2. Then, for each of the 400 original 2 × 2 data set triples (quarterly, monthly, weekly) from Boxes 1 and 2 of Figure 3, we fit 12 different model specifications at the three granularities using four different methods (L, WL, PL, SPLM), as given in Box 3(a) of Figure 3. We explain the models and methods in Sections 3.2.3 and 3.2.4, respectively. Lastly, we compared the results across methods to demonstrate SPLM's superior ability to select the best-fitting granularity-model pair (second objective) and to provide improved predictive accuracy and doubly robust predictions (third objective), as in Box 3(b) of Figure 3.

3.2.1. Monthly Data (Box 1 of Figure 3). We created monthly demand data for a period of two years (i.e., in-sample data for the current year and out-of-sample data for the upcoming year) following three steps. First, we generated monthly covariates (price and advertising) for a brand at store i in month t (= 1, 2, . . . , 24):[5]

$$\log(price_{\theta=\mathrm{m},it}) \sim N(\mu_{\mathrm{price,m}}, \sigma^2_{\mathrm{price,m}}); \tag{5}$$

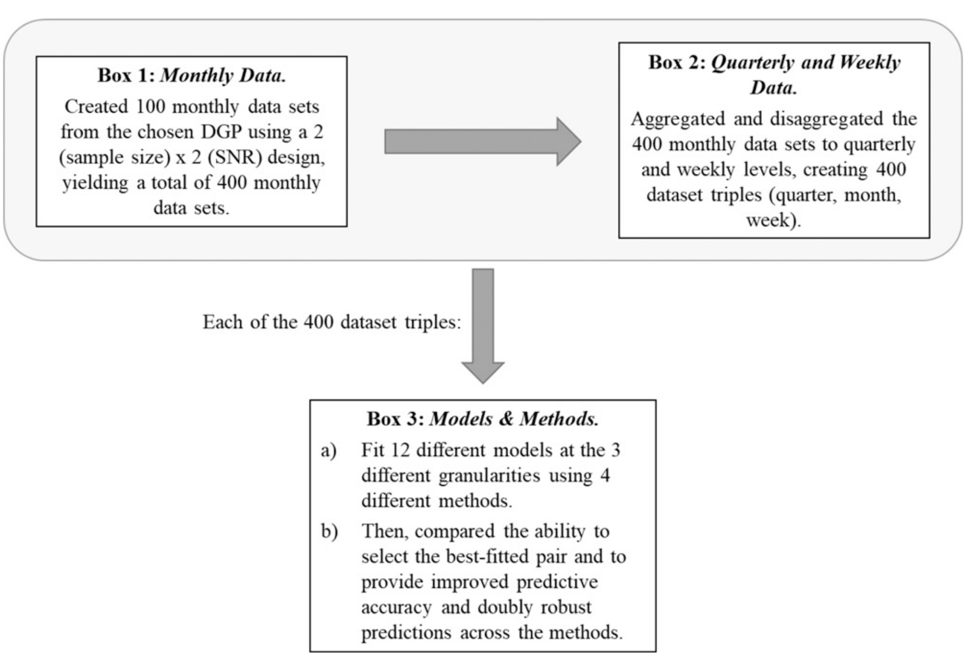
$$\log(ad_{\theta=\mathrm{m},it}) \sim N(\mu_{\mathrm{ad,m}}, \sigma^2_{\mathrm{ad,m}}). \tag{6}$$

Next, we generated brand-specific unit sales at store i in month t using a log-log model:

$$\begin{aligned}\log(sales_{\mathrm{m},it}) &= \mu_{\mathrm{m},it} + \varepsilon_{\mathrm{m},it}\\ &= \alpha_{\mathrm{m},i} + \beta_{\mathrm{m},i}\log \mathrm{price}_{\mathrm{m},it}\\ &\quad + \gamma_{\mathrm{m},i}\log \mathrm{ad}_{\mathrm{m},it} + \delta_{\mathrm{m},t} + \varepsilon_{\mathrm{m},it},\end{aligned} \tag{7}$$

with three heterogeneous parameters—namely, store-level baseline intercept $\alpha_{\mathrm{m},i}$, store-level price elasticity $\beta_{\mathrm{m},i}$, and store-level advertising elasticity $\gamma_{\mathrm{m},i}$— and 11 monthly fixed effects $\delta_{\mathrm{m},t}$, and an error term $\varepsilon_{\mathrm{m},it}$. The heterogeneous parameters ($\alpha_{\mathrm{m},i}$, $\beta_{\mathrm{m},i}$, $\gamma_{\mathrm{m},i}$) and error term $\varepsilon_{\mathrm{m},it}$ were assumed to follow independent Normal distributions—that is, $\alpha_{\mathrm{m},i} \sim N(\alpha_{\mathrm{m}}, \sigma^2_{\alpha,\mathrm{m}})$, $\beta_{\mathrm{m},i} \sim N(\beta_{\mathrm{m}}, \sigma^2_{\beta,\mathrm{m}})$, $\gamma_{\mathrm{m},i} \sim N(\gamma_{\mathrm{m}}, \sigma^2_{\gamma,\mathrm{m}})$, $\varepsilon_{\mathrm{m},it} \sim N(0, \sigma^2_{\varepsilon,\mathrm{m}})$.[6]

Figure 3. Simulation Design



Last, we used the first half of the data for in-sample estimation and the second half of the data for out-of-sample prediction.

When generating monthly data sets, we systematically varied the two factors—(a) sample size and (b) signal-to-noise ratio (SNR)—that have been shown to be important drivers for the performance of the extant methods in model selection and singly robust prediction (e.g., Mao 2013). First, we set the number of stores I at two levels: 25 (small) and 100 (large). Second, we varied SNR—that is, $\frac{\sigma^2(\mu_{m,it})}{\sigma^2(\varepsilon_{m,it})}$—at two levels: 1/9 (small) and 9 (large), which is equivalent to changing the average relative error variance—that is, $\frac{\sigma^2(\varepsilon_{m,it})}{\sigma^2(\mu_{m,it})+\sigma^2(\varepsilon_{m,it})}$ over 0.10 (small) and 0.90 (large).

3.2.2. Quarterly and Weekly Data (Aggregation and Disaggregation) (Box 2 of Figure 3). We aggregated and disaggregated the 400 monthly data sets to quarterly and weekly levels, respectively. Among the many possible aggregation and disaggregation processes, we selected ones that are likely to be observed in empirical data and utilized by researchers. (We explain our aggregation and disaggregation processes and their effects on the chosen DGP in the next few paragraphs.) Importantly, our aggregation process, linear aggregation as is commonly used, introduced bias but with no added stochastic error in the chosen DGP. In turn, and in contrast, our disaggregation process introduced random noise with no bias. This way of creating quarterly and weekly data allowed us to compare the performance of SPLM with that of extant methods under conditions when there is either bias or stochastic error introduced in the chosen DGP. As we generated two years of sales data, the quarterly, monthly, and weekly data sets have $n_q = 8 \cdot I$, $n_m = 24 \cdot I$, and $n_w = 104 \cdot I$ observations, where I is the number of stores, respectively.

(1) Quarterly Data (Aggregation)
We chose to linearly aggregate the monthly data to a quarterly level, as is commonly done in research and practice (e.g., Tellis and Franses 2006, Ivancic et al. 2011). Specifically, we summed up the monthly number of advertisements and monthly sales and took a simple average of monthly prices.[7] Since the chosen monthly DGP (log-log demand model) in Equation (7) for our simulated data was a nonlinear function, its linear aggregation introduced bias with no stochastic error (e.g., Christen et al. 1997). Thus, we expected that if the log-log model in Equation (7) is applied to both monthly and quarterly data, then the in-sample fit and out-of-sample prediction accuracy would be higher at the monthly level than at the quarterly level.

(2) Weekly Data (Disaggregation)
To create weekly data, we assumed that while each store sets price and feature advertising every month (as given in Equations (5) and (6)), its sales vary across weeks. This assumption is plausible because even if the marketing mix is determined at a monthly level, demand could vary across weeks due to random shocks. Formally, given monthly data (price, advertising, sales) of store i for month j, we assumed that price was the same across all weeks in a month and divided monthly advertising into equal weekly amounts. We then distributed $sales_{m,ij}$ for a given month into four weeks as follows:[8]

$$sales_{w,it} = sales_{m,ij} \cdot share_{w,it}, \tag{8}$$

where $\sum_{t=1}^{4} share_{w,it} = 1$. The disaggregation vector $\boldsymbol{share_{w,i}} = (share_{w,i1}, share_{w,i2}, share_{w,i3}, share_{w,i4})$ was drawn from a symmetric Dirichlet distribution, to add random, but unbiased, stochastic noise when generating weekly data from monthly data.[9] We used the Dirichlet distribution with a small concentration parameter (here, 1) to increase the noise in the week-to-week variation.

