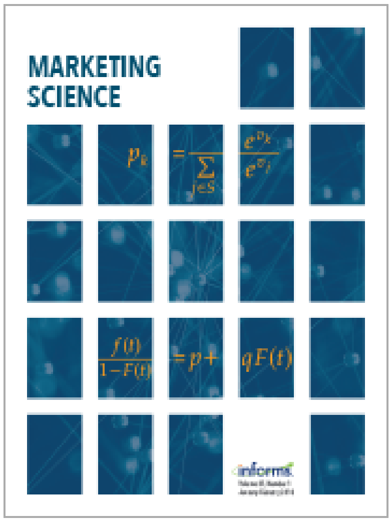


Marketing Science

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

Commentary—Reexamining Bayesian Model-Comparison Evidence of Cross-Brand Pass-Through

Jason A. Duan, Leigh McAlister, Shameek Sinha,

MARKETING SCIENCE
CELEBRATING 30 YEARS
Vol. 30, No. 3, May–June 2011, pp. 550–561
ISSN 0732-2399 | EISSN 1526-548X | 11 | 3003 | 0550

informs®
DOI 10.1287/mksc.1100.0611

Commentary

Reexamining Bayesian Model-Comparison Evidence of Cross-Brand Pass-Through

Jason A. Duan, Leigh McAlister, Shameek Sinha
Department of Marketing, University of Texas at Austin, Austin, Texas 78712
{duanj@mccombs.utexas.edu, leigh.mcalister@mccombs.utexas.edu, shameek.sinha@phd.mccombs.utexas.edu}

Using the Bayes factor estimated by harmonic mean [Newton, M. A., A. E. Raftery. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B.* **56**(1) 3–48] to compare models with and without cross-brand pass-through, Dubé and Gupta [Dubé, J.-P., S. Gupta. 2008. Cross-brand pass-through in supermarket pricing. *Marketing Sci.* **27**(3) 324–333] found that, in the refrigerated orange juice category, a model with cross-brand pass-through was selected 68% of the time. However, Lenk [Lenk, P. J. 2009. Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *J. Comput. Graph. Statist.* **18**(1) 941–960] has demonstrated that the infinite variance harmonic mean estimator often exhibits simulation pseudo-bias in favor of more complex models. We replicate the results of Dubé and Gupta in the refrigerated orange juice category and then show that any of three more stable finite variance estimators select the model with cross-brand pass-through less than 1% of the time. Relaxing the assumption that model errors are distributed normally eliminates all instances in which the cross-brand pass-through model is selected. In 10 additional categories, the harmonic-mean-estimated Bayes factor selects the model with cross-brand pass-through 69% of the time, whereas a finite variance estimator of the Bayes factor selects the model with cross-brand pass-through only 5% of the time. Applying arguments in McAlister [McAlister, L. 2007. Cross-brand pass-through: Fact or artifact? *Marketing Sci.* **26**(6) 876–898], these 5% of cases can be attributed to capitalization on chance. We conclude that Dubé and Gupta should not be interpreted as providing evidence of cross-brand pass-through.

Key words: cross-brand pass-through; promotion; retail and wholesale price; scanner data; Bayes factor; model comparison; Bayesian nonparametric
History: Received: March 23, 2010; accepted: October 14, 2010; Eric Bradlow served as the editor-in-chief and Michel Wedel served as associate editor for this article. Published online in *Articles in Advance* December 30, 2010.

1. Introduction

Each year, packaged goods manufacturers spend more than 60% of their marketing budgets (i.e., more than $75 billion) on trade promotions in an attempt to influence the merchandising support for and the prices of their brands at retail (Cannondale Associates 2001). A good deal of research has gone into the question of whether those trade promotion dollars are well spent. Early studies considered a few trade promotions offered to a retailer over a short period of time and came up with estimates of the extent to which trade promotions were passed through to end consumers (Armstrong 1991, Chevalier and Curhan 1976, Curhan and Kopp 1986, Walters 1989). Because details of trade promotions are not readily available and because researchers wanted to work with longer time windows, more recent promotion pass-through analyses use retail and wholesale price data, reasoning that the coefficient of a wholesale price in the reduced-form model of a retail price gives insight into the extent to which that wholesale price is passed through to that retail price (Besanko et al. 2005, McAlister 2007, Pauwels 2007, Dubé and Gupta 2008, Ailawadi and Harlam 2008, Nijs et al. 2010).

Recently, researchers have considered the possibility that one manufacturer's wholesale support is regularly being passed through to the retail price of competitors' brands, a practice known as cross-brand pass-through. Besanko et al. (2005) were thought to have provided evidence consistent with cross-brand pass-through. However, McAlister (2007) argued that they had inadvertently overcounted the number of independent observations. Adjusting statistical tests to reflect the actual number of independent observations, she showed that there was no evidence of cross-brand pass-through. Dubé and Gupta (2008, DG hereafter) used Bayesian inference to compare a model with cross-brand pass-through (an "unrestricted" model whose predictors include the wholesale prices of all brands in a category) to a model without cross-brand pass-through (a "restricted" model whose only predictor is the target brand's own wholesale price). Focusing on the refrigerated orange juice category, they found that harmonic-mean-estimated Bayes factors favored the unrestricted model 68% of the time. In 10 additional categories, investigated in less detail, their harmonic-mean-estimated Bayes factors favored the unrestricted model 69% of the time. They interpreted these findings as evidence of cross-brand pass-through.

Recent work by Lenk (2009) suggests the need to reexamine DG's findings. Lenk (2009) shows that the infinite variance harmonic-mean-estimated Bayes factor is often simulation pseudo-biased, and this pseudo-bias favors more complex models. To discover whether DG's results are an artifact of this pseudo-bias, we replicate DG's analysis of the refrigerated orange juice category and then reanalyze that category using three finite variance Bayes factor estimators: Gelfand and Dey (1994, GD hereafter), Raftery et al. (2007, RNSK hereafter), and Chib (1995). Following Chib and Jeliazkov (2001), we take agreement among these three differently structured finite variance estimators as evidence that these methods are not greatly influenced by simulation pseudo-bias and are relatively stable. We also reanalyze DG's 10 additional categories, contrasting results implied by the harmonic-mean-estimated Bayes factor to those implied by the Bayes factor estimated by Gelfand and Dey (1994).

