ELSEVIER

Intern. J. of Research in Marketing 25 (2008) 22 – 33

International Journal of Research in Marketing

www.elsevier.com/locate/ijresmar

Estimating the SCAN*PRO model of store sales: HB, FM or just OLS?

Rick L. Andrews [a,1], Imran S. Currim [b,*], Peter Leeflang [c,2], Jooseop Lim [d,3]

[a] Department of Business Administration, University of Delaware, Newark, DE 19716, USA
[b] Paul Merage School of Business, University of California, Irvine, California 92697-3125, USA
[c] University of Groningen, Faculty of Economics, PO Box 800, 9700 AV Groningen, The Netherlands
[d] John Molson School of Business, Concordia University, Montreal, Quebec, Canada H3G 1M8

First received in March 28, 2005 and was under review for 8.5 months, Area editor Marnik Dekimpe

Abstract

In this paper we investigate whether consideration of store-level heterogeneity in marketing mix effects improves the accuracy of the marketing mix elasticities, fit, and forecasting accuracy of the widely-applied SCAN*PRO model of store sales. Models with continuous and discrete representations of heterogeneity, estimated using hierarchical Bayes (HB) and finite mixture (FM) techniques, respectively, are empirically compared to the original model, which does not account for store-level heterogeneity in marketing mix effects, and is estimated using ordinary least squares (OLS). The empirical comparisons are conducted in two contexts: Dutch store-level scanner data for the shampoo product category, and an extensive simulation experiment. The simulation investigates how between- and within-segment variance in marketing mix effects, error variance, the number of weeks of data, and the number of stores impact the accuracy of marketing mix elasticities, model fit, and forecasting accuracy. Contrary to expectations, accommodating store-level heterogeneity does not improve the accuracy of marketing mix elasticities relative to the homogeneous SCAN*PRO model, suggesting that little may be lost by employing the original homogeneous SCAN*PRO model estimated using ordinary least squares. Improvements in fit and forecasting accuracy are also fairly modest. We pursue an explanation for this result since research in other contexts has shown clear advantages from assuming some type of heterogeneity in market response models. In an Afterthought section, we comment on the controversial nature of our result, distinguishing factors inherent to household-level data and associated models vs. general store-level data and associated models vs. the unique SCAN*PRO model specification.

Keywords: Store sales models; SCAN*PRO; Between-store heterogeneity; Marketing mix effects

1. Introduction

The SCAN*PRO and other related models of store-level product sales have seen more than 3000 reported commercial applications worldwide in addition to a number of academic applications (e.g., Christen, Gupta, Porter, Staelin, & Wittink, 1997; Foekens, Leeflang, & Wittink, 1994; Hoch, Kim, Montgomery, & Rossi, 1995; Horvath, Leeflang, Wieringa, & Wittink, 2005; Montgomery, 1997; Montgomery & Rossi, 1999; Van Heerde, Leeflang, & Wittink, 2000, 2001, 2002; Wittink, Addona, Hawkes, & Porter, 1988). In this paper we investigate whether consideration of store-level heterogeneity in marketing mix effects improves the accuracy of marketing mix elasticities, fit, and forecasting accuracy of the SCAN*PRO model. Our investigation is a natural response to Leeflang and Wittink's (2000) call for more research on store-level heterogeneity. In that our main innovation is to allow for store-level heterogeneity in the context of the widely applied SCAN*PRO model, our research contributes to an under-addressed corner of the store-level market response modeling literature, especially in commercial settings.[4] In addition, there is growing interest in the marketing science community on whether and why academic work affects marketing practice (Roberts, Kayande, & Stremersch 2007). Since the SCAN*PRO model is used widely in both communities there is relevance in knowing the

* Corresponding author. Tel.: +1 949 824 8368; fax: +1 949 725 2826.
E-mail addresses: andrews@lerner.udel.edu (R.L. Andrews), iscurrim@uci.edu (I.S. Currim), p.s.h.leeflang@eco.rug.nl (P. Leeflang), jlim@jmsb.concordia.ca (J. Lim).
[1] Tel.: +1 302 831 1190.
[2] Tel.: +31 0 503633696.
[3] Tel.: +1 514 848 2424x2772.

[4] We thank two anonymous reviewers for pointing this out.

0167-8116/$ - see front matter
doi:10.1016/j.ijresmar.2007.10.001

added value of further academically-driven extensions or complexities to the basic model.[5]

This study investigates heterogeneity in marketing mix effects between stores using two dominant methodologies for investigating heterogeneity: a random coefficients model estimated with hierarchical Bayes (HB) methods and a finite mixture (FM) model that derives *segments* of stores having different sales responses to prices and promotions. HB methods are frequently used to estimate models assuming continuous distributions of heterogeneity, while FM models assume discrete, nonparametric distributions of heterogeneity. For benchmarking purposes we also estimate the original widely-applied non-heterogeneous SCAN*PRO model estimated using ordinary least squares (OLS).

The models are empirically contrasted in two settings, actual scanner data and simulated data. An application to Dutch store-level scanner data for the shampoo category allows us to investigate whether allowing for heterogeneity in marketing mix effects between stores improves the fit and prediction of the original SCAN*PRO model, and the magnitude of improvement in fit and prediction. The empirical analyses of the real world scanner data do not, however, answer the question of how much improvement there is in the accuracy of store-level marketing mix elasticities, since the true store-level responses are unknown. To address this question, we design an extensive simulation study, wherein we experimentally manipulate variables that may impact model fit, forecasts, and the accuracy of marketing mix elasticities, such as the level of heterogeneity in the effects of the marketing mix, both between and within-segments of stores, the number of stores, the number of weeks, and the level of error variance.

From a managerial standpoint, if stores within a chain are truly heterogeneous in their responses to price and promotion, the use of a model that does not account for such heterogeneity could result in chain-level managers over- (or under-) estimating sales response to the marketing mix for particular stores. Such over- (or under-) estimation of sales response to the marketing mix is likely to influence adversely the sales expectations for marketing efforts at particular stores and hence incorrectly affect evaluations of store locations and customers, including the performance of store managers and other employees. Consequently, from a managerial viewpoint it is important to investigate whether there is heterogeneity in marketing mix effects between stores. In addition, finding individual stores or segments of stores having different responses to price and promotion activities can provide important information for both manufacturers (e.g., account managers) and retailers, who could subsequently re-allocate marketing effort to these stores in a way that is more efficient.

Andrews, Ansari, and Currim (2002) and Andrews, Ainslie, and Currim (2002)[6] investigate the benefits of using continuous vs. discrete representations of heterogeneity in consumer preference (e.g., conjoint regression) and consumer choice (e.g., multinomial logit) contexts, respectively. While both studies largely find equivalence between continuous and discrete representations of heterogeneity, they also find that the heterogeneous versions of the model provide substantial benefits over their homogeneous counterparts. Our results in the context of the SCAN*PRO model differ from these previous results, and in the discussion section of this paper we pursue an explanation for the difference. In doing so, we expand on what is known about the benefits of modeling heterogeneity using different kinds of market response models.

In the next section, we begin with the original SCAN*PRO econometric model. We then present the HB and FM extensions of the model. After an empirical application to Dutch store-level scanner data from the shampoo category, we present the simulation experiment. Subsequently, we summarize the results of both studies, including implications for marketing managers and modelers operating in commercial settings.

Finally, in an Afterthought section we comment on our controversial results in a more speculative manner, providing both sides of the controversy. We address what we could have done differently after the fact, what should others do, and how one could stretch the findings or implications of this paper. These discussions are meant to provoke thought, trigger debate, and perhaps speculate beyond the bounds of the actual findings (Stremersch & Lehmann, 2007).