Note that store-level price and advertising were assumed not to vary across weeks within a month and monthly sales were stochastically distributed to a weekly level. Thus, if the brand manager is given these weekly data and linearly aggregates the data to a monthly level, then the noise will be smaller in the monthly (aggregated) data than in the weekly (most granular) data. Hence, we expected that even if we apply the model in Equation (7) to both weekly and monthly data, the in-sample fit and out-of-sample prediction accuracy would be higher when the analysis is performed using the latter.

3.2.3. Models (Box 3 of Figure 3). We assessed the benefit of SPLM in identifying the chosen granularity-model pair and correcting for granularity-model misselection due to both the assumed temporal granularity and the model specification. For showcasing the latter, we systematically varied three factors while estimating the demand model—(a) the functional form, (b) seasonal fixed effects, and (c) the level of heterogeneity. For (a) the functional form, we assumed three levels: log-log (chosen), log-linear, and linear. The other two factors were varied at two levels—namely, (b) seasonal fixed effects, which are present in the model (chosen) or absent in the model, and (c) the level of heterogeneity, which is present in the model (chosen) or absent in the model. Thus, there were 12 variations of the estimated demand model (3 × 2 × 2) as described in Box 3 of Figure 3.

3.2.4. Methods (Box 3 of Figure 3). We used the 400 data set triples to estimate model parameters for the 12 specifications using four methods (L, WL, PL, SPLM).[10] Note that the extant methods have more restrictive weight constraints than SPLM does, as described in Table 1.

3.3. Simulation Results

To compare in-sample estimation and out-of-sample prediction results across granularity-model pairs and the four methods (L, WL, PL, SPLM), we adopted the following two steps, as described in Box 3 of Figure 3. First, as described in Box 3(a) of Figure 3, for each of the data set triples, we fit the aforementioned 12 model variations at the three different granularities using the four different methods. For each of the four methods, we chose the best-fitted model (in-sample) *given* the granularity, and then compared the in-sample fit *across* granularities to select "the best of the best" granularity-model pair. This approach is consistent with the panel on the right-hand side of Figure 1. Next, as described in Box 3(b) of Figure 3, we assessed our SPLM method's ability to select the best-fitting pair and to provide improved predictive accuracy and double robustness in comparison with the extant methods (L, WL, PL).

3.3.1. Granularity-Model Selection. As noted earlier, the cross-granularity weight, G_θ, is our proposed metric to identify the granularity-model pair that provides the best in-sample fit. The first panel of Figure 4(a) demonstrates that G_θ is higher for the monthly (chosen DGP) data than for either the quarterly (aggregated) or weekly (disaggregated) data. Accordingly, the second panel of Figure 4(a) and the first panel of Figure 4(b) show lower in-sample misfit (measured using weighted absolute percentage error, WAPE, as defined in the note of Figure 4 and detailed in Online Appendix B) when monthly data are used and when G_θ is higher, respectively. The in-sample results are extended to out-of-sample forecasting, since the out-of-sample period is stationary with the in-sample period in this simulation study. Specifically, the third panel of Figure 4(a) and the second panel of Figure 4(b) demonstrate lower out-of-sample misfit (measured using WAPE) when monthly data are used and when G_θ is higher, respectively.

Hence, Figure 4 supports the first objective by providing evidence against the common belief that the most granular data available always attain the best in-sample fit and out-of-sample predictions. It also demonstrates that G_θ is sensitive to the magnitude of the in-sample misfit. To extend this result and support the second objective, we compare G_θ with extant methods in terms of the ability to identify in-sample misfit in the next few paragraphs.

Next, as noted earlier, we examined whether SPLM is more sensitive in identifying in-sample misfit and so performs better in selecting the granularity-model pair with the smallest in-sample misfit (i.e., the best-fitted pair) than extant methods (L, WL, PL). Note that in-sample misfit in the data set triples (that we generated) varied within a relatively small range, as this was not the primary purpose of the simulation. For example, the second panel of Figure 4(a) shows that the 95% credible interval of quarterly in-sample misfit is [52%, 58%]. To explore and assess SPLM's ability to identify the best-fitted pair in comparison with the extant methods over a wider range of "signal" (in-sample misfit), we supplemented our simulation with an additional 100 data set triples, where we varied the levels of signal, as described in Online Appendix C.

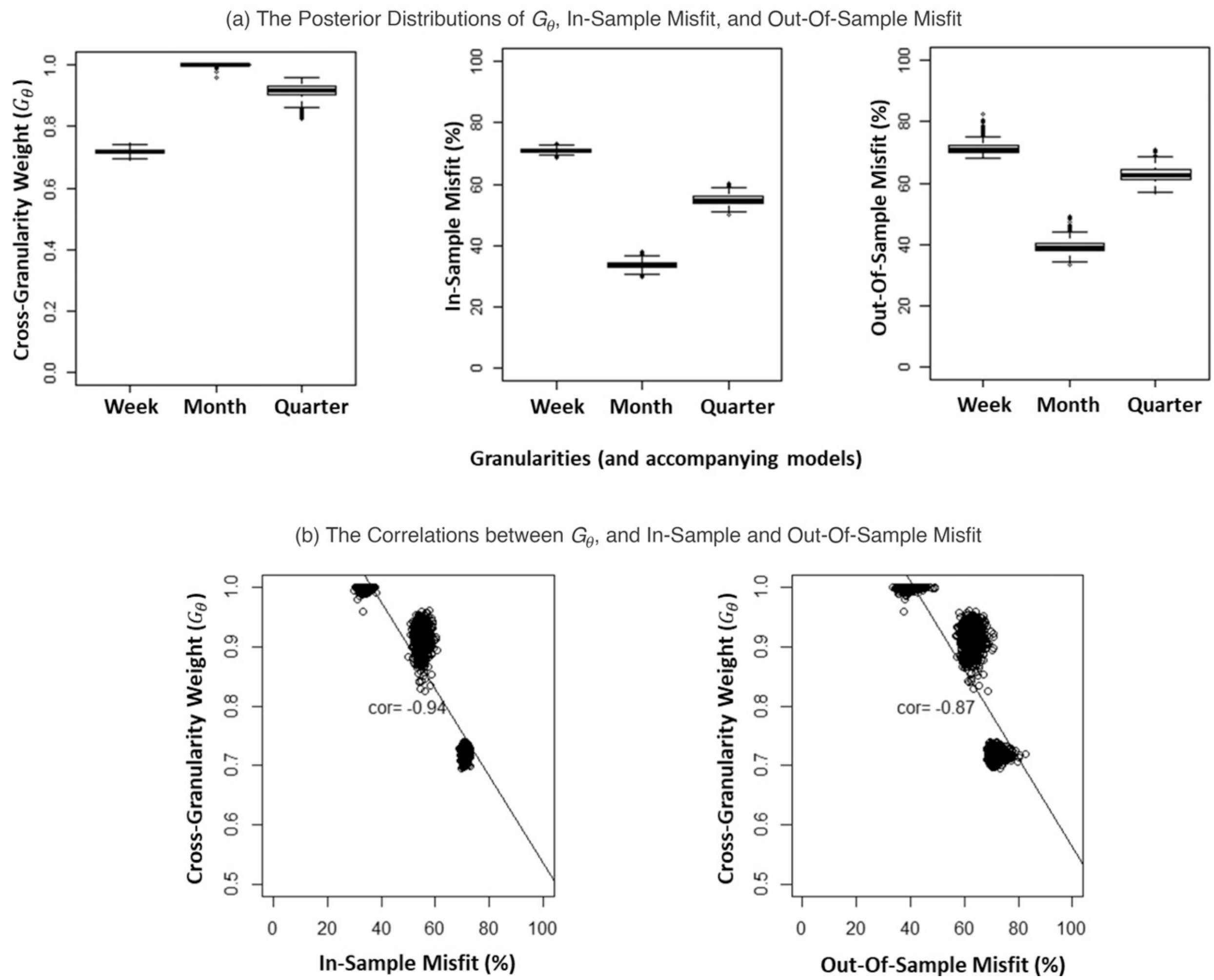
Parts (a) and (b) of Figure 5 contain the results. In Figure 5(a), the y-axis is the proportion of triples from which each of the four methods (L, WL, PL, SPLM) identifies the best-fitted granularity-model pair (in this case, the monthly DGP). The x-axis is a measure of in-sample misfit in a triple (in this case, the minimum of quarterly and weekly in-sample WAPE). The graph shows how well each method performs under varying levels of in-sample misfit. The x-axis and y-axis in Figure 5(b) are based on out-of-sample misfit (measured using out-of-sample WAPE) and out-of-sample prediction accuracy, respectively.

In Figure 5(a), for triples with small in-sample misfit (here, less than 40%), the value of the y-axis for our measure (G_θ of SPLM) is greater than that for extant in-sample measures, namely, the in-sample R^2 of L, WL, and PL. This indicates that when in-sample misfit is small, our measure (G_θ) is better in identifying the best-fitted granularity-model pair than the extant measures. Note that the value of the y-axis for the extant measures is zero, which indicates that extant measures always fail to select the best-fitted granularity-model pair.[11] When in-sample misfit is moderately large (here, between 40% and 60%), G_θ is still more likely to identify the best-fitted granularity-model pair than the extant measures. When the degree of in-sample misfit is sufficiently large (here, greater than 60%), all four measures can identify the best-fitted granularity-model pair.