In what follows, we first describe our application and the data employed. We then consider model selection implications of different methods for approximating the Bayes factor. We successfully replicate the results of DG in the refrigerated orange juice category, showing that the infinite variance harmonic-mean-estimated Bayes factors select the unrestricted model (with cross-brand pass-through) more frequently than the restricted model (without cross-brand pass-through), as one would expect given Lenk's (2009) demonstration that the harmonic mean estimator is simulation pseudo-biased in favor of more complex models. We also show that the three finite variance estimators (GD, RNSK, and Chib) yield Bayes factors that select the restricted model (without cross-brand pass-through) almost 100% of the time. In two of the 810 cases in our analysis, the Bayes factors estimated using GD, RNSK, and Chib select the unrestricted model (with cross-brand pass-through). Looking more closely at those two cases, we note extreme violations of the assumed normality of model errors. When we model the nonparametric error structure using Dirichlet process priors, we find that the resulting Bayes factors select the restricted model (without cross-brand pass-through). (Although Bayesian nonparametric models with Dirichlet process priors have been applied in the marketing literature—see Wedel and Zhang 2004, Ansari and Mela 2003—we demonstrate that a nonparametric error assumption can reverse managerial implications.) For the additional 10 categories (for which DG's harmonic-mean-estimated Bayes factors selected the unrestricted model 69% of the time), GD-estimated Bayes factors select the unrestricted model only 5% of the time, and cases are argued to be instances of capitalization on chance. In sum, we conclude that DG's evidence is an artifact of the simulation pseudo-bias associated with the harmonic-mean-estimated Bayes factors and, to a lesser extent, to the assumption of normal errors and to capitalization on chance. Using finite variance estimators, we find essentially no evidence consistent with cross-brand pass-through.

In addition, these findings imply that in retail-price model selection, researchers should realize the vulnerability of the harmonic mean estimator, consider the implications of inappropriate model error assumptions, and be vigilant for parameters driven by promotion-driven coincidences in pricing data. We end with a discussion of the implications of this work for researchers studying the determinants of retail grocery prices and for managers trying to understand the effectiveness of their trade promotion spending.

2. Data, Model, and Replication Using Harmonic-Mean-Estimated Bayes Factors

From 1989 to 1994, Chicago Booth and Dominick's Finer Foods (DFF) partnered to conduct store-level research into shelf management and pricing. Randomized experiments were conducted in more than 25 different categories throughout all stores in this 86-store chain. As a by-product of this research cooperation, approximately six years of store-level data on the sales of more than 3,500 UPCs have been made available on the Chicago Booth website.[1] These data are unique for their breadth of coverage and for the inclusion of wholesale prices.

Our data come from this DFF data set. We chose category, time frame, and models consistent with Dubé and Gupta's (2008) analysis of the retail prices of 10 brands of refrigerated orange juice. Dubé and Gupta (2008, p. 325) test "the posterior probability associated with an unrestricted model (i.e., with cross-brand pass-through) versus a restricted model

[1] http://research.chicagobooth.edu/marketing/databases/dominicks/index.aspx.

Table 1 Comparison of Dubé and Gupta's (2008) Model B to Our Replication of Model B

	Their model B	Our replication of their model B
Category	Refrigerated orange juice	
Brands	Top 10 brand sizes as defined by Montgomery (1997)	
Weeks of observation	224 (weeks not specified)	224 (weeks 53–276)
Number of stores	83	81
Average percentage of observations for which a brand is on promotion	37	39
Average value of own-brand coefficient for model A	0.56	0.58
Average value of cross-brand coefficient for model A	0.04	0.04
Percentage of cases in which log marginal likelihood for the model with cross-brand pass-through is higher than that for the model without cross-brand pass-through	68	56

with all the coefficients on substitute products' wholesale prices set to zero (i.e., without cross-brand pass-through)." Thus they cast the question of whether cross-brand pass-through occurs as a model selection problem.

Dubé and Gupta (2008, p. 328) estimate four different model formulations to test whether cross-brand pass-through exists in retail pricing. For simplicity of exposition, we focus on the unrestricted and restricted versions of their "model B," in which the models were estimated separately for each of the 10 brands in each of the 83 DFF stores they studied. The models and notation are as follows:

Unrestricted model (with cross-brand pass-through):

$$\log p_{ist} = \beta_{is}^0 + \beta_{is}^i \log c_{ist} + \sum_{j\neq i} \beta_{is}^j \log c_{jst} + \varepsilon_{ist}, \quad \text{where } \varepsilon_{ist} \sim \text{iid N}(0, \sigma_{is}^2); \quad (1)$$

Restricted model (without cross-brand pass-through):

$$\log p_{ist} = \beta_{is}^0 + \beta_{is}^i \log c_{ist} + \varepsilon_{ist}, \quad \text{where } \varepsilon_{ist} \sim \text{iid N}(0, \sigma_{is}^2); \quad (2)$$

where

$i = 1, 2, \ldots, I$: brand index;
$s = 1, 2, \ldots, S$: store index;
$t = 1, 2, \ldots, n$: time index;
p_{ist}: retail price of brand i in store s at week t;
c_{ist}: wholesale price of brand i in store s at week t;
$\beta_{is}^i =$ own-brand pass-through elasticity;
$\beta_{is}^j =$ cross-brand pass-through elasticity;
$\beta_{is} = \{\beta_{is}^i, \beta_{is}^j\}$.

Following the same priors used by Dubé and Gupta (2008), the priors for the parameters in the unrestricted model (Equation (1)) and restricted model (Equation (2)) are as follows:

$$\beta_{is} \sim \text{N}(\bar{\beta}, \bar{\Sigma}), \quad \text{where } \bar{\Sigma} = \begin{bmatrix} \bar{\sigma}^2 & & \\ & \ddots & \\ & & \bar{\sigma}^2 \end{bmatrix}_{\dim(\bar{\beta})}$$

with ($\bar{\beta} = 0$ and $\bar{\sigma}^2 = 100$), and $\tau_{is} = \sigma_{is}^{-2} \sim \text{Gamma}(a, a \cdot \text{var}(p_{is}))$ with $a = 1.5$, $p_{is} = \{p_{ist}, t = 1, 2, \ldots, n\}$, and $\text{var}(p_{is})$ is the sample variance of p_{is}.

Using the standard Bayesian linear regression formulation, we estimate Equations (1) and (2) for each of the 10 brands (i) in each of the 81 stores (s), using 224 weeks of data (t) for each brand–store combination.

We follow DG by estimating marginal likelihood using the harmonic mean estimator and assuming that model errors are normally distributed. Table 1 compares our replication of DG's model B to the statistics for that model that they report in their paper.

Both the original and the replication are built with 224 weeks of DFF data for the 10 brands of refrigerated orange juice defined by Montgomery (1997). We use weeks 53–276; DG do not report the specific 224 weeks that they use. We selected 81 of DFF's 86 stores because five stores had many missing observations for target brands. (We could not determine which three stores DG decided to drop.) Despite these small differences, sample statistics and parameters estimated are similar. DG's Table 3 reports that, on average, a brand is on promotion for 37% of its observations. In our replicated data, on average, a brand is on promotion for 39% of its observations. More importantly, when we look at the average value of model parameters estimated with the two data sets, we see that the average own-brand pass-through elasticity reported in DG's Table 5a is 0.56, the corresponding average in our replication is 0.58, and a paired-sample t-test shows that the difference is not statistically significant ($n = 10$, $t = 1.25$, $p = 0.24$). DG's Table 5a reports that average cross-brand pass-through elasticity is 0.04, the corresponding average from our replication is 0.04, and a paired-sample t-test shows that the difference is not statistically significant ($n = 90$, $t = 0.33$, $p = 0.74$). Finally, DG's Table 4 reports that (using the harmonic-mean-estimated Bayes factor), in 68% of the cases, the unrestricted version of model B (which included cross-brand pass-through) yielded higher posterior log marginal density than did the restricted version of model B (without cross-brand pass-through). In our replication (using the harmonic-mean-estimated Bayes factor), the version of model B that included cross-brand pass-through yielded higher posterior log marginal density 56% of the time.[2] Overall, our replication matches DG's pretty well.