2. Models

2.1. Original SCAN*PRO model with homogeneous marketing mix elasticities

The original SCAN*PRO model is a multiplicative model that decomposes sales for brand j (1,…,n) in store k (1,…,K) during week t (1,…,T) into (i) own- and cross-brand effects of price, feature advertising, and aisle displays, (ii) week effects, (iii) store effects, and (iv) a random component, as follows:

$$Q_{kjt} = \left[\prod_{r=1}^{n}\left(\frac{p_{krt}}{\overline{p}_{krt}}\right)^{\beta_{rj}}\prod_{l=1}^{3}\gamma_{lrj}^{D_{lkrt}}\right]\left[\prod_{t=1}^{T}\delta_{jt}^{X_t}\right]\left[\prod_{k=1}^{K}\lambda_{kj}^{Z_k}\right]e^{\varepsilon_{kjt}} \qquad (1.1)$$

where:

Q_{kjt} unit sales of brand j in store k, week t;
p_{krt} unit price for brand r in store k, week t;
$\overline{p}_{krt}$ regular unit price for brand r in store k, week t;
D_{1krt} an indicator variable for feature advertising: 1 if brand r is featured but not displayed by store k, in week t; 0 otherwise;
D_{2krt} an indicator variable for display: 1 if brand r is displayed but not featured by store k, in week t; 0 otherwise;
D_{3krt} an indicator variable for the simultaneous use of feature and display: 1 if brand r is featured and displayed; 0 otherwise;
X_t an indicator variable: 1 if observation is in week t;
Z_k an indicator variable: 1 if observation is in store k;
β_{rj} the own-brand (r=j) and cross-brand ($r \neq j$) price discount elasticities;
γ_{lrj} feature only (l=1), display only (l=2), feature &

[5] We thank the Area Editor for pointing this out.
[6] We select these works because they, too, are based on extensive simulations. Other works are by Otter, Tüchler, and Frühwirth-Schnatter (2004) and Moore (2004).

display ($l=3$) multipliers;
δ_{jt} seasonal multiplier for brand j, week t;
λ_{kj} store multiplier for brand j, store k;
ε_{kjt} disturbance term.

For identification purposes, we omit one week dummy and one store dummy and include an intercept. Model (1.1) can be estimated with OLS regression after taking the logs of both sides of the equation.

2.2. Model with continuous heterogeneity distribution for marketing mix elasticities

To simplify notation, rewrite model (1.1) as

$$q_{kjt} = X_{kjt}\beta + \varepsilon_{kjt} \tag{1.2}$$

where, $q_{kjt}=\ln(Q_{kjt})$, the marketing mix variables and weekly indicator variables[7] are included in the matrix X_{kjt}, and the elasticities and log-transformed multipliers are included in the coefficient vector β. Our goal is to allow β to vary between stores. A conventional HB formulation for a random coefficients regression model (e.g., Andrews et al., 2002; Lenk, DeSarbo, Green, & Young, 1996) is as follows:

$$\begin{aligned}
&q_{kjt} = X_{kjt}\beta_k + \varepsilon_{kjt}\\
&\varepsilon_{kjt} \sim N\left(0, \sigma_k^2\right)\\
&\beta_k \sim N\left(\overline{\beta}, \Lambda\right)\\
&\overline{\beta} \sim N(b_0, D_0)\\
&\Lambda^{-1} \sim W(v_0, S_0)\\
&\sigma_k^{-2} \sim G(a, b)\\
&a = 3,\ b = 1,\ b_0 = 0,\ D_0 = 10^3 I_r, v_0 = r+2,\ S_0 = (1/v_0)I_r
\end{aligned} \tag{1.3}$$

where r is the number of independent variables. We elaborate on this specification below.

Typically, a multivariate normal population distribution is used to specify the heterogeneity between entities, which are stores in our application. The mean vector $\overline{\beta}$ represents the mean marketing mix elasticities and log multipliers in the population of stores, whereas the covariance matrix Λ captures the extent of heterogeneity and the correlation in marketing mix elasticities and multipliers between stores. The error variances σ_k^2 are store-specific.

HB models require priors for the hyperparameters $\overline{\beta}$ and Λ and for the variances σ_k^2. It is customary to assume that σ_k^{-2} is distributed gamma $G(a,b)$, that $\overline{\beta}$ has a multivariate normal prior, $N(b_0, D_0)$, and that Λ^{-1} has a Wishart prior, $W(v_0, S_0)$. The values of a, b, b_0, D_0, v_0, and S_0 were chosen to obtain proper priors that are uninformative and consistent with the values used in previous studies.

[7] Since the HB method in effect estimates store-specific regressions, it is not feasible to include store-specific constants among the explanatory variables since they would be confounded with store-specific intercepts. Thus, the store-specific constants are removed from the HB analysis and are accounted for with an intercept that varies across stores.

2.3. Model with discrete heterogeneity distribution of marketing mix elasticities

FM models incorporate heterogeneity between entities (stores, in this application) by deriving segments having different response coefficients (marketing mix elasticities and multipliers in this application); store responses are assumed to be homogeneous within-segments. The density function for store k's sales of brand j is modeled as a mixture of distributions,

$$\sum_{m=1}^{M} \alpha_m \left[\prod_{t=1}^{T} g\left(q_{kjt}|X_{kjt}, \beta_m, \sigma_m^2\right)\right] \tag{1.4}$$

where the α_m are mixing weights, interpreted as segment sizes, such that $0<\alpha_m<1$ and $\sum_{m=1}^{M} \alpha_m = 1$. Assuming a normally distributed response variable q_{kjt},

$$g\left(q_{kjt}|X_{kjt}, \beta_m, \sigma_m^2\right) = \left(2\pi\sigma_m^2\right)^{-.5} \exp\left[\frac{\left(q_{kjt} - X_{kjt}\beta_m\right)^2}{-2\sigma_m^2}\right]. \tag{1.5}$$

For a sample of K stores, the log likelihood function for the brand j model is

$$\text{LOGL} = \sum_{k=1}^{K} \ln\left[\sum_{m=1}^{M} \alpha_m \left[\prod_{t=1}^{T} g\left(q_{kjt}|X_{kjt}, \beta_m, \sigma_m^2\right)\right]\right]. \tag{1.6}$$

Numerical optimization procedures are used to find the values of the parameters β_m and σ_m^2 that maximize Eq. (1.6) conditional on the number of segments M. A version of the Akaike Information Criterion (AIC3), which penalizes the log likelihood for the number of parameters required to estimate the model, is used to determine the number of segments M (following Andrews & Currim 2003). The number of segments to retain is that for which AIC3 is minimized.

3. Empirical application to market data

3.1. Data

We apply the models to Dutch shampoo scanner data from AC Nielsen, collected weekly at the store-level. Though there are eleven brands, only five engage in promotional activity. We focus our analyses on these five brands, which account for 83.4% of shampoo volume sold. We have 109 weekly observations for 28 stores in a national sample from one large supermarket chain, resulting in 3052 observations. Missing values and/or zero unit sales result in net sample sizes of 3034, 3040, 3040, 3039, and 3040 observations for the five brands.[8] To validate the models, we randomly sample about 75% of the data for use in estimation, reserving the balance of the data for model validation.[9] Descriptive statistics for the shampoo data appear in Table 1.

[8] We do not estimate store-level models because many of the X matrices are singular since stores do not promote certain brands. Hence a single set of predictors cannot be used for store-level models.

[9] This is done so we can estimate weekly constants for all weeks.