Figure 5(b) shows the superiority of our measure (G_θ) for out-of-sample forecasting as well. It identifies the pair with the best predictions better than extant out-of-sample measures (the out-of-sample R^2 of L, WL, and PL) do. Whereas the out-of-sample R^2 is calculated using out-of-sample data (by definition), it is statistically less powerful in identifying the pair with the best predictions than G_θ, due to its lack of comparability across granularities.

Last, whereas we assumed earlier (in Figures 4 and 5) that a firm does not prespecify the data granularity at which out-of-sample predictions have to be made, this may not always be the case. Typically, if managers want to make predictions at a certain fixed granularity, then they estimate the model at that "desired granularity." For example, Mela et al. (1997) analyze the

Figure 4. Relationship Between Cross-Granularity Weight (G_θ), In-Sample Misfit, and Out-of-Sample Misfit Under Large Sample Size and High SNR



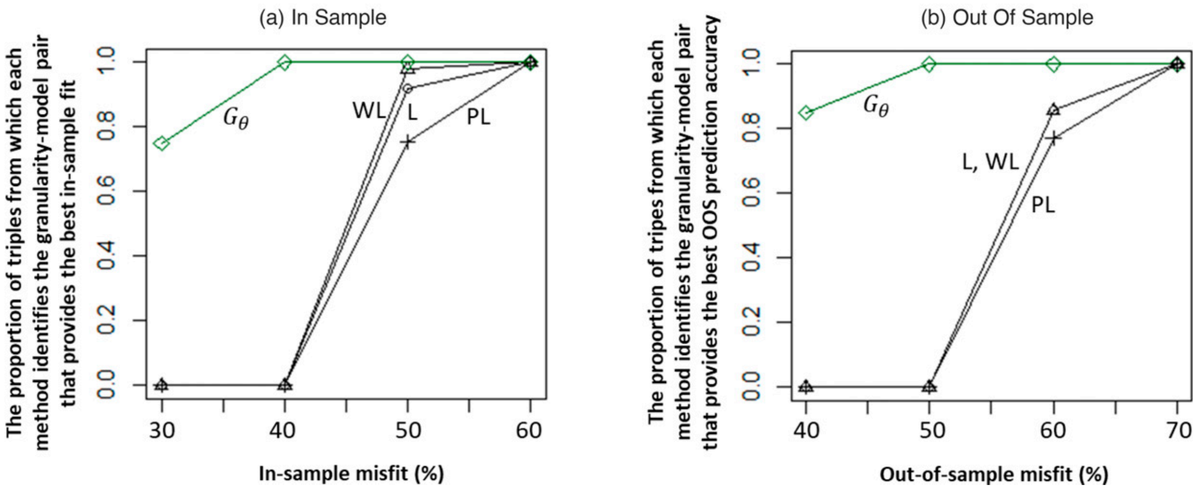
Notes. (1) We used weighted mean absolute percentage error (WAPE), which takes the sales-weighted average of the absolute percentage errors (APEs), to compare in-sample or out-of-sample misfit across granularities. We justified our use of WAPE in Online Appendix B. (2) Part (b) of the figure has three clusters, as it plots the results for the three granularities—week, month, and quarter.

effect of advertising on consumers' brand choice at a quarterly level, as their data provider is likely to make marketing decisions at that level.

Assume that a firm wanted to predict sales at a particular granularity (say, weekly, which in this case is the most granular). We applied the following three forecasting approaches for forecasting weekly sales. In the weekly approach, we estimated the demand model with the weekly data (i.e., desired granularity) and predicted sales at the weekly level. In the monthly approach, we estimated the model at the monthly level (i.e., SPLM-chosen granularity), predicted monthly sales, and then consistently disaggregated it to forecast sales at the weekly level. The definition and details of consistent disaggregation are in Online Appendix D. In the quarterly approach, we estimated the model using quarterly data, predicted quarterly sales, and then consistently disaggregated it to the weekly level. A similar procedure can be followed if the desired granularity is either monthly or quarterly.

Figure 6 shows that smaller out-of-sample prediction errors are obtained with the monthly approach (i.e., estimating the model at the SPLM-chosen granularity) than the other two approaches, regardless of the desired granularity and estimation method (i.e., M1 < W1, Q1 under L; M2 < W2, Q2 under WL; M3 < W3, Q3 under PL; M4 < W4, Q4 under SPLM). This result demonstrates that estimating and predicting at the SPLM-chosen granularity-model pair, even if that is not the managerially desired granularity (e.g., if the manager has to make weekly or quarterly predictions to set weekly or quarterly budgets) can improve predictions. This finding is highly relevant for managers who may erroneously believe that their granularity

Figure 5. (Color online) Comparison of G_θ of SPLM and Extant Measures in Terms of the Ability to Identify a Granularity-Model Pair That Provides the Best In-Sample Fit and Out-of-Sample Prediction Accuracy



Notes. (1) In part (a) of the figure, L, WL, and PL indicate the in-sample R^2 of L, WL, and PL. In part (b), L, WL, and PL indicate the out-of-sample R^2 of L, WL, and PL. (2) We measured in-sample and out-of-sample misfit with weighted mean absolute percentage error (WAPE), which takes the sales-weighted average of the absolute percentage errors (APEs), as explained in Online Appendix B.

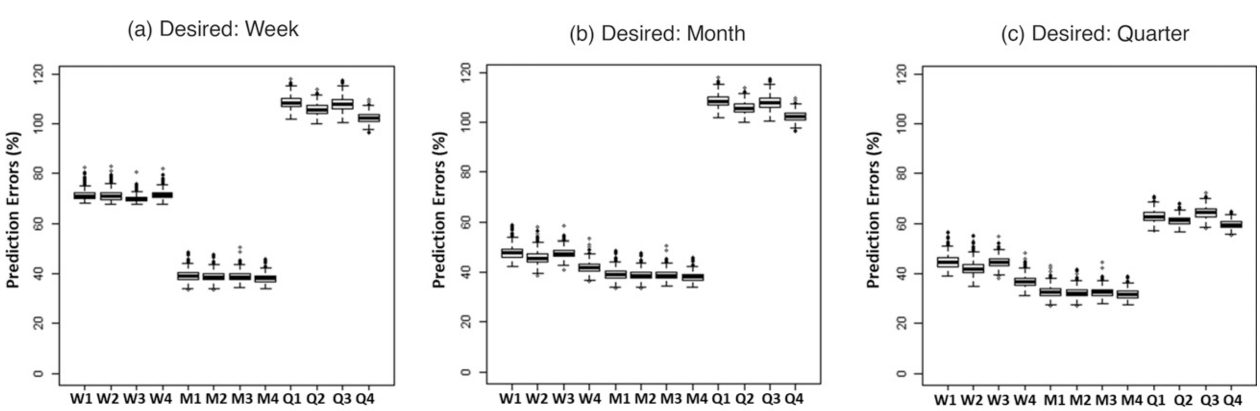
used for estimating the demand model must match with the one used for forecasting.

3.3.2. Doubly Robust Prediction.

We have shown that SPLM improves prediction accuracy by choosing the best-fitted granularity-model pair (as in Figure 4). We now demonstrate that SPLM provides predictions that are robust against both granularity and model misspecifications in comparison with those from extant methods (L, WL, PL), supporting our third objective.

First, to assess SPLM's robustness to granularity misspecification, assume that the manager erroneously selected a weekly or quarterly granularity and chose the best-fitted model given this misspecified granularity. Figure 6 demonstrates that SPLM reduces prediction errors due to the granularity misspecification by a greater amount than do extant methods (i.e., W4 < W1, W2, W3; Q4 < Q1, Q2, Q3 in parts (b) and (c) of Figure 6). SPLM accomplishes this via down-weighting observations that are discrepant from a chosen model (as explained in Section 2). However, if the out-of-sample data at the desired granularity are too noisy (here, weekly data), then SPLM will provide prediction accuracy similar to that under the extant methods (e.g., W4 is similar to W1–W3 in Figure 6(a)). Note that this

Figure 6. Posterior Distributions of Percentage Errors of Predicting Sales at Different Desired Granularities Under Large Sample Size and High SNR



Notes. (1) The labels in the x-axis indicate a forecasting approach and estimation method pair. From left to right, they are (W1) week + L, (W2) week + WL, (W3) week + PL, (W4) week + SPLM, (M1) month + L, (M2) month + WL, (M3) month + PL, (M4) month + SPLM, (Q1) quarter + L, (Q2) quarter + WL, (Q3) quarter + PL, and (Q4) quarter + SPLM. (2) We measured out-of-sample prediction errors with weighted mean absolute percentage error (WAPE), which takes the sales-weighted average of the absolute percentage errors (APEs), as explained in Online Appendix B.

boundary condition (i.e., the out-of-sample data should not be too noisy) may not heavily restrict our application of SPLM in practice, as much empirical research will satisfy this condition, since it is necessary for reliable prediction.