3. Model Comparison Using Finite Variance Estimators of the Bayes Factor

Marginal likelihoods are the building blocks of the Bayes factor. Newton and Raftery's (1994) harmonic mean estimator for marginal likelihood, used by DG and used in the replication just reported, is the simplest and most commonly applied approach to calculate marginal likelihood. However, the harmonic mean estimator is known to have infinite variance and, hence, to be unstable. In addition, Lenk (2009) shows that this estimator often suffers from substantial simulation pseudo-bias. The pseudo-bias causes overestimation of the marginal likelihood, especially for more complex models. Hence the harmonic mean estimator can be expected to incorrectly select more complex models. Because the model with cross-brand pass-through in Equation (1) is more complex than the model without cross-brand pass-through in Equation (2), the harmonic mean estimator is expected to wrongly favor the cross-brand pass-through model. To assess the extent of bias introduced by the harmonic mean estimator, we estimate Bayes factors with three additional, finite-variance estimators: GD, RNSK, and Chib.[3] In what follows, we give a brief description of the three finite variance methods, whose detailed derivations are given in Chapter 6 of Rossi et al. (2005).

(1) GD. Gelfand and Dey (1994) propose the following estimator for the marginal likelihood:

$$\hat{p}(y \mid M_k) = \left\{ \frac{1}{R} \sum_{r=1}^{R} \frac{q(\theta^r)}{p(y \mid \theta^r, M_k) p(\theta^r \mid M_k)} \right\}^{-1}, \qquad (3)$$

where $r = 1, 2, \ldots, R$ are the number of Monte Carlo draws from the posterior, $p(y \mid \theta^r, M_k)$ is the likelihood function under model M_k, and $p(\theta^r \mid M_k)$ is the prior density. The importance sampling function, $q(\theta^r)$, is the density of an arbitrary distribution which has the same support of $p(\theta^r \mid M_k)$. For (3) to have a finite variance, $q(\theta)$ is selected to have lighter tails than the posterior $p(\theta \mid y, M_k)$. Gelfand and Dey (1994) propose using a multivariate normal density for $q(\theta)$, whose mean and covariance matrix are set to be equal to the posterior mean and covariance estimated from the posterior sample. Lenk (2009) applied this estimator and found it to be quite accurate. The GD method has also been applied in the marketing literature by Albuquerque et al. (2007), Duan and Mela (2009), and Lenk and DeSarbo (2006). Note the harmonic mean estimator is a special case of GD where $q(\theta)$ is selected to equal the prior $p(\theta \mid M_k)$, which causes the infinite variance problem.

(2) RNSK. Raftery et al. (2007) propose an extension of GD by marginalizing a subvector of θ and maintaining closed-form for the likelihood function. They prove that this estimator is more stable than the harmonic mean estimator and that it also allows easier selection of the importance sampling function $q(\theta)$ in the GD estimator. We apply this method in our paper by integrating out σ_{is}^2 in our regression, resulting in a t-distribution likelihood function.

(3) Chib. As the harmonic mean, GD and RNSK are all based on importance sampling; Chib (1995), however, approaches the problem in a completely different way by proposing to evaluate the marginal likelihood based on a simple observation that

$$p(y \mid M_k) = \frac{p(y \mid \theta^*, M_k) p(\theta^* | M_k)}{p(\theta^* \mid y, M_k)} \qquad (4)$$

holds for any fixed choice of θ^*. The posterior density, $p(\theta^* \mid y, M_k)$, is often computed from repeated posterior samples holding sequentially various subvectors of θ constant and equal to θ^*. Recent applications of Chib's method in marketing include Aribarg et al. (2010), Choi et al. (2010), van der Lans et al. (2009), and Teixeira et al. (2010).

It is useful to note that the linear, nonhierarchical structure of the models in Equations (1) and (2) facilitates implementation of the three finite variance methods. For RNSK, implementation is further eased by the fact that errors are assumed to be normal, facilitating the integration needed to reduce parameters and thereby ensure "heavy tails" for the importance sampling density. For Chib (1995), implementation is further eased by the assumed conjugate structure.

The results from our harmonic-mean-based replication of DG (which selected the model with cross-brand pass-through 56% of the time) are compared to the results of model selection based on GD's estimator with multivariate normal density distribution for $q(\theta)$, RNSK's method, and Chib's method. Given the distinctive strategies that underlie the three finite variance estimators, we will take agreement among

[2] The discrepancy between DG's and our estimates in the percentages of the cases for which the unrestricted model has higher log marginal likelihood is likely due to the differences in the data and the instability of the harmonic mean estimator.

[3] We only apply various estimators of the Bayes factor because our models are estimated with Bayesian methods. Hence the results can be directly compared with DG. Alternative model selection criteria are surveyed and compared in Rust et al. (1995).

Table 2 Log Marginal Likelihoods for the Unrestricted Model (with Cross-Brand Pass-Through) and Restricted Model (Without Cross-Brand Pass-Through) Using Different Estimators (Estimated with Data from Store 41 and Assuming Model Errors Are Normal)

	Log marginal likelihood estimator	Unrestricted model	Restricted model
Tropicana Premium 64 oz	HM	**95.6**	94.5
	GD	54.0	**84.2**
	RNSK	54.2	**84.3**
	Chib	54.0	**84.2**
Tropicana Premium 96 oz	HM	**268.5**	268.3
	GD	218.9	**257.0**
	RNSK	219.0	**257.1**
	Chib	218.9	**257.0**
Florida Natural 64 oz	HM	**154.0**	152.1
	GD	110.5	**143.6**
	RNSK	110.6	**143.5**
	Chib	110.5	**143.6**
Tropicana 64 oz	HM	74.8	**83.2**
	GD	38.0	**73.2**
	RNSK	38.0	**73.2**
	Chib	38.0	**73.2**
Minute Maid 64 oz	HM	81.0	**82.1**
	GD	39.4	**71.4**
	RNSK	39.6	**71.5**
	Chib	39.4	**71.4**
Minute Maid 96 oz	HM	**240.5**	237.6
	GD	195.6	**225.8**
	RNSK	195.6	**225.9**
	Chib	195.6	**225.8**
Tree Fresh 64 oz	HM	34.1	**35.6**
	GD	−3.9	**24.7**
	RNSK	−3.7	**24.7**
	Chib	−3.9	**24.7**
Dominick's 64 oz	HM	**42.6**	40.0
	GD	3.8	**31.0**
	RNSK	4.0	**31.0**
	Chib	3.8	**31.0**
Dominick's 128 oz	HM	**156.1**	133.6
	GD	113.1	**122.6**
	RNSK	113.1	**122.7**
	Chib	113.1	**122.6**
Tropicana Premium 32 oz	HM	**438.2**	427.7
	GD	380.4	**415.7**
	RNSK	380.4	**415.7**
	Chib	380.4	**415.7**

Note. HM, harmonic mean.

these methods' estimates as evidence of the estimates' accuracy (as was done in Chib and Jeliazkov 2001).