Table 1
Descriptive statistics for the shampoo data

	Brand				
	2	3	4	5	8
Number of stores	28	28	28	28	28
Number of weeks	109	109	109	109	109
Number of observations, estimation	2250	2258	2279	2312	2279
Number of observations, validation	784	782	761	727	761
Brand share percentage (by volume)	23.9	9.7	6.3	31.5	12.0
Average percentage price discount	26.3	22.2	14.8	13.0	19.1
% weeks regular price	91.1	81.7	93.6	94.1	93.6
% weeks feature only	1.6	1.7	0.7	0.8	0.8
% weeks display only	1.8	1.9	2.6	1.8	1.8
% weeks feature+display	4.7	2.3	1.9	2.5	1.9
% weeks price promotion without support	2.5	12.6	2.0	1.5	2.3
% weeks non-price promotion	1.6	0.3	1.1	0.7	0.4

3.2. Models and estimation

3.2.1. Model with continuous heterogeneity distribution

Though the specification of the HB model with continuous heterogeneity distribution in the prior section is more general, we found that the model predicted slightly better out of sample (but fitted the sample data less well) when we constrained Λ to be diagonal, indicating no covariances among parameters between stores.[10] Therefore we present the results for the constrained models below. Since the univariate special case of the inverted Wishart distribution is the inverted gamma distribution, the diagonal elements of Λ^{-1} were assumed to have gamma priors, $G(a', b')$ with $a'=3$ and $b'=.01$, corresponding to a conservative mean and standard deviation of .005 for the random coefficient variances (and a coefficient range of about 0.4 between stores). Unlike classical estimation methods, with HB estimation there is very little difference in the computational burden resulting from the assumption of a diagonal or full covariance matrix.

We used standard Gibbs sampling methods for inference. Gibbs sampling is an algorithm for taking draws from the posterior distribution of the parameters. The means of these draws become the estimated elasticities. We allowed 3000 iterations for burn-in, which appeared to be adequate based on plots of the parameters by iteration, and collected results from 3000 iterations for inference. For the empirical application, one model was estimated for each of the five brands (approximately 2300 observations per brand), and each model required about 45 min to run on a personal computer. For the simulation study, the number of observations varied with the number of stores and weeks, and so the estimation time varied by data condition.

[10] This implies, for example, that price sensitivity of one store is unrelated to display sensitivity in another store, a reasonable assumption when the data is on one chain of stores, as in this study, because consumers in store k almost never visit store l, $l \neq k$ and l and k are stores of the same supermarket chain. However, if the data were across multiple chains, a more general model could allow for marketing mix sensitivities to be related.

3.2.2. Model with discrete heterogeneity distribution

Though the specification of the FM model with discrete heterogeneity distribution in the prior section is more general, we found that the models had a better parameter-adjusted fit when only the own- and cross-brand price elasticities and feature multipliers (20 parameters) were allowed to vary between segments, restricting the store and week multipliers to be constant between segments. This saves considerable estimation time as well. For the empirical application, a 2-segment model takes about 15 min to estimate on a personal computer, while a 3-segment model takes up to 40 min. For the simulation, the number of observations varies across data conditions, and so does the estimation time required.

3.3. Results

Table 2 shows the estimation and validation results for the original SCAN*PRO model and its HB and FM extensions, applied to the store-level shampoo scanner data. For the estimation data, the values of the log likelihood function (Logl), R^2, and a version of Akaike's Information Criterion (known as AIC3) are used to measure the degree of model fit.[11] Larger values of Logl and R^2 and smaller values of AIC3 are preferred. Logl and R^2 are absolute measures of fit, while AIC3 penalizes the log likelihood value for the number of parameters required to estimate the model. We use AIC3 because research shows it to be the best model selection criterion for finite mixture regression models (Andrews & Currim, 2003). AIC3 is used primarily to determine the number of store segments to retain for the FM model, so we did not compute such a penalized measure for the HB models. Two- and three-store segment specifications of the FM model are presented.

The estimation results show that the HB model has the best fit for all brands according to both Logl and R^2 (as indicated by the **bold** entries in Table 2). The improvements in R^2 compared to the original SCAN*PRO model range from 4% (brand 2) to 8% (brand 5). The FM model does not offer improvements in R^2 over the original SCAN*PRO model (AIC3 indicates that a 3-store segment model is preferred for brand 2 but that 2-store segment models are preferred for all other brands).

To assess the performance of the model forecasts on validation data, Logl–V and the Root Mean Squared Error, RMSE(Y), of the sales forecasts are computed. Logl–V is quite good as a validation sample criterion for assessing regression models (e.g., Andrews & Currim, 2003). RMSE is computed using the natural metric of the data, volume sales, correcting for bias in the conditional mean predictions due to estimation in the log space (see Van Heerde et al., 2001).

For four of the five brands, the HB model has the best Logl–V values. For brand 5, the HB model fails to improve on the original model but at least does not perform appreciably worse. As for RMSE(Y), the HB models have the lowest values for

[11] The fit and validation statistics are computed only once using the parameter averages from the sampler, not at every iteration of the sampler, in both the study with actual data and in the simulation study that follows (see Andrews, Ainslie, & Currim (2008) for a comparison of these procedures).

Table 2
Shampoo study results for SCAN*PRO models: original (OLS) and heterogeneous (HB and FM) versions

Estimation	SCAN*PRO: Original			SCAN*PRO: HB			SCAN*PRO: FM 2-Segment			SCAN*PRO: FM 3-Segment	
	Logl	R^2	AIC3	Logl	R^2	Logl	R^2	AIC3	Logl	R^2	AIC3
Brand 2	−485	0.86	1441	−156	0.90	−379	0.86	1295	−343	0.86	1288
Brand 3	−1005	0.72	2480	−738	0.77	−930	0.72	2397	−906	0.72	2415
Brand 4	−1206	0.57	2883	−957	0.64	−1132	0.57	2802	−1102	0.58	2807
Brand 5	−201	0.70	874	−6	0.78	−107	0.71	750	−73	0.71	750
Brand 8	−844	0.71	2160	−607	0.76	−776	0.71	2089		−749 0.71	2101
Validation	Logl–V	RMSE(Y)		Logl–V	RMSE(Y)		Logl–V	RMSE(Y)		Logl–V	RMSE(Y)
Brand 2	−311	3.44		−268	3.34		−329	3.39		−340	3.47
Brand 3	−474	2.59		−437	2.40		−474	2.32		−473	2.50
Brand 4	−500	1.01		−462	0.91		−489	0.94		−498	1.17
Brand 5	−139	3.57		−141	3.57		−146	3.49		−155	7.40
Brand 8	−406	1.67		−367	1.61		−403	1.76		−421	1.86

Logl is the value of the log likelihood function on the estimation sample.
Logl–V is the value of the log likelihood function on the validation sample.
R^2 is the percentage of variance explained.
RMSE(Y) is the root mean squared error of validation sample sales forecasts, in the metric of the raw data.
AIC3 = −2 Logl + 3P, where P is the number of parameters required for estimation.

three of the five brands (2, 4, and 8), while the 2-store segment FM models have the lowest RMSE values for the other two brands (3 and 5). However, the magnitude of reduction in RMSE(Y) for the five brands, arising even from the use of the best heterogeneous model for each brand, over the original SCAN*PRO model is small (3.34 vs. 3.44; 2.32 vs. 2.59; 0.91 vs. 1.01; 3.49 vs. 3.57; 1.61 vs. 1.67). If only the HB or FM model were considered in place of the original model (not both), the differences in RMSE(Y) would be smaller.