Similarly, researchers may (unknowingly) select a misspecified model given the granularity. For example, in our simulation study, researchers may select either a linear or log-linear demand model, estimate it using monthly data, and then predict sales at their desired granularity. Figure 7 demonstrates that SPLM reduces prediction errors due to model misspecification more so than extant methods regardless of the desired granularity. Since SPLM uses more flexible weights than the extant methods (as described in Table 1), it can provide more "protection" against model misspecification than the extant ones. Figure 7 also shows that researchers may well get even bigger benefits from using SPLM when their assumed model is more discrepant from the data. For example, SPLM reduces out-of-sample prediction errors by a greater amount for the linear misspecification (as in Figure 7(b)) than for the log-linear misspecification (as in Figure 7(a)).

4. Secondary Data Application

4.1. Objectives

We provide evidence that the benefits of our proposed method (SPLM) can be realized within secondary data contexts as well. In particular, we show the following: (1) SPLM is more discriminating in selecting a granularity-model pair that leads to better in-sample fit and out-of-sample predictions as compared with the

Figure 7. Comparison of the Posterior Distributions of Prediction Errors Across the Estimation Methods When Monthly Data Are Estimated with a Model in a Misspecified Functional Form Under Large Sample Size and High SNR

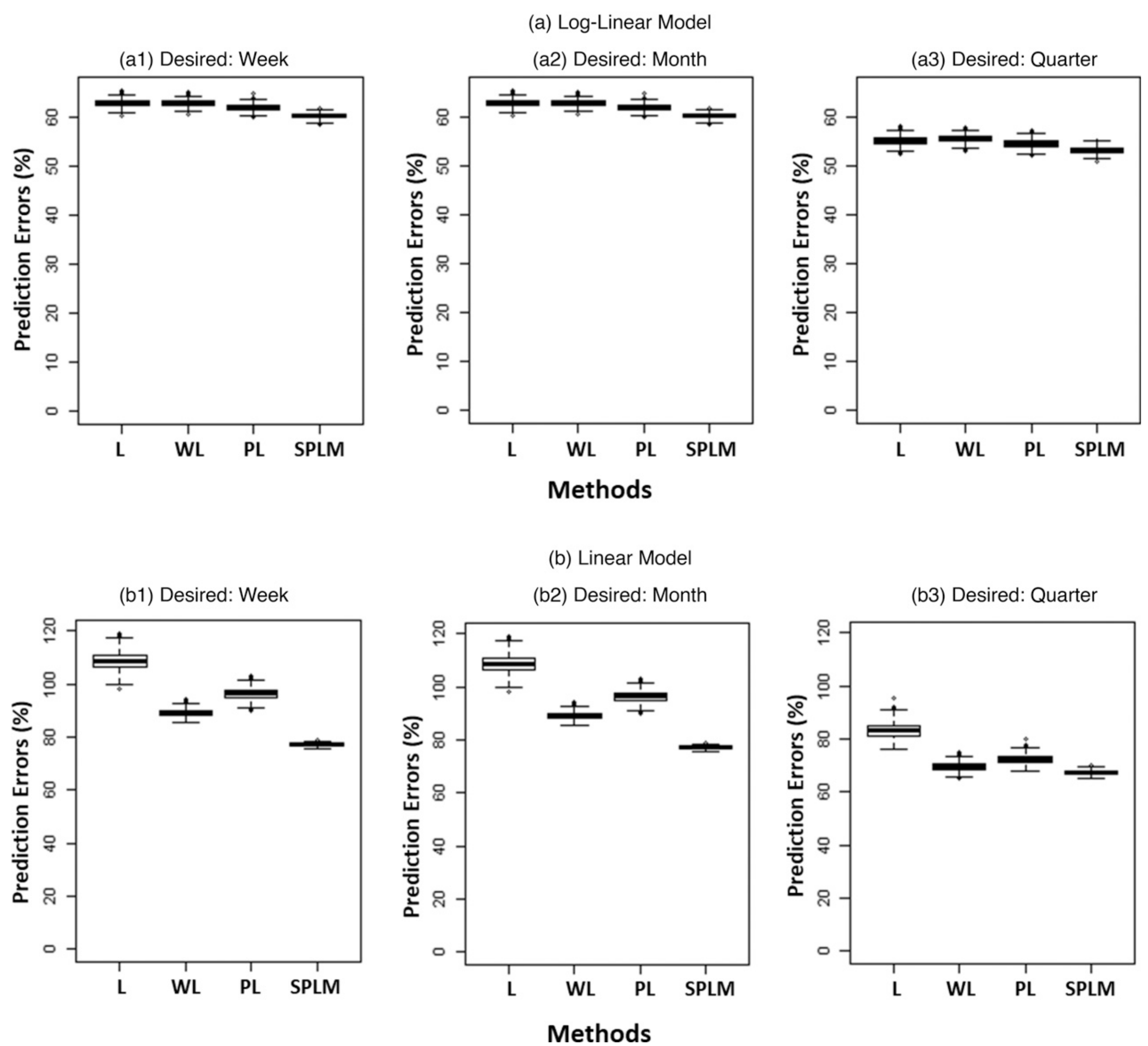


Table 2. Means (Standard Deviations) of 2015 Nielsen Orange Juice Data at Varying Granularities

	Week	Month	Quarter
Log unit sales	2.498 (0.863)	4.030 (0.740)	5.156 (0.696)
Log price	0.569 (0.067)	0.570 (0.054)	0.570 (0.051)
Log number of advertisements	0.042 (0.208)	0.180 (0.403)	0.382 (0.692)
Number of observations	2,600	600	200

three aforementioned extant methods—that is, likelihood (L), weighted likelihood (WL), and power likelihood (PL); (2) relatedly, conducting a demand analysis at the SPLM-chosen granularity improves predictive accuracy even if the chosen granularity is neither the most granular nor the managerially desired granularity (i.e., the granularity at which sales forecasts have to be made); and (3) SPLM yields doubly robust predictions as compared with the same three extant methods.

4.2. Analysis

4.2.1. Data. We used data, made available by AC Nielsen, at the universal product code (UPC)–store–week level. The data include 671 UPCs and 34,759 stores across a variety of formats (grocery, mass merchandise, drug stores) with sales, prices, and feature advertising over a period of 12 years (2006–2017). We illustrate the use of our proposed framework by analyzing the demand for the leading brand-size combination in the refrigerated orange juice category (Tropicana, 12 ounces) in drug stores across U.S. markets over a period of two years (2015 and 2016).[12] Our choice of analyzing sales in comparable stores lowers the possibility of omitted variables.

To create data at the store-week level from the data at the UPC-store-week level, we summed up unit sales and the incidence of feature advertising, and sales-weighted averaged price over all UPCs for the Tropicana brand. Thus, at the store-week level, our advertising variable indicates the number of UPCs with feature advertising at a given store and week. To facilitate the comparison of results across three granularities (weekly, monthly, quarterly), we set the month and quarter of each week as the month and quarter of the first day (Sunday) of the week. We then linearly aggregated the store-week data (i.e., summed up sales and advertisements, and took a simple average of weekly prices) to the corresponding store-month and store-quarter levels, respectively.

Table 2 summarizes the distributions of log unit sales, log price, and log feature advertising across the three granularities. The summaries are based on the data from 2015, which was used for (in-sample) model calibration, whereas the year 2016 was used for (out-of-sample) forecasting.

4.2.2. Model Specifications. Demand analysis tries to capture underlying drivers, and doing so at different granularities may require different model specifications. For instance, whereas dynamic purchase effects may need to be considered when analyzing weekly data, these effects may get averaged out and thus may not be important when it comes to using quarterly data. Thus, weekly variation in sales may be captured using an autoregressive (AR) log-log model, whereas quarterly data may not need such dynamic effects. With this underlying idea, at each granularity, we chose the best-fitted model among three model specifications—that is, an AR(p) log-log demand model, where $p = 0$, 1, or 2.[13] If $p = 0$, then we obtain a nonautoregressive log-log demand model, which does not include lagged dependent variables as covariates. The most generalized model—the AR(2) log-log model—is shown in Equation (9):

$$\begin{aligned}\log(sales_{\theta,it}) &= \alpha_{\theta,i} + \beta_{\theta,i}\log(price_{\theta,it}) + \gamma_{\theta,i}\log(ad_{\theta,it}+1) \\ &\quad + \sum_{p=1}^{2}\rho_{\theta,p}\log(sales_{\theta,it-p}) + \delta_{\theta,t} + \varepsilon_{\theta,it} \\ &= \mu_{\theta,it} + \varepsilon_{\theta,it}, \end{aligned} \tag{9}$$