For illustration purposes, Table 2 reports log marginal likelihood values for all 10 brands, for both unrestricted (including cross-brand pass-through) and restricted (no cross-brand pass-through) models in Store 41, selected at random. We see that the harmonic mean estimator selects the unrestricted model for 70% of the cases (i.e., for all brands except Tropicana 64 oz, Minute Maid 64 oz, and Tree Fresh 64 oz), consistent with DG's (2008, Table 4) finding that the unrestricted model (with cross-brand pass-through) was selected in 68% of the cases. Note, however, that the three finite variance methods (GD, RNSK, and Chib) imply a different pattern of model selection. For all three alternative methods and for all 10 brands, the restricted model (without cross-brand pass-through) is chosen. Moreover, these three methods yield nearly identical log marginal likelihood estimates, making it reasonable to assume that these methods' estimates are accurate (Chib and Jeliazkov 2001). Consequently, as predicted by Lenk (2009), we see that the harmonic mean estimator produces upwardly biased log marginal likelihood estimates for the simple, restricted model, and it produces even more upwardly biased log marginal likelihood estimates for the more complex, unrestricted model. Consistent with Lenk, this shows that the more complex model with cross-brand pass-through is incorrectly favored by the harmonic mean estimator.

Looking across all 81 DFF stores we have studied, Table 3 presents a similar pattern of results to that presented in detail for Store 41 in Table 2.[4] Each column in Table 3 corresponds to a log marginal likelihood estimation method; each row corresponds to a brand. The table entries report the number of stores (out of 81) for which the log marginal likelihoods favored the unrestricted, cross-brand pass-through model. The "Harmonic mean" column shows that the unrestricted model has a higher log marginal likelihood in 56% (450 out of 810) of the cases, analogous to DG's findings using the harmonic mean estimator. For all three finite variance methods and for 9 out of the 10 brands, the unrestricted model *never* has a higher log marginal likelihood than the restricted model. For one brand (Dominick's 128 oz refrigerated orange juice), in two stores (Stores 23 and 79), all four estimation methods produce a higher log marginal likelihood for the unrestricted model. In the next section, we look at the nature of models' errors to consider the question of whether one should interpret these two (out of 810) observations as evidence of cross-brand pass-through.

4. Model Comparison Assuming Nonparametric Errors

In this section we consider the implication of assuming models' errors to be distributed normally when, in

[4] Note that Table 2's pattern, for each brand, of essentially equal log marginal likelihood estimates across the three finite variance methods holds up for all brands in all stores. We suppress that detail in Table 3, but the data are available from the authors upon request.

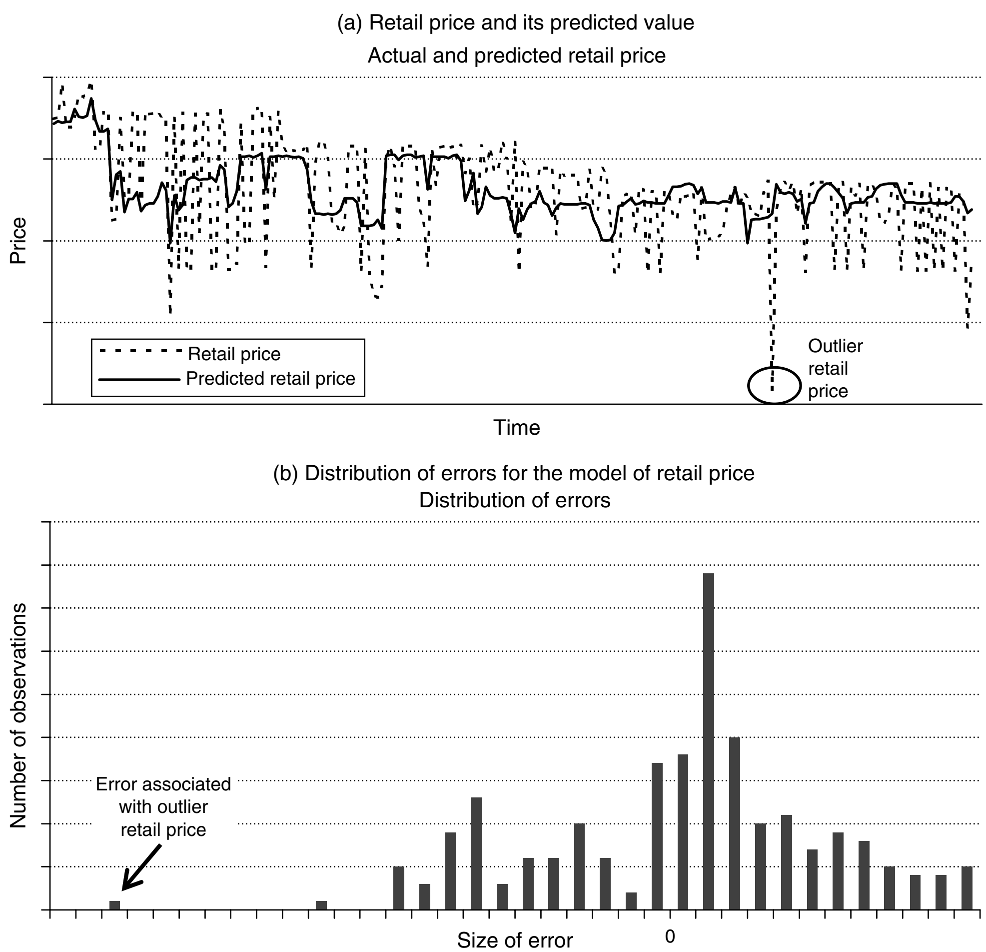
Table 3 **Fraction of Cases in Which the Unrestricted Model (with Cross-Brand Pass-Through, Equation (1)) Has Higher Log Marginal Likelihood Than the Restricted Model (Without Cross-Brand Pass-Through, Equation (2)), Given Different Log Marginal Likelihood Estimation Methods and Normally Distributed Model Errors**

Refrigerated orange juice brands	Fraction of cases in which $\log(\hat{p}(y \mid M_i)_{\text{Unrestricted}}) > \log(\hat{p}(y \mid M_i)_{\text{Restricted}})$			
	Harmonic mean	GD	RNSK	Chib
Tropicana Premium 64 oz	20/81	0	0	0
Tropicana Premium 96 oz	37/81	0	0	0
Florida Natural 64 oz	52/81	0	0	0
Tropicana 64 oz	18/81	0	0	0
Minute Maid 64 oz	5/81	0	0	0
Minute Maid 96 oz	80/81	0	0	0
Tree Fresh 64 oz	39/81	0	0	0
Dominick's 64 oz	52/81	0	0	0
Dominick's 128 oz	81/81	2/81	2/81	2/81
Tropicana Premium 32 oz	66/81	0	0	0
Total	450/810	2/810	2/810	2/810

fact, those errors are not normal. We begin by explaining why one might expect models of retail price to have nonnormal errors. We then show that it is the nonnormality of the model errors that leads to higher log marginal likelihood for the unrestricted model for Dominick's 128 oz refrigerated orange juice in Stores 23 and 79.