While our analysis of Dutch store-level scanner data for the shampoo product category provides important insights into whether consideration of heterogeneity in marketing mix effects between stores improves fit and forecasting accuracy in a real world setting, there are two shortcomings. First, since the true store-level marketing mix sensitivities are unknown, it is unclear whether the store-level marketing mix elasticity estimates from the HB or FM heterogeneous models are more or less accurate than the corresponding OLS based estimates of the original homogeneous SCAN*PRO model. Second, with real world data we are limited to one dataset or setting, which is characterized by a certain level of heterogeneity in marketing mix effects between and within stores, a certain number of stores, a certain time period, etc. Thus, it is not possible to determine whether the benefits of modeling store-level heterogeneity in marketing mix elasticities depend on the amount of heterogeneity in marketing mix effects between and within stores, the number of stores, the number of weeks, etc., or across a variety of settings. To further investigate these issues we designed the following simulation.

4. Simulation

4.1. Description

In this section we first describe the structural and data factors varied in the simulation and provide rationale for the selection of levels for each factor. Second, we briefly provide expectations about which model (continuous heterogeneity (HB), discrete heterogeneity (FM), or no heterogeneity (OLS)) is expected to be preferred in each experimental condition. Third, we present the performance measures used to gauge model performance.

4.1.1. Factors

We vary 3 structural factors (S) and 2 data factors (D) which are expected to impact the fit, forecasting accuracy, and/or accuracy of the marketing mix elasticities estimated by the SCAN*PRO model:

- S1. Within-segment variance (WS) in marketing mix effects: none (0) or high (0.25);
- S2. Between-segment variance (BS), operationalized as separation or distance between store segment means on effects of the marketing mix: none (0) or high (1.0);
- S3. Error variance: base or two times base;
- D1. Number of stores: 30 or 60;
- D2. Number of weeks: 60 or 120;
- R. Replications: 5 datasets based on each of 5 actual shampoo brands.

4.1.2. Selection of levels

Since within- and between-segment heterogeneity are manipulated in a factorial design, the resulting datasets have, in equal numbers, (i) a unimodal distribution with no variance, (ii) a unimodal distribution with large variance, (iii) a bimodal distribution with each mode having no variance, and (iv) a bimodal distribution with each mode having a large variance. Levels of within- and between-segment variance were chosen to be consistent with previous studies in the literature (in particular Vriens, Wedel, & Wilms 1996; Andrews et al., 2002). The segment sizes vary uniformly in the range of 20–80%, so these conditions encompass a very wide range of market scenarios.

The manipulation of the error variance has to be done in conjunction with the scale of the dependent variable in order to produce reasonable R^2 values (similar to what is observed in practice). We chose the levels of the error variance to produce R^2 values in the 60–85% range, which is consistent with the empirical example and our experience with store-level response models. Finally, the manipulation of the number of stores and weeks is based on the nature of datasets we have encountered in practice and in the literature.

4.1.3. Expectations

The original homogeneous SCAN*PRO model (OLS) should perform well when within-store segment variance in marketing mix effects is zero (WS = 0), and when there is no separation between store segment means of marketing mix effects (BS = 0). FM should be preferred when within-store segment variance on marketing mix effects is zero (WS = 0) and there is high separation between store segment means of marketing mix effects (BS = 1.0). HB should be preferred when within-store segment variance on marketing mix effects is high (WS = 0.25) and there is no separation between store segment means of marketing mix effects (BS = 0). Which model specification will be preferred when there is high within-store segment variance on marketing mix effects (WS = 0.25) and high separation between store segment means (BS = 1.0) is not clear and remains an important empirical question.

Though all models should perform better with more stores, we might expect the heterogeneous models (FM and HB) to perform better relative to the original model when the number of stores is larger since both techniques are thought to perform less well with smaller sample sizes, especially when the number of observations per entity is also smaller (e.g., Andrews et al., 2002). And though we would normally expect the models to perform better when the number of weeks is larger, it is not clear in this case since the number of weekly constants requiring estimation will increase as well.[12] And though all models should perform better with less error variance, it is not clear that having more error variance will affect the models in different ways. Andrews et al. (2002) observed that an HB conjoint model had poor parameter recovery overall when the error variance was large.

4.1.4. Models and performance measures

For each combination of the above factors, we generate five replications corresponding to the five actual brands in the shampoo product category, resulting in a total of 2 × 2 × 2 × 2 × 2 × 5 = 160 artificial datasets. For each dataset, the original SCAN*PRO model and the HB and FM extensions of the model are estimated for a total of 480 model–dataset combinations. The performance measures calculated for each model–dataset combination include (i) RMSE(β), the root mean squared error between the actual and estimated elasticities and log multipliers for the store-level own price, own feature, own display, and own feature plus display coefficients;[13] (ii) Logl, the estimation sample log likelihood (higher is better); (iii) R^2, the percentage of variation in sales explained by the predictors for the estimation sample, (iv) Logl–V, the validation sample log likelihood, (v) R^2–V, the percentage of variation in sales explained by the predictors for the validation sample, and (vi) RMSE (Y), the root mean square error between the actual and estimated sales on the validation sample. Logl and Logl–V are expressed on a per-observation basis for improved interpretation. AIC3 is used to determine the number of segments to retain for the FM models.

4.2. Results

Table 3 shows the results of six ANOVAs or meta-analyses conducted across all 160 datasets, one for each performance measure. The independent variables are in the form of dummy variables which take on a value of 1 for the higher level (of number of stores, weeks, within- and between-segment heterogeneity, and error). The base level for the HB and FM dummies is the OLS model. Our initial observation is that there are a fair number of statistically significant effects indicating that our manipulations worked well. Number of stores, number of weeks, between-segment heterogeneity, and the FM model dummies are significant for 3 of 6 measures, error variance and the HB model dummies are significant for 5 of 6 performance measures, and within-segment heterogeneity is significant for all measures. The signs of all significant effects are in the expected direction.

The HB model provides a statistically significant improvement over the original SCAN*PRO model, both on the estimation (Logl, R^2) and validation samples (Logl–V, R^2–V, RMSE(Y)), but not parameter accuracy (RMSE(β)). The improvement in R^2 and R^2–V are 0.071 and 0.022, respectively. The FM model, on average, provides an improvement over the original OLS estimation on the estimation (Logl, R^2) and the validation sample (RMSE(Y)), but not parameter accuracy (RMSE(β)). However, the magnitude of improvement in R^2 is small (0.018). We provide an explanation for these findings in the next section.

In order to get insight into how different models perform under different structural and data conditions, we turn to Table 4, which shows the means of each performance measure (the best values are highlighted in **bold** font) across the eight data conditions (No or High Within-segment variance (WS) × No or High Between-segment variance (BS) × Low or High Error (E)). The means are presented in four panels, each having two data conditions. The panels correspond to our hypotheses regarding which model is expected to perform better or expected to be the correct model in certain data conditions.

Data conditions 1 and 2 (panel A) comprise low within-segment variance and low between-segment variance, so that OLS is expected to perform best. OLS is the preferred model

[12] We generate the data such that each week has a unique effect and specify the models accordingly, but an alternative strategy would be to generate “week of the year” effects, such that only 51 constants would be estimated.

[13] The cross effects were not included in the calculation of RMSE(b) because the vast majority of cross effects were not statistically significant with the shampoo category models.