where the dependent variable $sales_{\theta,it}$ indicates the unit sales for Tropicana at drug store i for the t th observation at granularity θ, with $i = 1,\ldots,I$ (where I indicates the total number of drug stores); θ = week (w), month (m), or quarter (q); $t = 1,\ldots,T_\theta$ (where $T_{\mathrm{w}} = 52$, $T_{\mathrm{m}} = 12$, $T_{\mathrm{q}} = 4$). The two covariates $price_{\theta,it}$ and $ad_{\theta,it}$ captured the price and number of feature advertisements of the brand at store i for the tth observation at granularity θ, respectively. In order to control for dynamic effects at granularity θ, we included two lagged dependent variables—$\log(sales_{\theta,it-1})$ and $\log(sales_{\theta,it-2})$.[14] Seasonal patterns at the given granularity were accommodated by including a time fixed effect $\delta_{\theta,t}$ (e.g., monthly fixed effect if we analyze monthly level data). Finally, the store-level heterogeneous parameters ($\alpha_{\theta,i}$, $\beta_{\theta,i}$, $\gamma_{\theta,i}$) and the error term ($\varepsilon_{\theta,it}$) were assumed to follow independent Normal distributions—that is, $\alpha_{\theta,i} \sim N(\alpha_\theta, \sigma^2_{\alpha,\theta})$, $\beta_{\theta,i} \sim N(\beta_\theta, \sigma^2_{\beta,\theta})$, $\gamma_{\theta,i} \sim N(\gamma_\theta, \sigma^2_{\gamma,\theta})$, $\varepsilon_{\theta,it} \sim N(0, \sigma^2_{\varepsilon,\theta})$.

Note that whereas the inclusion of lagged sales as covariates may improve model fit (as well as the out-of-sample predictions) and account for dynamic purchase effects, it can create an issue of endogeneity, because lagged sales (as well as current sales) are functions of unobserved store-level effects. We addressed the issue by applying the Arellano and Bover (1995) method that used differences in further lags of the dependent variable as instruments for a lagged term (e.g., Yoganarasimhan 2012, Gamper-Rabindran and Finger 2013, Uetake and Yang 2020). Following Uetake and Yang (2020), we used six lagged differences as instruments for a lagged term. That is, we used $\Delta\log(sales_{\theta,it-k}) = \log(sales_{\theta,it-k}) - \log(sales_{\theta,it-k-1})$, where $k = 2, 3, \ldots, 7$, and $\Delta\log(sales_{\theta,it-k})$, where $k = 3, 4, \ldots, 8$, as instruments for $\log(sales_{\theta,it-1})$ and $\log(sales_{\theta,it-2})$, respectively.[15] Since the Arellano and Bover (1995) method requires that there be no serial correlation, for each of the model specifications, we performed the Arellano-Bond test and confirmed that there was no evidence of serial correlation (e.g., Arellano and Bond 1991, Yoganarasimhan 2012). We also tested the power of the instruments via the Sanderson-Windmeijer multivariate F-tests for multiple endogenous variables (e.g., Andrews and Stock 2005, Sanderson and Windmeijer 2016, Uetake and Yang 2020) and rejected a null hypothesis of having weak instruments.

4.2.3. Approaches We selected a granularity-model pair and performed in-sample estimation and out-of-sample forecasting under both extant approaches and our framework (SPLM), as shown in Table 3. Besides using three formal extant likelihood-based approaches—approaches (3), (4), and (5)—we also used two simpler methods—approaches (1) and (2)—that are commonly used (e.g., Leone 1995, Bell and Song 2007).

1. Interpurchase time (IPT)

We set IPT as the temporal granularity based on past work (e.g., Bass and Leone 1983, Leone 1995). For example, Bass and Leone (1983) proposed to use the IPT as the data interval for analysis, as the level of the IPT may provide maximum information per observation. (They did not offer any support for the assumption.) Then, conditional on the IPT-chosen granularity, we estimated the aforementioned model specifications using the standard likelihood and selected the one with the maximum in-sample and/or out-of-sample R^2 (InR^2_θ and/or $OutR^2_\theta$).

2. The most granular data

The Nielsen data are weekly in their most granular state. Then, conditional on the finest granularity (here, week), we estimated the model specifications under the standard likelihood and selected the one with the highest InR^2_θ and/or $OutR^2_\theta$.

3. Likelihood (L)

We followed two steps: (a) At each of the three granularities, we estimated the model specifications under the standard likelihood and selected the model that maximizes InR^2_θ and/or $OutR^2_\theta$. (b) Then, among the chosen three granularity-model pairs (i.e., quarterly, weekly, and monthly models), we selected the one that maximizes InR^2_θ and/or $OutR^2_\theta$.

4. Weighted likelihood (WL)

We repeated the two steps described in (3) but used the weighted likelihood for estimation.

5. Power likelihood (PL)

We repeated the two steps in (3) but used the power likelihood for estimation.

6. Our method (SPLM)

We repeated two steps in (3) but used SPLM for estimation. We used the cross-granularity weight, G_θ, for granularity-model selection.

Table 3. Comparison of SPLM with the Extant Approaches

Approaches	Selection criteria: Granularity	Selection criteria: Model	Estimation methods
(1) IPT	IPT	InR^2_θ, $OutR^2_\theta$	L
(2) The finest granularity	The finest granularity	InR^2_θ, $OutR^2_\theta$	L
(3) L	InR^2_θ, $OutR^2_\theta$	InR^2_θ, $OutR^2_\theta$	L
(4) WL	InR^2_θ, $OutR^2_\theta$	InR^2_θ, $OutR^2_\theta$	WL
(5) PL	InR^2_θ, $OutR^2_\theta$	InR^2_θ, $OutR^2_\theta$	PL
(6) SPLM	G_θ	G_θ	SPLM

Notes. (1) IPT, L, WL, PL, and SPLM stand for the interpurchase time, likelihood, weighted likelihood, power likelihood, and scaled power likelihood with multiple weights, respectively. (2) InR^2_θ, $OutR^2_\theta$, and G_θ indicate the in-sample R^2, out-of-sample R^2, and cross-granularity weight, respectively, at granularity θ.

4.3. Results

We show the results for both the first two simpler methods and the more formal likelihood-based methods (L, WL, PL, SPLM) in Table 4. Although not explicitly shown in the table, the results for the simpler approaches based on IPT (approach 1) and using the most granular data (approach 2) are covered by our findings obtained under the L approach in Table 4, Panel A. The average IPT for the refrigerated orange juice category is 69 days, which suggests possibly quarterly granularity (albeit that is "more art than science"), and weekly data are the most granular.

For the likelihood-based approaches, we compared the "zero robust" (L), "singly robust" (WL and PL), and "doubly robust" (SPLM) methods using the two steps described in Section 4.2.3.[16] At the first step, we obtained similar results across the four approaches. That is, at each granularity, the chosen best-fitted model is the same across the four approaches (L, WL, PL, and SPLM). Specifically, the nonautoregressive model with quarterly fixed effects, the AR(1) model with monthly fixed effects, and the AR(2) model with weekly fixed effects were chosen for the quarterly, monthly, and weekly data, respectively, regardless of the approaches. This result suggests that (a) there is a lot of signals in the data; (b) as the data get more granular, an autoregressive structure is needed; and (c) it is important to use different models at different levels of granularity.

The results obtained in the second step highlight the differences between our method and the extant methods (L, WL, PL), thus supporting the first objective of the secondary data application. First, SPLM is more discriminating in selecting the best-fitted granularity-model pair (among the granularity-model candidates of

Table 4. Comparison of the Fit Measures Under the Likelihood (L), Weighted Likelihood (WL), Power Likelihood (PL), and Scaled Power Likelihood with Multiple Weights (SPLM) at Weekly, Monthly, and Quarterly Levels

	Week $\theta = \text{w}$	Month $\theta = \text{m}$	Quarter $\theta = \text{q}$
	Panel A: Likelihood (L)		
InR^2_θ	0.693	0.859	0.883
	[0.678, 0.709]	[0.846, 0.870]	[0.854, 0.905]
$OutR^2_\theta$	0.677	0.793	0.804
	[0.663, 0.691]	[0.774, 0.812]	[0.769, 0.840]
	Panel B: Weighted likelihood (WL)		
InR^2_θ	0.697	0.846	0.882
	[0.682, 0.710]	[0.828, 0.861]	[0.848, 0.904]
$OutR^2_\theta$	0.680	0.763	0.803
	[0.667, 0.694]	[0.737, 0.787]	[0.767, 0.836]
	Panel C: Power likelihood (PL)		
InR^2_θ	0.703	0.853	0.869
	[0.687, 0.716]	[0.837, 0.866]	[0.836, 0.895]
$OutR^2_\theta$	0.685	0.791	0.790
	[0.669, 0.699]	[0.767, 0.811]	[0.749, 0.826]
	Panel D: SPLM		
$\hat{G}_\theta$	0.746	0.948	0.549
	[0.722, 0.769]	[0.914, 0.971]	[0.466, 0.632]