Note that a typical grocery brand will have a "regular" unpromoted retail price and will also have noticeable short-term price reductions, probably

Figure 1 **Implication of Retail Price Outliers for the Model Error Distribution**



associated with the brand's retail promotions. The thin line in Figure 1(a) represents the retail price pattern for a grocery brand for 224 weeks. Although the brand's regular retail price goes down over the 224 weeks, there is an episodically stable regular price, punctuated by frequent short-term price reductions. Figure 1(a) presents actual and predicted retail prices, and Figure 1(b) presents the error distribution associated with those actual and predicted prices. Given the outliers, multiple modes, and heavy tails, the errors may not have a normal distribution. Indeed, normality tests (the Anderson–Darling test, the Cramér–von-Mises criterion, the Lilliefors test, and the Shapiro–Wilk test) reject the hypothesis that this error distribution is normal. If, in analyzing these data, one assumes that model error has a normal distribution (as do most Bayesian models of retail price), then marginal likelihood values may be distorted. As an alternative, we suggest a flexible, nonparametric distribution of errors for the underlying models and show that when we use a flexible nonparametric error structure rather than imposing a normal error structure, Bayes factor-based model selections change.

We examine the error structure of the model with cross-brand pass-through in Equation (1) and the model without cross-brand pass-though in Equation (2)

Figure 2 **Normal Q–Q Plots for the Residuals of the Unrestricted and Restricted Models for Dominick's 128 oz Refrigerated Orange Juice in Stores 23 and 79**

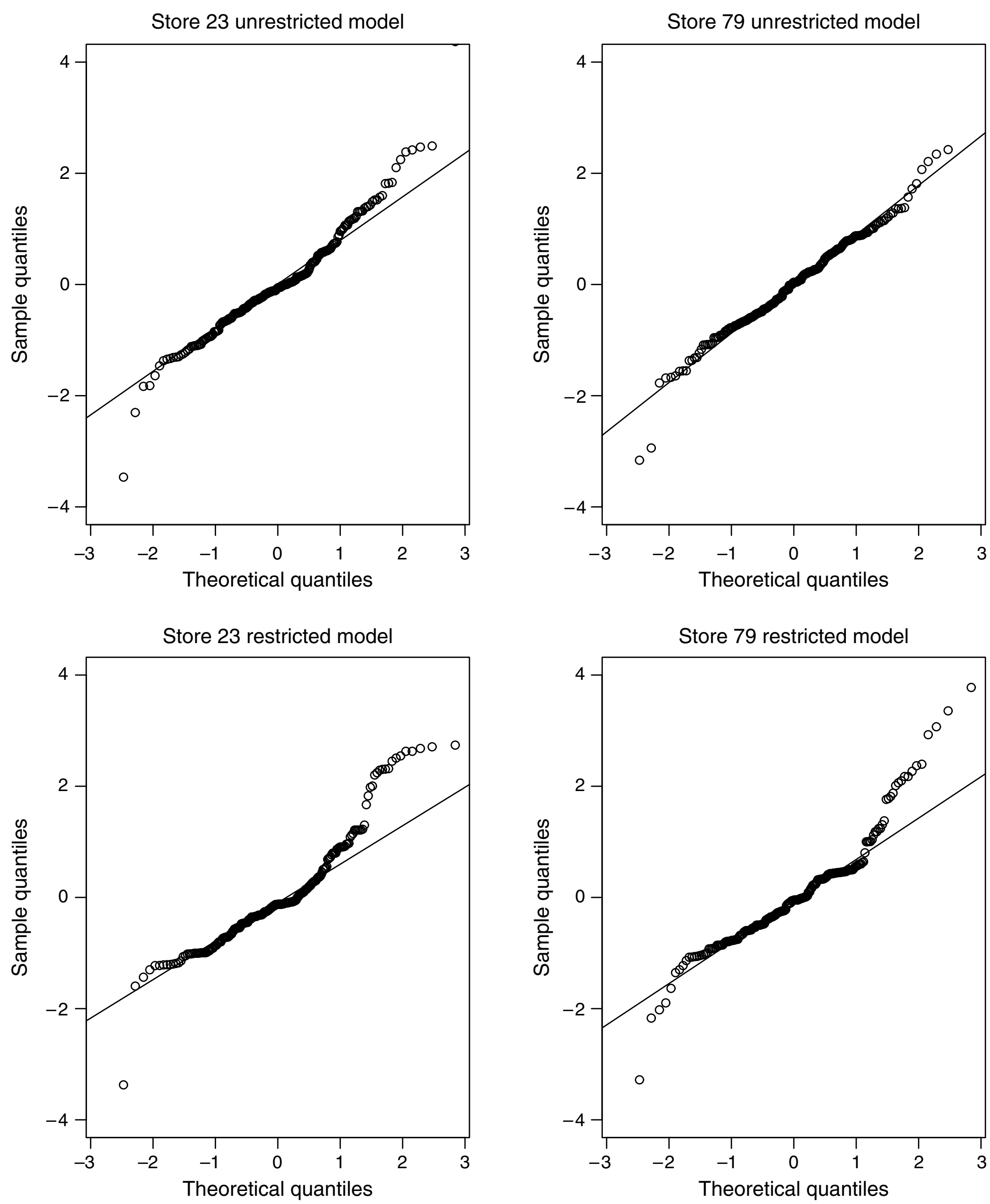
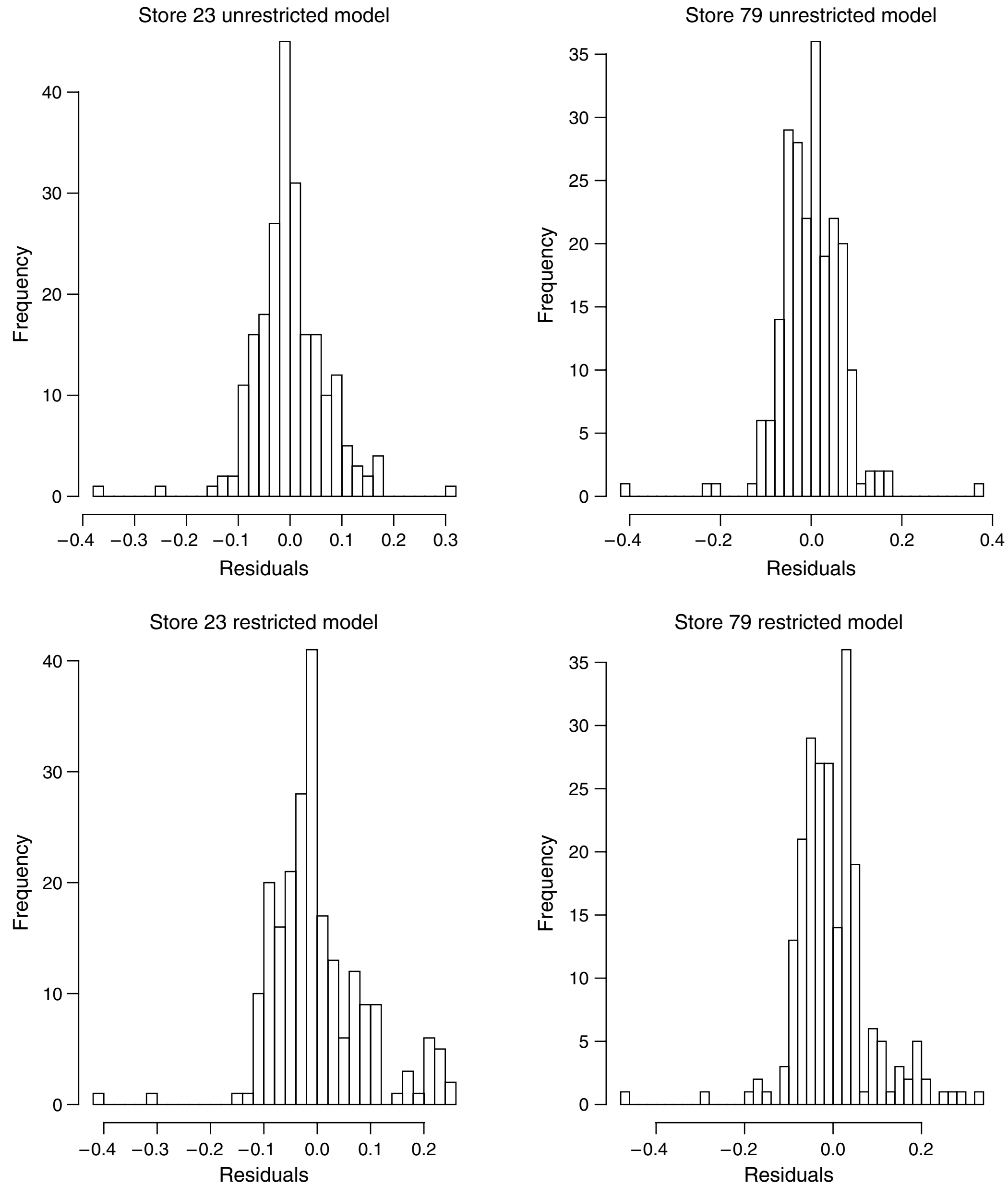