Table 3
ANOVA results from simulation across 160 datasets

	RMSE(β)				Logl				R^2			
	Estimate	S.E.	*t*-value	*p*-value	Estimate	S.E.	*t*-value	*p*-value	Estimate	S.E.	*t*-value	*p*-value
Constant	0.568	0.058	9.822	0.000	−0.518	0.014	−37.295	0.000	0.772	0.006	122.436	0.000
Number of stores	−0.090	0.041	−2.199	0.028	−0.007	0.010	−0.700	0.484	0.005	0.004	1.132	0.258
Number of weeks	−0.254	0.041	−6.204	0.000	0.001	0.010	0.091	0.928	−0.008	0.004	−1.889	0.059
Within-segment Heterogeneity	0.453	0.041	11.082	0.000	−0.462	0.010	−46.994	0.000	−0.107	0.004	−23.999	0.000
Between-segment Heterogeneity	0.120	0.041	2.938	0.003	−0.018	0.010	−1.828	0.068	−0.001	0.004	−0.221	0.825
Error variance	0.003	0.041	0.082	0.935	−0.243	0.010	−24.705	0.000	−0.097	0.004	−21.792	0.000
HB model dummy	0.033	0.050	0.666	0.506	0.121	0.012	10.079	0.000	0.071	0.005	13.042	0.000
FM2 model dummy	−0.008	0.050	−0.151	0.880	0.028	0.012	2.309	0.021	0.018	0.005	3.219	0.001
R^2	0.27				0.86				0.72			
Adjusted R^2	0.26				0.86				0.72			

	Logl−V				R^2−V				RMSE(Y)−Validation			
	Estimate	S.E.	*t*-value	*p*-value	Estimate	S.E.	*t*-value	*p*-value	Estimate	S.E.	*t*-value	*p*-value
Constant	−0.598	0.014	−43.112	0.000	0.726	0.007	97.419	0.000	0.407	0.007	59.254	0.000
Number of stores	0.032	0.010	3.233	0.001	0.031	0.005	5.976	0.000	0.005	0.005	1.088	0.277
Number of weeks	0.019	0.010	1.957	0.051	0.007	0.005	1.309	0.191	0.003	0.005	0.540	0.590
Within-segment Heterogeneity	−0.484	0.010	−49.308	0.000	−0.134	0.005	−25.345	0.000	0.254	0.005	52.272	0.000
Between-segment Heterogeneity	−0.028	0.010	−2.860	0.004	−0.005	0.005	−0.859	0.391	0.010	0.005	1.991	0.047
Error variance	−0.236	0.010	−24.060	0.000	−0.106	0.005	−20.199	0.000	0.131	0.005	27.002	0.000
HB model dummy	0.034	0.012	2.801	0.005	0.022	0.006	3.435	0.001	−0.070	0.006	−11.833	0.000
FM2 model dummy	0.009	0.012	0.711	0.477	0.007	0.006	1.104	0.270	−0.017	0.006	−2.819	0.005
R^2	0.87				0.70				0.88			
Adjusted R^2	0.86				0.70				0.88			

according to RMSE(β), having the lowest bias in 7 out of 8 experimental conditions. However, on the basis of fit and forecasting accuracy, OLS appears to be the best only according to R^2–V. HB appears to be the best model according to other measures. Even with R^2–V, the differences between models are very small.

Data conditions 3 and 4 (panel B) comprise low within-segment variance and high between-segment variance, so that FM is expected to be the correct model. FM does in fact appear to be the best model according to RMSE(β) and R^2–V, while HB is preferred according to the other measures. Even with RMSE(β) and R^2–V, the differences between the correct FM model and other models are small.

Data conditions 5 and 6 (panel C) comprise high within-segment variance and low between-segment variance, so that HB is expected to be the correct model. HB is indicated to be the best model according to all measures of fit and forecasting accuracy, but importantly, not according to the parameter accuracy measure RMSE(β).

Data conditions 7 and 8 (panel D) comprise high within-segment variance and high between-segment variance, so that it is not clear which model will perform better (although clearly the OLS model should not perform best). HB is found to be the preferred model on most fit, forecasting accuracy, and parameter accuracy measures. However, again the differences across models are small.

Based on the analysis of the four panels, we observe only weak confirmation of our *a priori* expectations of model performance, and even when expectations are confirmed, the improvement resulting from the use of a heterogeneous model is fairly modest. Consistent with Table 3, we see that using a heterogeneous model does not improve the accuracy of the elasticities. At this juncture it is important to note that the judgment on whether differences in fit, forecasting accuracy and parameter recovery are considered small or large enough to necessitate the more complicated estimation procedure could be subjective. Although we perceive the differences to be minor, other researchers may perceive the differences as more substantial than we do. Our results are presented in a way that readers can make their own judgments.

5. Discussion and implications for marketing modelers and managers in commercial settings

The homogeneous SCAN*PRO model of store-level product sales is one of the most widely applied models in industry. Hence, our main objective is to guide work on this model in commercial settings. In this paper we investigate whether allowance for store-level heterogeneity in marketing mix elasticities using HB and FM methods improves the performance of the original model, which is estimated using simpler OLS regression.

The main result based on Dutch store-level scanner data is that the HB model has the best estimation sample fit for all brands, and both heterogeneous models (HB and FM) offer advantages in prediction accuracy over the original homogeneous model estimated using OLS. In the simulation study, both HB and FM models provide improvements over the original model in terms of fit and prediction accuracy, but not parameter accuracy. However, the magnitude of the improvements in fit and prediction accuracy are fairly modest, so little may be lost by using the original SCAN*PRO in commercial settings. Since our simulation manipulated between and within-segment variance in marketing mix effects and error variance at fairly extreme levels, it is likely that in most practical or intermediate settings the advantages of using heterogeneous versions of SCAN*PRO could be

Table 4
Simulation results on performance of models in each of 8 experimental data conditions

Panel A

		Data condition 1: Low WS, Low BS, Low E (OLS correct)				Data condition 2: Low WS, Low BS, High E (OLS correct)			
		# of Stores=30	# of Stores=60	# of Weeks=60	# of Weeks=120	# of Stores=30	# of Stores=60	# of Weeks=60	# of Weeks=120
RMSE(β)	OLS	0.31	**0.17**	**0.24**	**0.24**	**0.21**	**0.42**	**0.47**	**0.17**
	HB	**0.23**	0.26	0.25	0.24	0.32	0.54	0.68	0.18
	FM	0.35	0.20	0.27	0.28	0.38	0.43	0.63	0.18
	Means	0.30	0.21	0.25	0.25	0.31	0.46	0.59	0.18
Logl	OLS	−0.43	−0.45	−0.44	−0.44	−0.80	−0.80	−0.79	−0.80
	HB	**−0.37**	**−0.38**	**−0.40**	**−0.35**	**−0.71**	**−0.71**	**−0.72**	**−0.70**
	FM	−0.43	−0.44	−0.43	−0.44	−0.79	−0.79	−0.78	−0.80
	Means	−0.41	−0.42	−0.42	−0.41	−0.76	−0.77	−0.76	−0.77
R^2	OLS	0.82	0.80	0.82	0.79	0.67	0.67	0.68	0.66
	HB	**0.85**	**0.83**	**0.85**	**0.83**	**0.71**	**0.71**	**0.71**	**0.70**
	FM	0.82	0.80	0.83	0.80	0.67	0.68	0.68	0.66
	Means	0.83	0.81	0.83	0.81	0.68	0.69	0.69	0.68
Logl–V	OLS	−0.51	−0.49	−0.51	−0.50	−0.88	−0.83	−0.86	−0.85
	HB	**−0.50**	**−0.48**	**−0.50**	**−0.48**	**−0.86**	**−0.81**	**−0.85**	**−0.83**
	FM	−0.52	−0.49	−0.51	−0.50	−0.89	−0.84	−0.88	−0.85
	Means	−0.51	−0.49	−0.51	−0.49	−0.88	−0.83	−0.86	−0.84
R^2–V	OLS	**0.79**	**0.78**	**0.79**	**0.78**	0.60	**0.64**	**0.61**	0.63
	HB	0.78	0.78	0.79	0.77	**0.60**	0.64	0.61	**0.63**
	FM	0.78	0.78	0.79	0.77	0.59	0.64	0.60	0.63
	Means	0.78	0.78	0.79	0.77	0.60	0.64	0.61	0.63
RMSE(Y)	OLS	0.38	0.38	0.38	0.38	0.54	0.54	0.54	0.54
	HB	**0.34**	**0.35**	**0.35**	**0.35**	**0.51**	**0.51**	**0.51**	**0.51**
	FM	0.37	0.38	0.37	0.38	0.54	0.54	0.53	**0.54**
	Means	0.36	0.37	0.37	0.37	0.53	0.53	0.52	0.53