Notes. (1) In L, WL, and PL, InR^2_θ and $OutR^2_\theta$ indicate Bayesian in-sample R^2 at granularity θ, and Bayesian out-of-sample R^2 at granularity θ, respectively. In SPLM, $\hat{G}_\theta$ represents cross-granularity weight. The values in cells indicate the posterior means of these measures. The 95% credible intervals are in brackets. (2) We computed the credible interval of the Bayesian in-sample R^2 (InR^2_θ) in the following way. For each posterior draw s, we computed the vector of $\hat{\mu}_{\theta,it,s}$ (i.e., estimated $\log y_{\theta,it}$) and $\hat{\varepsilon}_{\theta,it,s}$ (i.e., $\log y_{\theta,it} - \hat{\mu}_{\theta,it,s}$), and then $InR^2_\theta = \frac{\sigma^2(\hat{\mu}_{\theta,it,s})}{\sigma^2(\hat{\mu}_{\theta,it,s})+\sigma^2(\hat{\varepsilon}_{\theta,it,s})}$ following Gelman et al. (2019). Then, we used the vector of InR^2_θ (with size equal to the number of posterior draws) to calculate its credible interval. The credible interval of the Bayesian out-of-sample R^2 ($OutR^2_\theta$) was computed in the same way.

interest) than the extant methods are. If we were to use SPLM, then we would select to analyze the demand data at a monthly level. The monthly data (and its accompanying model) yield significantly higher cross-granularity weight ($\hat{G}_\text{m} = 0.948$) and thus better fit than the other two granularity-model pairs (i.e., quarterly and weekly models). According to our simulation study, it might be because the month-to-quarter (linear) aggregation created bias and because the weekly data provided much larger stochastic error than the monthly data. However, under the L, WL, and PL methods, there is no single "winning" granularity-model pair. Unlike the cross-granularity weight, the selection criteria (InR^2_θ and $OutR^2_\theta$) under the extant methods are not significantly higher at a monthly level than at a quarterly level. It is because both InR^2_θ and $OutR^2_\theta$ lack of cross-granularity comparability and particularly favor coarser granularity, as explained in Section 3 (specifically Figure 5).

Next, in order to support the second objective, we examined whether estimating and predicting sales at the SPLM-chosen granularity (here, monthly) could improve out-of-sample predictive performance, even when the chosen granularity and the managerially desired granularity (i.e., the granularity at which sales forecasts have to be made) do not match. As in the simulation study, we assumed that a brand manager wanted to forecast quarterly sales and considered two forecasting approaches. In a quarterly approach, we estimated the best-fitted model at a quarterly level (i.e., the desired granularity) and predicted quarterly sales. In a monthly approach, we estimated the best-fitted model at a monthly level (i.e., the SPLM-chosen granularity-model pair), predicted monthly sales, and summed them up to quarterly sales. In Figure 8, we obtained smaller out-of-sample prediction errors under the monthly approach than under the quarterly approach (i.e., M1 < Q1 under L; M2 < Q2 under WL; M3 < Q3 under PL; M4 < Q4 under SPLM). For instance, under L, by using the monthly approach instead of the quarterly one, we reduced the posterior mean of prediction errors from 26.6% (with a 95% credible interval [23.8%, 30.0%]) to 21.8% (with a 95% credible interval [19.9%, 23.9%]).

One may argue that the manager can reduce prediction errors just because the monthly data are more granular (and may have more information) than quarterly data. This argument is in line with the common belief that the use of the most granular data should lead to the most accurate out-of-sample predictions. To test this intuition, we introduced a weekly approach—that is, estimating the best-fitted model at a weekly level (i.e., the finest granularity), predicting weekly sales, and summing them up to a quarterly level—and compared out-of-sample prediction errors under the monthly and weekly approaches. Figure 8 provides evidence against this common belief by showing that prediction errors under the weekly approach are larger than those under the monthly approach (i.e., M1 < W1 under L; M2 < W2 under WL; M3 < W3 under PL). In summary, Figure 8 extends our simulation results on granularity-model selection (specifically Figure 6) to the real empirical setting. It shows that out-of-sample prediction accuracy is improved by analyzing at the SPLM-chosen granularity (here, monthly), which provides the best in-sample fit, even if the chosen granularity is not at the desired granularity (here, quarterly) or the most granular level (here, weekly).

Also, consistent with the simulation study, we now assumed that the brand manager was interested in forecasting weekly sales (the most granular level) and that the manager considered three forecasting approaches: (1) estimating the best-fitted model at a weekly level (i.e., desired granularity) and predicting weekly sales; (2) estimating the best-fitted model at a monthly level

Table 5. Comparison of Percentage Errors of Predicting Weekly Sales Across the Forecasting Approaches and the Estimation Methods

	Forecasting approaches		
Estimation methods	Week	Month	Quarter
Likelihood (L)	35.4 [34.4, 36.5]	35.7 [34.5, 37.1]	38.5 [36.8, 41.0]
Weighted likelihood (WL)	36.1 [34.9, 37.5]	35.5 [34.3, 36.6]	38.1 [36.3, 40.5]
Power likelihood (PL)	36.0 [35.0, 37.2]	35.7 [34.5, 37.2]	39.1 [37.0, 41.9]
SPLM	35.0 [34.1, 36.0]	34.1 [33.1, 35.2]	36.7 [34.9, 39.0]

Notes. (1) We measured out-of-sample prediction errors with weighted mean absolute percentage error (WAPE), which takes the sales-weighted average of the absolute percentage errors (APEs), which is explained in Online Appendix B. (2) The values in cells indicate the posterior means of WAPE. The 95% credible intervals are in brackets.

(i.e., the SPLM-chosen granularity-model pair), predicting monthly sales, and consistently disaggregating them into weekly sales; and (3) estimating the best-fitted model at a quarterly level, predicting quarterly sales, and consistently disaggregating them into weekly sales. (Consistent disaggregation is defined and described in Online Appendix D.) The monthly and quarterly approaches must assume month-to-week and quarter-to-week disaggregation processes, respectively, which are typically (if not always) unknown in empirical settings. Thus, researchers might believe that those approaches would provide considerably higher out-of-sample prediction errors than those from the weekly approach. However, Table 5 shows that this belief does not hold under the monthly (i.e., SPLM-chosen) approach. Whereas prediction errors under the quarterly approach are greater than those under the weekly approach (e.g., 38.5% versus 35.4% under L), the monthly and weekly approaches show similar results (e.g., 35.7% versus 35.4% under L). This finding suggests that forecasts at the SPLM-chosen granularity (where $\hat{G}_\theta$ is maximized), which provides the best in-sample fit, even when it is coarser than the managerially desired one, can be as accurate as those obtained from more granular data.

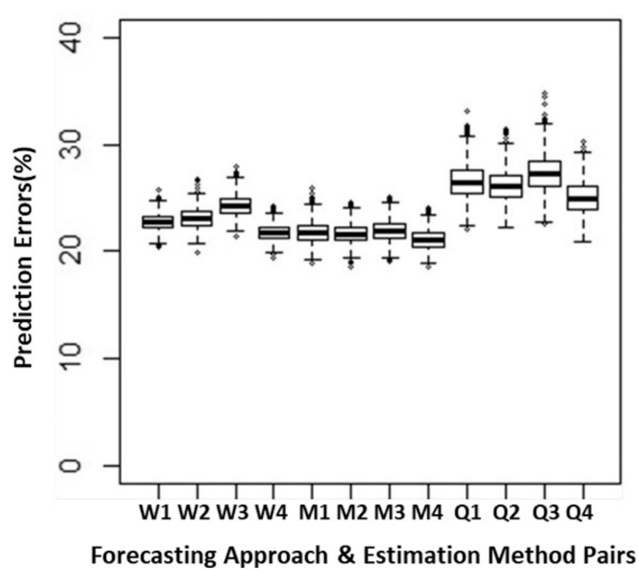
Last, Table 5 shows the role of SPLM in doubly robust prediction, supporting the third objective of the empirical study. The errors in weekly sales predictions are marginally smaller under SPLM than under extant methods (L, WL, PL), regardless of the forecasting approaches. This finding is consistent with the simulation results (specifically Figure 6), except that SPLM marginally improves predictions, even at the SPLM-chosen granularity-model pair (e.g., 35.7% under L versus 34.1% under SPLM). Whereas the SPLM-chosen granularity-model pair provides the best out-of-sample predictions among a set of granularity-model pairs of our interest, prediction errors still exist. Thus, the prediction errors at the SPLM-chosen pair can be reduced even more so by applying SPLM, which down-weights observations and data that are discrepant from the chosen model.

5. Conclusions and Future Research

Researchers frequently use historical demand data for forecasting and other managerial goals. For any such analysis, two decisions play a key role in determining how managerially useful, accurate, and stable the results would be—(a) the temporal granularity for the analysis and (b) the demand model given the chosen granularity. For the former, extant research has employed a variety of heuristics to justify the choice (e.g., use the most granular data, or use the granularity at which sales forecasts have to be made). In addition, the same demand model specification may not be appropriate across granularities—the best-fitted model for the analysis of demand at the weekly level may differ from that at the monthly level. Thus, the two decisions are intertwined and should be decided jointly. How to do so, however, is not straightforward.