Figure 3 Residual Histograms for the Unrestricted Model (Including Cross-Brand Pass-Through) and Restricted Model (Not Including Cross-Brand Pass-Through) for Dominick's 128 oz Refrigerated Orange Juice in Stores 23 and 79



by estimating linear regression models and recording the resulting residuals. Figure 2 illustrates the normal Q–Q plots of the residuals from the corresponding unrestricted and restricted regression models for the retail prices of Dominick's 128 oz orange juice in Stores 23 and 79. The nonnormality of the errors for both models in both stores, easily seen in Figure 2's Q–Q plots, is also confirmed with various statistical tests. (The Anderson–Darling test, the Cramér–von-Mises criterion, the Lilliefors test, and the Shapiro–Wilk test all reject normality for the model errors of these two stores.)

To see the implication of nonnormality, consider Figure 3, with residual histograms for Dominick's 128 oz refrigerated orange juice in stores 23 and 79. The residuals from the restricted models have heavier tails and may even have multiple modes, increasing error variance. Such an increase in error variance will cause the normal distribution to assign lower density values to the majority of the observations clustering near the mode of the distribution. Therefore, the restricted model is penalized by having lower likelihood (density) for most of the residuals near the mode, which leads to a lower marginal likelihood. The unrestricted model (with nine more parameters than the restricted model) does not have as many outliers, and its outliers are less extreme, thereby reducing the outlier penalty.

Table 4 Implications of Normal and Nonparametric Error Assumptions: Log Marginal Likelihoods for Unrestricted Model (with Cross-Brand Pass-Through, Equation (1)) and Restricted Model (Without Cross-Brand Pass-Through, Equation (2))

	Normal error distribution (log marginal likelihood estimated with Chib 1995)		Nonparametric error distribution (log marginal likelihood estimated with Basu and Chib 2003)	
	Log marginal likelihood for unrestricted model (with cross-brand pass-through)	Log marginal likelihood for restricted model (without cross-brand pass-through)	Log marginal likelihood for unrestricted model (with cross-brand pass-through)	Log marginal likelihood for restricted model (without cross-brand pass-through)
Store 23	217.2	208.8	221.4	**254.9**
Store 79	215.5	213.2	229.5	**232.6**

Note. Estimated with data for Dominick's 128 oz refrigerated orange juice in Stores 23 and 79.

Theoretically, based on the information inequality (or Kullback-Leibler divergence),

$$D_{\mathrm{KL}}(p(x), q(x)) = \int p(x) \log \frac{p(x)}{q(x)}\, dx \geq 0,$$

where $p(x)$ and $q(x)$ are densities; the greater the likelihood function deviates from the real distribution from which the data are drawn, the lower the log likelihood. If the restricted model's errors are "more" nonnormal than the unrestricted model's errors in these two cases (Dominick's 128 oz refrigerated orange juice in stores 23 and 79), then the restricted model's marginal likelihood may be being penalized more than the unrestricted model's marginal likelihood by the normality assumption. If this is the case, employing a more appropriate error structure will eliminate likelihood distortion.