Panel B

		Data condition 3: Low WS, High BS, Low E (FM correct)				Data condition 4: Low WS, High BS, High E (FM correct)			
		# of Stores=30	# of Stores=60	# of Weeks=60	# of Weeks=120	# of Stores=30	# of Stores=60	# of Weeks=60	# of Weeks=120
RMSE(β)	OLS	0.70	0.63	0.75	0.58	0.63	0.61	0.64	0.60
	HB	0.66	0.57	**0.69**	0.54	0.66	0.64	**0.57**	0.72
	FM	**0.64**	**0.48**	0.72	**0.40**	**0.59**	**0.55**	0.61	**0.52**
	Means	0.67	0.56	0.72	0.51	0.63	0.60	0.61	0.62
Logl	OLS	−0.51	−0.50	−0.50	−0.51	−0.82	−0.83	−0.82	−0.83
	HB	**−0.39**	**−0.38**	**−0.41**	**−0.36**	**−0.72**	**−0.72**	**−0.73**	**−0.71**
	FM	−0.45	−0.44	−0.44	−0.45	−0.80	−0.81	−0.80	−0.81
	Means	−0.45	−0.44	−0.45	−0.44	−0.78	−0.78	−0.78	−0.78
R^2	OLS	0.79	0.79	0.80	0.79	0.65	0.67	0.66	0.66
	HB	**0.84**	**0.84**	**0.84**	**0.84**	**0.71**	**0.72**	**0.71**	**0.72**
	FM	0.81	0.81	0.82	0.81	0.66	0.68	0.67	0.68
	Means	0.82	0.82	0.82	0.81	0.67	0.69	0.68	0.69
Logl–V	OLS	−0.58	−0.55	−0.57	−0.55	−0.90	−0.88	−0.89	−0.89
	HB	**−0.54**	−0.50	−0.55	**−0.49**	**−0.87**	**−0.85**	**−0.87**	**−0.85**
	FM	−0.55	**−0.50**	**−0.55**	−0.50	−0.89	−0.86	−0.89	−0.87
	Means	−0.56	−0.52	−0.56	−0.52	−0.88	−0.86	−0.88	−0.87
R^2–V	OLS	0.76	0.77	0.76	0.76	0.59	0.64	0.60	0.63
	HB	0.77	0.79	0.77	0.78	0.60	**0.66**	**0.61**	**0.64**
	FM	**0.77**	**0.79**	**0.78**	**0.79**	**0.60**	0.65	0.61	0.64
	Means	0.77	0.78	0.77	0.78	0.59	0.65	0.61	0.64
RMSE(Y)	OLS	0.40	0.40	0.40	0.40	0.55	0.56	0.55	0.56
	HB	**0.35**	**0.35**	**0.35**	**0.35**	**0.51**	**0.51**	**0.51**	**0.51**
	FM	0.38	0.38	0.38	0.38	0.54	0.55	0.54	0.55
	Means	0.38	0.38	0.38	0.38	0.53	0.54	0.53	0.54

(continued on next page)

Table 4 (*continued*)

Panel C

		Data condition 5: High WS, Low BS, Low E (HB correct)				Data condition 6: High WS, Low BS, High E (HB correct)			
		# of Stores = 30	# of Stores = 60	# of Weeks = 60	# of Weeks = 120	# of Stores = 30	# of Stores = 60	# of Weeks = 60	# of Weeks = 120
RMSE(β)	OLS	**1.16**	**0.70**	**1.14**	0.72	0.88	0.90	**1.15**	0.63
	HB	1.46	0.76	1.52	**0.70**	**0.84**	1.00	1.25	**0.59**
	FM	1.17	0.70	1.16	0.72	0.87	**0.90**	1.15	0.62
	Means	1.27	0.72	1.27	0.72	0.86	0.93	1.19	0.61
Logl	OLS	−1.04	−1.06	−1.04	−1.06	−1.17	−1.19	−1.18	−1.18
	HB	**−0.90**	**−0.89**	**−0.91**	**−0.89**	**−1.06**	**−1.05**	**−1.06**	**−1.04**
	FM	−1.00	−1.02	−1.00	−1.03	−1.15	−1.16	−1.15	−1.16
	Means	−0.98	−0.99	−0.98	−0.99	−1.13	−1.13	−1.13	−1.13
R^2	OLS	0.64	0.63	0.64	0.63	0.57	0.58	0.58	0.57
	HB	**0.73**	**0.73**	**0.73**	**0.74**	**0.65**	**0.68**	**0.66**	**0.67**
	FM	0.67	0.65	0.67	0.65	0.59	0.60	0.60	0.59
	Means	0.68	0.67	0.68	0.67	0.60	0.62	0.61	0.61
Logl–V	OLS	−1.12	−1.10	−1.10	−1.11	−1.26	−1.22	−1.24	−1.24
	HB	**−1.08**	**−1.05**	**−1.07**	**−1.06**	**−1.23**	**−1.20**	**−1.22**	**−1.21**
	FM	−1.12	−1.09	−1.11	−1.10	−1.25	−1.21	−1.24	−1.23
	Means	−1.10	−1.08	−1.10	−1.09	−1.25	−1.21	−1.24	−1.22
R^2–V	OLS	0.58	0.59	0.59	0.58	0.51	0.55	0.53	0.52
	HB	**0.62**	**0.63**	**0.61**	**0.63**	**0.54**	**0.57**	**0.55**	**0.56**
	FM	0.58	0.60	0.58	0.60	0.52	0.55	0.54	0.54
	Means	0.59	0.61	0.59	0.60	0.52	0.56	0.54	0.54
RMSE(Y)	OLS	0.68	0.70	0.68	0.70	0.78	0.80	0.79	0.79
	HB	**0.59**	**0.59**	**0.60**	**0.59**	**0.70**	**0.70**	**0.71**	**0.69**
	FM	0.66	0.67	0.65	0.68	0.76	0.78	0.76	0.77
	Means	0.64	0.65	0.64	0.65	0.75	0.76	0.75	0.75