The problem of what granularity to choose for analysis, and which model to employ given the chosen

Figure 8. Posterior Distributions of Percentage Errors of Predicting Quarterly Sales Across the Forecasting Approaches and the Estimation Methods



Notes. (1) The *x*-axis indicates a forecasting approach and estimation method pair. From left to right, they are (W1) week + L, (W2) week + WL, (W3) week + PL, (W4) week + SPLM, (M1) month + L, (M2) month + WL, (M3) month + PL, (M4) month + SPLM, (Q1) quarter + L, (Q2) quarter + WL, (Q3) quarter + PL, (Q4) quarter + SPLM. (2) We measured out-of-sample prediction errors with weighted mean absolute percentage error (WAPE), which takes the sales-weighted average of the absolute percentage errors (APEs), which is explained in Online Appendix B.

Marketing Science, 2022, vol. 41, no. 4, pp. 848–866, © 2022 INFORMS

granularity, is managerially quite important as well, given that past research has shown that out-of-sample forecasts may vary across granularities and models (e.g., Christen et al. 1997). Consider an organization wherein the sales division is using data at a monthly level to forecast sales for the upcoming year, while the operations division is doing so at a weekly level to accomplish the same goal. These two divisions may have different sales forecasts, and therefore doing so can create friction within the organization in terms of determining how much inventory to maintain and what level of advertising needs to be done. To ameliorate this issue, it would be ideal if one could obtain demand forecasts that are not as sensitive to both the chosen granularity and model.

From both a theoretical and a practical perspective, how to select the best-fitted granularity-model pair and provide robust forecasts across pairs is an important problem. One important piece in this puzzle is a way to compare model fit across granularity-model pairs as the number of observations also differs across such pairs (in which case the typical metrics of AIC and BIC fail to be applicable). In this paper, we address all of these issues by proposing a tool, called *scaled power likelihood with multiple weights* (SPLM), that allows a researcher to (1) jointly select the granularity-model pair that provides the best in-sample fit (comparable across granularities, which allows for selection) and (2) provide doubly robust protection against granularity misselection and model misspecification (conditional on the granularity).

We highlight the benefits of our proposed approach as compared with other extant methods such as likelihood, weighted likelihood, and power likelihood using both a large-scale simulation study and a secondary data application. The simulation study provides three main results: (1) conducting analysis at the SPLM-chosen granularity (even if the chosen one is coarser than the most granular data available) can provide better in-sample fit and out-of-sample predictions; (2) SPLM improves out-of-sample forecasts more so than the extant methods due to its ability to select the best-fitted granularity-model pair and conduct doubly robust prediction; and (3) even when the managerial goal is to predict sales at a certain granularity (e.g., week or quarter), prediction accuracy can be improved by performing model estimation at the granularity chosen by SPLM (e.g., month) and then appropriately transforming the predictions. The application of our proposed framework to Nielsen store-level scanner data confirms the findings from the simulation study. Notably, SPLM selects a level of analysis (i.e., store-monthly) that is neither the most granular nor one that would be selected by simpler extant approaches.

Our simulation and secondary data application also suggest the following three guidelines to researchers who are going to use our method (SPLM). First, SPLM is most beneficial as a doubly robust granularity-model selection tool when (a) sample size is large, (b) estimated signal-to-noise ratio (SNR) is high, and (c) the out-of-sample period is stationary with respect to the in-sample period. Second, SPLM can be used when researchers apply a model in which our proposed scaling step (described in Section 2) can preserve the functional form of the distribution of the dependent variable. For instance, in our simulation studies, we used log-log, log-linear, and linear demand models. Lastly, researchers should note that, whereas the SPLM-chosen granularity-model pair is the best-fitted one for an application, we do not imply that it is the true data generating process (DGP) for that context. For example, in a grocery context, the true DGP is likely at the customer level, but if we model store-level sales (as in our empirical analysis), then we will not uncover the true DGP, even if we analyze the data at the best-fitted pair.

Future research may consider substantive extensions. First, although we motivate a granularity-model selection issue with respect to time granularity (e.g., weekly or monthly), researchers could extend the application of SPLM to other types of granularities. For instance, much past research has investigated the impact of independent variables (e.g., price) on the choices made by decision makers, but the level at which the decision maker is modeled (a single individual or a dyad) is typically assumed a priori (e.g., Adamowicz et al. 2005). Another context is spatial analysis, where researchers must select the spatial units to analyze (e.g., census tract, zip code) and build a model based on that chosen spatial unit (e.g., Choi et al. 2010). Second, our proposed framework can also be helpful in other applications (e.g., data privacy), where it is important to understand the trade-off between loss of information (via data aggregation) versus privacy (keeping data at a finer granularity). Recent work in the area of data privacy and the estimation of marketing mix models offers an interesting opportunity for the application of SPLM as a metric to choose the appropriate level of data aggregation (e.g., Schneider et al. 2018). Lastly, whereas our research focuses on developing an innovative approach for identifying the best-fitted granularity that can be used for out-of-sample forecasting, it could serve as a foundation for more normative research on when and why granularity matters. For example, researchers could uncover new and interesting mechanisms through which more granular data might provide worse fit and forecasts. Also, they could consider managerial goals that could not be attained by selecting the best-fitted granularity and uncover an interplay between these managerial goals and granularity.

Future research may consider methodological extensions as well. First, whereas our research assumes a single granularity for the entire data set, there are certainly contexts where multiple granularities may be warranted. For instance, consider a manager who is analyzing sales after a new product introduction. In this case, demand analysis at a weekly level may be preferable soon after the product is introduced, whereas a coarser level of granularity may be preferable as the product matures. This intuition is consistent with an extant finding that the speed of consumer's decision making (e.g., adoption, disadoption) varies by product age (e.g., Arora et al. 2017). Second, it will be interesting to consider dynamics in a fast-changing market, by updating (online learning) the best-fitted granularity-model pair as more data are added. Third, future research can propose a scaling step that is more generalizable than our proposed one (described in Section 2). Finally, whereas our proposed method is likelihood and data-size agnostic in theory, the application to large data sets can be computationally prohibitive and may require more scalable methods.

In conclusion, SPLM may help improve out-of-sample predictions (e.g., sales forecasting) due to its ability to select the best-fitted granularity-model pair and conducting doubly robust prediction. The issues related to the appropriate level of data aggregation and the corresponding model are quite general, and we hope this generality enhances the attractiveness of our proposed tool to researchers across a wide variety of domains.

Acknowledgments
Researchers own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Endnotes

[1] In the literature on missing data and causal inference models, researchers have used a term "doubly robust" in order to offer protection against misspecification of either of two models—for example, one for a missingness mechanism and the other for a treatment effect (e.g., Bang and Robins 2005). However, in this paper, we use the term "doubly robust" for a different purpose but in a similar spirit—that is, to provide protection against granularity misselection and model misspecification conditional on the granularity.

[2] If one posits a prior that constrains weights to be at most 1 (e.g., spike-and-slab beta), as in our simulations, then observations that fit an assumed model well receive $w_{\theta,it} = 1$ (the prior mode) and the others are down-weighted ($0 < w_{\theta,it} < 1$). If one assumes a prior without this constraint, then observations can be either down- or up-weighted ($w_{\theta,it} > 1$).

[3] If G_θ and $f_{\theta,it}$ were fully unconstrained, then the decomposition of $w_{\theta,it}$ in Equation (3) into G_θ and $f_{\theta,it}$ in Equation (4) would not be identified. However, as we now state, G_θ and $f_{\theta,it}$ have different impact on $w_{\theta,it}$ (and hence are identified) since they meet the aforementioned conditions (e.g., $0 < G_\theta \le 1$ and $\sum_{it=1}^{n_\theta} f_{\theta,it} = 1$). First, a decrease in G_θ lowers $w_{\theta,it}$ for *all* observations. This cannot be done by lowering $f_{\theta,it}$, because we cannot lower $f_{\theta,it}$ for all the observations due to its constraint ($\sum_{it=1}^{n_\theta} f_{\theta,it} = 1$). Second, a decrease in $f_{\theta,it}$ for some observations lowers $w_{\theta,it}$ for those observations. This cannot be achieved by lowering G_θ, because a decrease in G_θ would lower $w_{\theta,it}$ for all observations.

[4] The scaled in-sample fit does not depend on granularity θ, because it is an average of the log of the scaled likelihood terms, $p_{1(\theta)}(sales_{\theta,it} | \boldsymbol{\beta_\theta})$, which transform the standard likelihood terms, $p_\theta(sales_{\theta,it} | \boldsymbol{\beta_\theta})$, to the same granularity and thus do not depend on granularity θ.

[5] The means and variances of monthly price and advertising variables were similar to those from the real-data application—that is, $\mu_{\text{price,m}} = \mu_{\text{ad,m}} = 0$ and $\sigma^2_{\text{price,m}} = \sigma^2_{\text{ad,m}} = 0.4$.