To achieve that goal, we propose to use the Dirichlet process mixture (DPM) model from Bayesian nonparametrics for the error distribution, which is specified as follows:

$$\begin{aligned}
&\log p_{ist} = \beta_{is}^0 + \beta_{is}^i \log c_{ist} + \sum_{j\neq i} \beta_{is}^j \log c_{jst} + \varepsilon_{ist},\\
&\varepsilon_{ist} \sim \mathrm{N}(\mu_{ist}, \tau_{ist}^{-1}),\\
&\qquad \text{where } \tau_{ist} = \sigma_{ist}^{-2} \text{ is the precision parameter},\\
&\mu_{ist}, \tau_{ist} \sim G(\mu, \tau) \quad \text{and} \quad G \sim DP(\alpha G_0),\\
&G_0 = \mathrm{N}(\mu_{ist} \mid 0, (g\tau_{ist})^{-1}) \mathrm{Gamma}(\tau_{ist} \mid a_\tau, a_\tau b_\tau),\\
&\alpha \mid a_\alpha, b_\alpha \sim \mathrm{Gamma}(a_\alpha, b_\alpha) \quad \text{and} \quad b_\tau \sim \mathrm{Gamma}(a_b, b_b),\\
&\beta_{is} = (\beta_{is}^j, j = 0, 1, \ldots, S) \sim \mathrm{N}(\bar{\beta}, \bar{\Sigma}). \qquad (5)
\end{aligned}$$

The values $(g, a_\tau, a_b, b_b, a_\alpha, b_\alpha)$ are the hyperparameters with given values. The restricted version of model (8) has $\beta_{is}^j = 0$ for all $j \neq i$. This is a DPM model for the error ε_{ist} that allows us to approximate any unknown distribution.

Following Basu and Chib (2003), we calculate marginal likelihood for DPM models where the likelihood is evaluated by a sequential Monte Carlo method. The marginal likelihood for the unrestricted and restricted versions of model (5) can be directly compared with the marginal likelihood for the unrestricted and restricted models with normal error distribution in §3. The details of Bayesian inference and marginal likelihood estimation for model (5) are available in the electronic companion, available as part of the online version that can be found at http://mktsci.pubs.informs.org.

The results in Table 4 focus on Dominick's 128 oz refrigerated orange juice in stores 23 and 79, the two cases out of 810 for which all estimation methods yielded higher log marginal likelihood for the unrestricted model than for the restricted model, given that errors were assumed to be normal (see Table 3). Assuming that normal error structure, columns 2 and 3 in Table 4 represent the log marginal likelihood values for unrestricted and restricted models estimated using Chib (1995). Assuming the flexible nonparametric error distribution described above, columns 4 and 5 in Table 4 represent the log marginal likelihood values for unrestricted and restricted models estimated using Basu and Chib (2003). As we saw earlier, assuming normal errors results in the unrestricted model having a higher log marginal likelihood than the restricted model. However, after relaxing the assumption of normal errors, it is the restricted model that has a higher log marginal density for Dominick's 128 oz refrigerated orange juice in stores 23 and 79 (shown in bold). Given that we earlier showed that model errors in these two cases are highly nonnormal, and given the fact that the log marginal likelihoods are higher for nonparametric models than for the models assuming normal errors, we conclude that this evidence should be interpreted as being consistent with the restricted model (which does not include cross-brand pass-through) being superior to the unrestricted model (which does include cross-brand pass-through), even for Dominick's 128 oz refrigerated orange juice in stores 23 and 79.

Table 5 Ten Additional Categories: Fraction of Cases in Which the Unrestricted Model (with Cross-Brand Pass-Through, Equation (1)) Has Higher Log Marginal Likelihood Than the Restricted Model (Without Cross-Brand Pass-Through, Equation (2)) for Harmonic-Mean-Estimated and Gelfand and Dey Estimated Bayes Factors

	Fraction of cases in which $\log(\hat{p}(y \mid M_i)_{\text{Unrestricted}}) > \log(\hat{p}(y \mid M_i)_{\text{Restricted}})$	
Category	Harmonic-mean-estimated Bayes factors (as reported in Dubé and Gupta 2008)	Bayes factors (estimated by Gelfand and Dey 1994)
Bath tissue	50/90	0/90
Beer	53/105	20/90[a]
Crackers	19/105	0/105
Dish detergent	139/165	2/165
Frozen orange juice	62/75	0/75
Laundry detergent	128/180	0/180
Oat cereal	29/45	6/45
Paper towels	77/90	11/75[a]
Refrigerated orange juice	83/105	0/105
Toothpaste	146/150	0/150
Tuna (canned)	25/60	2/60
Average (%)	69	5

[a]McAlister (2007) was only able to identify six of the Besanko et al. (2005) seven beer brands and five of their six paper towel brands.

5. Analysis of 10 Additional Categories

To analyze the refrigerated orange juice category, DG expanded the window of observation to 224 weeks from the 52 weeks of data that were considered in Besanko et al. (2005). In the remaining 10 categories, DG analyzed the original 52 weeks of data. For our reanalysis of those 10 categories, we use the corresponding 52 weeks of data constructed and analyzed by McAlister (2007). Within a category, restricted and unrestricted models were estimated for each brand in each price zone yielding ({number of brands in the category} × 15 zones) model comparisons for each category. The first column of Table 5 reports the category. The second column reports, for the harmonic-mean-estimated Bayes factors, the fraction of model comparisons for which the unrestricted model was chosen over the restricted model. The last column reports, for the GD-estimated Bayes factors, the fraction of model comparisons for which the unrestricted model was chosen over the restricted model. On average, the harmonic-mean-estimated Bayes factors selected the unrestricted model 69% of the time. Consistent with the harmonic-mean-estimated Bayes factor exhibiting simulation pseudo-bias in favor of more complex models, on average, the GD-estimated Bayes factors selected the unrestricted model only 5% of the time.

To decide whether the 5% of cases in which the GD-estimated Bayes factors selected the unrestricted model should be taken as evidence consistent with cross-brand pass-through, we turn to evidence provided in McAlister (2007, Table F1). That analysis showed that, on average for these categories, approximately 17% of the estimated cross-brand pass-through coefficients were statistically significant when the unrestricted model was estimated with all 52 weeks of data but that only 1% of the estimated cross-brand pass-through coefficients were stable (i.e., only 1% of estimated coefficients that were significant when the model was estimated with the first half of the data remained significant with the same sign when the model was estimated with the second half of the data). McAlister (2007, Figure 6) explained these unstable coefficients as instances in which the unrestricted model exploits promotion-driven coincidences in the price data. Phenomena documented in McAlister (2007) can be taken as an explanation for the 5% of cases, on average, in which the GD-estimated Bayes factors selected the unrestricted model. Plots of price data[5] confirm this interpretation. Consequently, we attribute these instances to capitalization on chance.

6. Discussion and Conclusion

The question of whether evidence that cross-brand pass-through has been practiced has been cast as a model selection problem: Does a model that includes cross-brand pass-through—an unrestricted model—have a higher marginal likelihood than a model that does not include cross-brand pass-through—a restricted model? To address this question, DG used the infinite variance harmonic mean estimator of the Bayes factor to choose between unrestricted and restricted Bayesian linear regression models. We have shown that the harmonic mean estimator artificially inflated the marginal likelihood of the more complex, cross-brand pass-through model as pointed out by Lenk (2009). Finite variance estimation methods for the marginal likelihood (Gelfand and Dey 1994, Raftery et al. 2007, Chib 1995) are in agreement, and those methods' estimates differ dramatically from harmonic mean estimates. Following Chib and Jeliazkov (2001), we conclude that the evidence provided by the three finite variance methods is more convincing. In fact, for 808 out of 810 cases, the restricted model has a higher marginal likelihood than the unrestricted model.