Panel D

		Data condition 7: High WS, High BS, Low E (Not clear)				Data condition 8: High WS, High BS, High E (Not clear)			
		# of Stores = 30	# of Stores = 60	# of Weeks = 60	# of Weeks = 120	# of Stores = 30	# of Stores = 60	# of Weeks = 60	# of Weeks = 120
RMSE(β)	OLS	0.88	0.89	0.91	0.86	1.02	0.81	0.98	0.85
	HB	0.99	0.91	1.11	**0.80**	**0.86**	**0.77**	**0.86**	**0.77**
	FM	**0.87**	**0.87**	**0.90**	0.84	1.01	0.81	0.99	0.83
	Means	0.91	0.89	0.97	0.83	0.96	0.80	0.94	0.82
Logl	OLS	−1.07	−1.08	−1.08	−1.07	−1.19	−1.20	−1.19	−1.20
	HB	**−0.90**	**−0.89**	**−0.91**	**−0.88**	**−1.06**	**−1.05**	**−1.06**	**−1.05**
	FM	−1.02	−1.04	−1.03	−1.04	−1.16	−1.18	−1.16	−1.18
	Means	−1.00	−1.00	−1.01	−0.99	−1.13	−1.15	−1.14	−1.15
R^2	OLS	0.62	0.63	0.63	0.62	0.62	0.58	0.58	0.57
	HB	**0.73**	**0.75**	**0.74**	**0.74**	**0.74**	**0.68**	**0.67**	**0.67**
	FM	0.65	0.66	0.67	0.64	0.64	0.60	0.61	0.59
	Means	0.67	0.68	0.68	0.67	0.67	0.62	0.62	0.61
Logl–V	OLS	−1.17	−1.14	−1.16	−1.15	−1.28	−1.25	−1.27	−1.26
	HB	**−1.11**	**−1.08**	**−1.11**	**−1.08**	**−1.24**	**−1.21**	**−1.24**	**−1.21**
	FM	−1.16	−1.12	−1.15	−1.13	−1.27	−1.24	−1.27	−1.24
	Means	−1.14	−1.11	−1.14	−1.12	−1.26	−1.23	−1.26	−1.24
R^2–V	OLS	0.54	0.59	0.58	0.56	0.50	0.55	0.52	0.52
	HB	**0.59**	**0.64**	**0.61**	**0.62**	**0.54**	**0.58**	**0.55**	**0.57**
	FM	0.56	0.60	0.59	0.57	0.51	0.55	0.52	0.54
	Means	0.56	0.61	0.59	0.59	0.51	0.56	0.53	0.54
RMSE(Y)	OLS	0.70	0.71	0.71	0.70	0.80	0.81	0.80	0.81
	HB	**0.59**	**0.59**	**0.60**	**0.58**	**0.70**	**0.70**	**0.71**	**0.70**
	FM	0.67	0.69	0.68	0.68	0.77	0.79	0.77	0.79
	Means	0.66	0.66	0.66	0.66	0.76	0.77	0.76	0.77

WS is within-segment heterogeneity, BS is between-store heterogeneity, E is error.

even smaller. At this point it is important to note that our empirical and simulation studies were conducted in the SCAN*PRO model setting to investigate whether the allowance for heterogeneity improves performance of the widely used model on fit, forecasting accuracy, and parameter recovery. Therefore, our conclusion should be restricted to the SCAN*PRO model only and should not immediately be extended to other settings.

In particular, our empirical result is different from other studies comparing FM and HB models applied to scanner panel choice data and conjoint ratings data (e.g., Andrews et al., 2002 for choice models; Andrews et al., 2002 (AAC) for conjoint regression-based models), which have shown the two methods to be equally valuable in modeling heterogeneity and significantly better than their homogeneous counterparts. The AAC study is especially relevant because it analyzes data having a normally-distributed dependent variable using OLS regression, as we do in this study.

One hypothesis as to why we did not find a significant improvement in the context of the SCAN*PRO model is that the number of parameters required for the traditional model is extremely large, creating a difficult environment for recovering heterogeneity in parameters and resulting in decreasing returns from using additional parameters to explain more variation in sales. For the 60-store, 120-week condition, the SCAN*PRO model requires 199 parameters, in contrast to the 13 parameters required for the conjoint models examined in the AAC study. To test our hypothesis, we re-generated the five brand-level datasets in each of two conditions (High WS, High BS, Low E, 60 stores, 120 weeks and High WS, High BS, High E, 60 stores, 120 weeks), except that no store- or week-specific effects were generated. Likewise, we removed the store- and week-specific constants from the OLS, HB, and FM models, resulting in only 20 parameters plus an intercept. The results of this exercise are shown in Table 5. There we see far more significant improvements in the accuracy of elasticities and massive improvements in fit and forecasting accuracy. Of particular interest is the improvement in RMSE(β). Comparing the OLS and HB-estimated SCAN*PRO models, we see that using the HB model results in almost a 20% reduction in RMSE(β) for the low error variance condition and almost a 34% reduction in the high error variance condition. Unfortunately, commercial applications of SCAN*PRO will require the week and store constants that we assumed away for this exercise, so the main conclusion of our study is still that there is little incentive for developing heterogeneous versions of SCAN*PRO for commercial applications.

There is a parallel between our findings and the empirical observation in many hazard-rate models. That is, if one works with a semi-parametric estimation of the baseline hazard (i.e. step-wise, period-by-period approximation), the estimated variance for the mixing distribution (capturing the extent of heterogeneity) most often converges to zero. For example, Vanhuele, Dekimpe, Sharma, and Morrison (1995) found that the estimated variance for the mixing distribution, which captures the heterogeneity in the hazard-rate model, becomes small when they include time dummy variables and all the available covariates.

Table 5
Model performance with reduced parameter counts

	High WS, High BS, Low E, 60 stores, 120 weeks			High WS, High BS, High E, 60 stores, 120 weeks		
	OLS	HB	FM2	OLS	HB	FM2
RMSE(β)	1.03	**0.83**	1.04	1.12	**0.74**	1.11
Logl	−1.43	**−0.93**	−1.23	−1.50	**−1.09**	−1.32
R^2	0.08	**0.65**	0.38	0.07	**0.58**	0.34
Logl–V	−1.43	**−1.03**	−1.24	−1.50	**−1.18**	−1.34
R^2–V	0.07	**0.58**	0.37	0.06	**0.51**	0.32
RMSE(Y)	1.01	**0.62**	0.83	1.08	**0.73**	0.91

In this study we limited ourselves to one store-level sales model, though it is well established and documented with a sizeable number of commercial and academic applications. We also limited ourselves to the variables and specification proposed in the basic model; recent refinements to the basic specification are not considered (Van Heerde et al., 2000, 2001).

One potentially valuable direction for future research is to investigate more complex models. For example, one could estimate a model that permits pre- and post-promotion dips or a model that allows for complex nonlinearities and interactions in the deal effect curve (Van Heerde et al., 2000, 2001). Second, one could allow for non-zero covariances across stores in the error term of the sales equation. This is feasible when the number of observations per store is the same and there are no missing observations. Third, it might be useful to investigate non-zero covariances among parameters as well, as this could affect the relative benefits of using FM and HB-estimated SCAN*PRO models. Finally, it would be valuable to further address our hypothesis on the effects of the number of intercepts on the model performance by estimating the SCAN*PRO models with a reduced number of intercepts, for example, assuming the same weekly effects across years or seasonal patterns, and where store constants are replaced with the effect of intrinsic store characteristics, e.g., store size. We are hopeful that our work will encourage such efforts.

6. Afterthought

Our results will undoubtedly be controversial. On the one hand, some scholars who have strong prior beliefs about heterogeneity are likely to conclude quickly that we are advocating the use of a mis-specified model over a more correctly specified model. They may argue that just because the correct model is more complex and the resulting benefits may be small according to some that this is not sufficient justification to ignore heterogeneity. Some are likely to look at differences we have deemed small and judge them to be large, especially when the differences favor the heterogeneous model. These conclusions and arguments are expected, not only given their experience and orientation, but also because the empirical results in the household and store data modeling literatures, to be revisited below, have supported the use of the heterogeneous model over its homogeneous counterpart.