[6] We designed our simulation study following the related literature on price and advertising elasticities (e.g., Christen et al. 1997, Bijmolt et al. 2005, Hanssens 2015) and the results from our real-data application (in Section 4). We set the intercept $\alpha_{\text{m}} = 0$, price elasticity $\beta_{\text{m}} = \beta = -2$, and advertising elasticity $\gamma_{\text{m}} = \gamma = 0.1$, following the related literature. We set all of the parameter variances $\sigma^2_{\alpha,\text{m}}$, $\sigma^2_{\beta,\text{m}}$, and $\sigma^2_{\gamma,\text{m}}$ as 0.01, which are similar to those estimated from the empirical application. Month-level fixed effects ($\delta_{\text{m},t}$) were also set based on our real-data application results. Note that the error variance $\sigma^2_{\varepsilon,\text{m}}$ had two levels—high and low—depending on the level of the SNR, as explained in Section 3.2.1.

[7] We did not take a sales-weighted average of prices for the following reason. Since we were interested in forecasting sales, which was unknown when prediction was conducted, it did not make sense to aggregate price to more-aggregated time units using "unknown" sales information.

[8] When disaggregating monthly sales into five weeks, we followed the same steps but used a disaggregation vector that had five elements that summed up to 1.

[9] Our disaggregation process introduced additional noise that was negatively correlated across weeks within a month but uncorrelated across months. This structure of the additional noise may not match the empirical patterns of many data sets.

[10] Across all the four methods, all intercepts and slope parameters in Equation (7)—that is, α_θ, β_θ,γ_θ—had Gaussian priors, $N(0,3)$. Parameter and error variances—that is, $\sigma^2_{\alpha,\theta}$, $\sigma^2_{\beta,\theta}$, $\sigma^2_{\gamma,\theta}$, $\sigma^2_{\varepsilon,\theta}$—had weakly informative inverse-gamma priors, $IG(0.5, 0.5)$. As for SPLM, we imposed the priors that we derived in Section A1 of Online Appendix A to the cross-granularity weight (G_θ) and the vector of the within-granularity weights ($\boldsymbol{f_\theta}$). Specifically, G_θ was assumed to have a spike-and-slab prior using $Beta(2, 1)$, and $\boldsymbol{f_\theta}$ was assumed to follow a symmetric Dirichlet prior with a concentration parameter of 2.

[11] When the degree of in-sample misfit is very small (less than 40%), the extant in-sample measures ($InR^2_{\text{L},\theta}$, $InR^2_{\text{WL},\theta}$, and $InR^2_{\text{PL},\theta}$) are always maximized at a quarterly level (i.e., the coarsest granularity). This is because the distribution of the in-sample R^2 depends on the number of observations and particularly favors a coarser granularity (the smaller number of observations). To be specific, since it follows a beta distribution, $Beta(a, b_\theta)$, whose shape parameter (b_θ) increases with the number of observations (n_θ), its distribution at a coarser granularity is more right-skewed and thus favors larger values (of $InR^2_{\text{L},\theta}$, $InR^2_{\text{WL},\theta}$, and $InR^2_{\text{PL},\theta}$). For the same reason mentioned earlier, the out-of-sample R^2 favors a coarser granularity as well.

[12] We excluded drug stores where information about feature advertising was not recorded in either of those two years. After this exclusion, there were 50 drug stores in the sample. The leading brand, Tropicana, accounted for 42%, the largest market share, of the total category sales

volume of drug stores. The sales of the size 12-ounce carton accounted for 59% of the sales volume of the leading brand at drug stores.

[13] Our analyses applied an AR(p) log-log demand model where p = 0, 1, or 2; however, whether adding more lags ($p > 2$) and/or using different functional forms improves the in-sample fit and out-of-sample prediction accuracy should be tested based on the empirical context.

[14] We used the data set of year 2014 (as well as the data sets of years 2015 and 2016) to create lagged dependent variables. For example, one-period lagged unit sales at store i for the first quarter ($t = 1$, $\theta = \text{q}$) of year 2015 indicates unit sales at store i for the fourth quarter ($t = 4$, $\theta = \text{q}$) of year 2014.

[15] We used the data sets of years 2013 and 2014 (as well as the data sets of years 2015 and 2016) to create instrument variables. Our analyses used differences in lagged variables as instruments; however, other types of instruments should be tested based on the empirical context.

[16] We expected that our method would have good discrimination, as this empirical study satisfies the boundary conditions suggested by the simulation studies—namely, (1) larger sample size (here, larger than the small sample size [25] of the simulation) and (2) high estimated signal-to-noise ratio (SNR) (here, 8, 6, and 2 at the quarterly, monthly, and weekly levels, all of which were higher than the low SNR [1/9] of the simulation). Note that estimated SNR at granularity θ was estimated as $\frac{\sigma^2(\hat{\mu}_{\theta,it})}{\sigma^2(\hat{\varepsilon}_{\theta,it})}$, where $\hat{\mu}_{\theta,it}$ replaced parameters in Equation (9) with their estimates and $\hat{\varepsilon}_{\theta,it} = \log y_{\theta,it} - \hat{\mu}_{\theta,it}$.

References

Adamowicz W, Hanemann M, Swait J, Johnson R, Layton D, Regenwetter M, Reimer T, Sorkin R (2005) Decision strategy and structure in households: A "groups" perspective. *Marketing Lett.* 16(3–4):387–399.

Andrews DWK, Stock J (2005) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (Cambridge University Press, New York).

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. *J. Econometrics* 68(1):29–51.

Arora S, Hofstede FT, Mahajan V (2017) The implications of offering free versions for the performance of paid mobile apps. *J. Marketing* 81(6):62–78.

Bang H, Robins JM (2005) Doubly robust estimation in missing data and causal inference models. *Biometrics* 61(4):962–972.

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

Bell DR, Song S (2007) Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quant. Marketing Econom.* 5(4):361–400.

Bijmolt THA, Van Heerde HJ, Pieters RGM (2005) New empirical generalizations on the determinants of price elasticity. *J. Marketing Res.* 42(2):141–156.

Choi J, Hui SK, Bell DR (2010) Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. *J. Marketing Res.* 47(1):75–89.

Christen M, Gupta S, Porter JC, Staelin R, Wittink DR (1997) Using market-level data to understand promotion effects in a nonlinear model. *J. Marketing Res.* 34(3):322–334.

Gamper-Rabindran S, Finger SR (2013) Does industry self-regulation reduce pollution? Responsible care in the chemical industry. *J. Regulatory Econom.* 43:1–30.

Gelman A, Goodrich B, Gabry J, Vehtari A (2019) R-squared for Bayesian regression models. *Amer. Statistician* 73(3):307–309.

Hanssens DM (2015) Empirical generalizations about marketing impact. Webinar, Marketing Science Institute, New York.

Hui S, Fader PS, Bradlow ET (2009) Path data in marketing: An integrative framework and prospectus for model building. *Marketing Sci.* 28(2):320–335.

Ivancic L, Diewert WE, Fox KJ (2011) Scanner data, time aggregation and the construction of price indexes. *J. Econometrics* 161(1):24–35.

Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei DM (2017) Automatic differentiation variational inference. *J. Machine Learning Res.* 18(14):1–45.

Leone RP (1995) Generalizing what is known about temporal aggregation and advertising carryover. *Marketing Sci.* 14(3):G141–G150.

Mao G (2013) Model selection for regression with heteroskedastic and autocorrelated errors. *Econom. Lett.* 118(3):497–501.

McCarthy D, Jensen S (2016) Power-weighted densities for the time series data. *Ann. Appl. Stat.* 10(1):305–334.

Mela CF, Gupta S, Lehmann DR (1997) The long-term impact of promotion and advertising on consumer brand choice. *J. Marketing Res.* 34:248–261.

Miller JW, Dunson DB (2019) Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* 114(527):1113–1125.

Ruiz FJR, Athey S, Blei DM (2020) SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *Ann. Appl. Statist.* 14(1):1–27.

Sanderson E, Windmeijer F (2016) A weak instrument *F*-test in linear IV models with multiple endogenous variables. *J. Econometrics* 190(2):212–221.

Schneider M, Jagpal S, Gupta S, Li S, Yu Y (2018) A flexible method for protecting marketing data: An application to point-of-sale data. *Marketing Sci.* 37(1):153–171.

Tellis GJ, Franses PH (2006) Optimal data interval for estimating advertising response. *Marketing Sci.* 25(3):217–229.

Uetake K, Yang N (2020) Inspiration from the "Biggest Loser": Social interactions in a weight loss program. *Marketing Sci.* 39(3):487–499.

Wang Y, Kucukelbir A, Blei DM (2017) Robust probabilistic modeling with Bayesian data reweighting. *Proc. 34th Internat. Conf. Machine Learning* 70:3646–3655.

Yoganarasimhan H (2012) Impact of social network structure on content propagation: A study using YouTube data. *Quant. Marketing Econom.* 10(1):111–150.