Other readers will recognize that we are interested in studying the benefits of capturing store-level heterogeneity only in the context of the widely used SCAN*PRO model of store sales, not store sales models in general (including for example, logit-based models of store sales). They will recognize that we neither developed the homogeneous SCAN*PRO model nor have we published work on the value of heterogeneous store models, so we do not have vested interests. We are only interested in empirically investigating the performance of heterogeneous and homogeneous versions of the SCAN*PRO model across a variety of conditions that may affect comparative performance of the models. These readers are likely to conclude that the paper offers valuable insights for practitioners and commercial users of the SCAN*PRO model.

We attempt to summarize the previous and current findings in different settings by distinguishing factors inherent to (1) household-level data and associated models, (2) general store-level data and associated models, and (3) the unique SCAN*PRO specification. This research follows the research tradition of comparing the performance of alternative models under varying degrees of household heterogeneity. For example, Andrews et al., 2002 analyze household-level choice data using logit analysis, and Andrews et al. (2002) analyze individual-level preference data using OLS regression in a conjoint analysis setting. Given the different findings in this study (that a homogeneous model recovers parameters as well as the heterogeneous model) versus those works (which find that different specifications of heterogeneous models perform equivalently but better than a homogeneous model), it is worthwhile to discuss when or under which condition it is likely better to adopt the more complex heterogeneous model over the simpler homogeneous model and vice versa. Since the two earlier works focused on consumer heterogeneity and this work focuses on store heterogeneity, comparing these works involves a stretch beyond the bounds of the actual findings in each study and is somewhat more speculative than a standard discussion.

Similarly, our findings are also in contrast to the findings from the store heterogeneity literature, which have shown significant benefits from modeling across-store heterogeneity (though none use SCAN*PRO, e.g., Hoch et al., 1995; Montgomery, 1997; Montgomery & Rossi, 1999). Hoch et al. (1995) search for determinants of store-level price sensitivities among variables describing the demographics of the trading area and competitive variables. We did not investigate the effects of such covariates on store-level sensitivities, which could have hindered the heterogeneous models.

Montgomery (1997) fits a model similar in concept to the SCAN*PRO model but does not use weekly intercepts. Recall that we found that the benefits of modeling across-store heterogeneity in simulated data were much more apparent when there were no week-specific effects, reducing the parameter count considerably. However, we have not applied a model without weekly intercepts to real data. Perhaps the benefits of modeling heterogeneity would be much larger since there would be more unexplained variation that can be explained by the heterogeneity specification. This is not obvious, however, since the model captures across-store heterogeneity, and the missing effects would be week-specific. Montgomery (1997) and Montgomery and Rossi (1999) analyze 11 refrigerated orange juice products simultaneously, giving rise to a covariance matrix of error terms, whereas we analyze the brands separately. It is not clear what effect this might have on the recovery of heterogeneity, but it is an interesting question for future research to ponder.

There is another potential explanation of our findings relative to previous studies. Store-level data is different from household level data in that (a) the number of stores (say about 50) is typically much smaller than the number of households considered (which typically runs from a few hundred up to over a thousand households per application), and (b) the number of data points per household is typically much lower than the number of observations per store. This makes it more possible to estimate store-level constants, whereas estimating fixed effects for households is much more difficult. At best, some proxies such as household preferences based on the (few) purchases in an initialization period can be used. Hence, while allowing for heterogeneity in household level models serves to capture differences in household-level fixed effects, store-level models may already capture these in a systematic fashion, so that the additional advantage from allowing for heterogeneity across stores only pertains to marketing mix effectiveness. This may be the case even with the HB models, for which the store-specific constants were removed and replaced with a distribution of store intercepts, since the larger quantity of data available at the store-level vs. the household level could have resulted in quite accurate estimates of store intercepts.

Concerning the SCAN*PRO model specifics, one question that remains is to what extent does the use of separate constants for every week in the data period and the consequent use of degrees of freedom hamper the possibility of assessing true differences in stores' promotion effectiveness? Would the use of more judicious and less parameter consuming seasonal patterns (for instance, week of year effects, holiday weeks, end of month weeks, and other patterns that recur on a yearly basis) not allow for a better recovery of heterogeneity in promotion response across stores using HB or FM models? And, would the adoption of such heterogeneous models without separate constants for each week lead to improved forecasts for future periods compared to the traditional SCAN*PRO model?

We believe that this is a fruitful avenue for future research, and hope that our work motivates such research.

References

Andrews, R. L., Ainslie, A., & Currim, I. S. (2002, November). An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *Journal of Marketing Research*, *39*, 479−487.

Andrews, R. L., Ainslie, A., & Currim, I. S. (2008, January). On the recoverability of choice behaviors with random coefficients choice models in the context of limited data and unobserved effects. *Management Science*, *54*, 83−99.

Andrews, R. L., Ansari, A., & Currim, I. S. (2002, February). Hierarchical Bayes versus finite mixture conjoint analysis models: a comparison of fit, prediction, and partworth recovery. *Journal of Marketing Research*, *39*, 87−98.

Andrews, R. L., & Currim, I. S. (2003, December). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, *20*, 315−322.

Christen, M., Gupta, S., Porter, J. C., Staelin, R., & Wittink, D. R. (1997, August). Using market-level data to understand the promotion effects in a nonlinear model. *Journal of Marketing Research*, *36*, 322−334.

Foekens, E. W., Leeflang, P. S. H., & Wittink, D. R. (1994). A comparison and an exploration of the forecasting accuracy of a loglinear model at different levels of aggregation. *International Journal of Forecasting*, *10*(2), 245−261.

Hoch, S. J., Kim, B. D., Montgomery, A. L., & Rossi, P. E. (1995, February). Determinants of store-level price elasticity. *Journal of Marketing Research*, *32*, 17−29.

Horvath, C., Leeflang, P. S. H., Wieringa, J. E., & Wittink, D. R. (2005). Competitive reaction- and feedback effects based on VARX models of pooled store data. *International Journal of Research in Marketing*, *22*, 415−426.

Leeflang, P. S. H., & Wittink, D. R. (2000). Building models for marketing decisions: past, present and future. *International Journal of Research in Marketing*, *17*, 105−126.

Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, *15*(2), 173−191.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science*, *16*(4), 315−337.

Montgomery, A. L., & Rossi, P. E. (1999). Estimating price elasticities with theory-based priors. *Journal of Marketing Research, 36*, *4*, 413−423.

Moore, W. (2004). A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing*, *21*(3), 299−312.

Otter, T., Tüchler, R., & Frühwirth-Schnatter, S. (2004). Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. *International Journal of Research in Marketing*, *21*(3), 285−297.

Stremersch, S., & Lehmann, D. (2007). Editorial. *International Journal of Research in Marketing*, *24*(1), 1−2.

Roberts, J., Kayande, U., & Stremersch, S. (2007). *Impact of marketing science on practice.* European Marketing Association Conference Presentation.

Vanhuele, M., Dekimpe, M., Sharma, S., & Morrison, D. (1995, April). Probability models for duration: the data don't tell the whole story. *Organizational Behavior and Human Decision Processes*, *62*, 1−13.

Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2000, August). The estimation of pre- and post-promotion dips with store-level scanner data. *Journal of Marketing Research*, *37*, 383−395.

Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2001, May). Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research*, *38*, 197−215.

Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2002, July). How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Business Review*, *54*, 198−220.

Vriens, M., Wedel, M., & Wilms, T. (1996, February). Metric conjoint segmentation methods: a Monte Carlo comparison. *Journal of Marketing Research*, *33*, 73−85.

Wittink, D.R., Addona, M.J., Hawkes, W.J., & Porter, J.C. (1988). The estimation, validation, and use of promotional effects based on scanner data. Working paper, Johnson Graduate School of Management, Cornell University.