For two of the 810 cases, all of the four methods yielded higher marginal likelihood estimates for the unrestricted model than for the restricted model. To determine whether these two cases might be evidence

[5] Price plots, suppressed here in the interest of space, are available from the authors upon request.

consistent with cross-brand pass-through, we looked at them more closely. In both cases, for both the unrestricted model and the restricted model, the hypothesis that model errors were distributed normally was rejected. Consistent with that, when we reestimated models using a flexible nonparametric error structure, we found that marginal likelihoods improved for both models but that the improvement was greater for the restricted model. In a demonstration that a nonparametric error structure can reverse managerial implications, the restricted model's marginal likelihood was shown to be greater than the unrestricted model's marginal likelihood.

Finally, we compared restricted and unrestricted models for the 10 additional categories analyzed by DG. Whereas the harmonic-mean-estimated Bayes factors suggested that, on average, 69% of the model comparisons favored the unrestricted model, the GD-estimated Bayes factors suggested that, on average, only 5% of the model comparisons favored the unrestricted model. Based on evidence in McAlister (2007), we attribute the 5% of instances favoring the unrestricted model to capitalization on chance.

In summary, the evidence from Dubé and Gupta (2008), which was thought to have been consistent with cross-brand pass-through, has been shown in this paper to be an artifact of a simulation pseudo-biased estimator and an inappropriate error structure. This does not prove that cross-brand pass-through was not practiced. To support such a claim, one would need to consider a large number of linear and nonlinear model structures involving the wholesale prices of every possible combination of brands within a category and across all other categories in the store. Ailawadi and Harlam (2008) confirm that one manufacturer's promotion money is sometimes spent on other manufacturers' brands. They describe a situation in which a retailer, each week, pools promotion money across all brands in the store and then allocates promotional support from that pool of funds to those brands that the retailer has selected for exceptional support that week. What we do not have is managerial or statistical support for an interpretation of cross-brand pass-through as the transfer of one brand's wholesale promotional support to a specific competitive brand's retail price every time the first brand offers wholesale promotional support. Given the lack of evidence for such within-category, systematic use of one manufacturer's promotion dollars to support a specific competitive brand's retail price change, a manufacturer should probably focus its attention on understanding the extent to which the wholesale promotional support that it provides to a retailer is passed through to its own brands at retail.

For researchers, this work suggests that one should carefully consider properties of the estimation method when using the Bayes factor to select models. The simulation pseudo-bias inherent in the harmonic mean estimator, which was proven theoretically by Lenk (2009), has been shown by the example of cross-brand pass-through in this paper to give misleading answers. Researchers selecting Bayesian models using the harmonic mean estimator should confirm their selection results using at least one of the more reliable estimators (GD, RNSK, or Chib). Because the pseudo-bias tends to underpenalize more complex models, the harmonic mean estimator ought not to be used to compare nested models. In addition, researchers should carefully consider error structure because nonnormality of errors has been shown to inappropriately influence model choice when errors are assumed to be normal. When error structures exhibit unusual patterns (as in our data set, where episodically stable regular prices are punctuated by frequent short-term price reductions), more flexible nonparametric methods need to be employed.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mktsci.pubs.informs.org/.

Acknowledgments

The authors thank Ty Henderson, Garrett Sonnier, Greg Allenby, Peter Fader, Carl Mela, the University of Houston Doctoral Consortium 2009, and anonymous reviewers for helpful comments.

References

Ailawadi, K. L., B. A. Harlam. 2008. Retailer promotion pass-through: A measure, its magnitude, and its determinants. *Marketing Sci.* **28**(4) 782–791.

Albuquerque, P., B. J. Bronnenberg, C. J. Corbett. 2007. A spatiotemporal analysis of the global diffusion of ISO 9000 and ISO 14000 certification. *Management Sci.* **53**(3) 451–468.

Ansari, A., C. F. Mela. 2003. E-customization. *J. Marketing Res.* **40**(2) 131–145.

Aribarg, A., R. Pieters, M. Wedel. 2010. Raising the BAR: Bias adjustment of recognition tests in advertising. *J. Marketing Res.* **47**(3) 387–400.

Armstrong, M. K. 1991. Retail response to trade promotion: An incremental analysis of forward buying and retail promotion. Unpublished doctoral dissertation, School of Management, University of Texas at Dallas, Dallas.

Basu, S., S. Chib. 2003. Marginal likelihood and Bayes factors for Dirichlet process mixture models. *J. Amer. Statist. Assoc.* **98**(461) 224–235.

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

Cannondale Associates. 2001. Trade promotion spending and merchandising study. Report, Cannondale Associates, Wilton, CT.

Chevalier, M., R. C. Curhan. 1976. Retail promotions as a function of trade promotions: A descriptive analysis. *Sloan Management Rev.* **18**(3) 19–32.

Chib, S. 1995. Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90**(432) 1313–1321.

Chib, S., I. Jeliazkov. 2001. Marginal likelihood from the Metropolis-Hastings output. *J. Amer. Statist. Assoc.* **96**(453) 270–281.

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

Curhan, R. C., R. J. Kopp. 1986. Factors influencing grocery retailers' support of trade promotions. Report 86-104, Marketing Science Institute, Cambridge, MA.

Duan, J. A., C. F. Mela. 2009. The role of spatial demand on outlet location and pricing. *J. Marketing Res.* **46**(2) 260–278.

Dubé, J.-P., S. Gupta. 2008. Cross-brand pass-through in supermarket pricing. *Marketing Sci.* **27**(3) 324–333.

Gelfand, A. E., D. K. Dey. 1994. Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56**(3) 501–514.

Lenk, P. J. 2009. Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *J. Comput. Graph. Statist.* **18**(4) 941–960.

Lenk, P. J., W. J. DeSarbo. 2006. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* **65**(1) 93–119.

McAlister, L. 2007. Cross-brand pass-through: Fact or artifact? *Marketing Sci.* **26**(6) 876–898.

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

Newton, M. A., A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* **56**(1) 3–48.

Nijs, V., K. Misra, E. T. Anderson, K. Hansen, L. Krishnamurthi. 2010. Channel pass-through of trade promotions. *Marketing Sci.* **29**(2) 250–267.

Pauwels, K. 2007. How retailer and competitor decisions drive the long-term effectiveness of manufacturer promotions for fast moving consumer goods. *J. Retailing* **83**(3) 297–308.

Raftery, A. E., M. A. Newton, J. M. Satagopan, P. N. Krivitsky. 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West, eds. *Bayesian Statistics*, Vol. 8. Oxford University Press, New York, 1–45.

Rossi, P. E., G. M. Allenby, R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons, Chichester, UK.

Rust, R. T., D. Simester, R. J. Brodie, V. Nilikant. 1995. Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Management Sci.* **41**(2) 322–333.

Teixeira, T. S., M. Wedel, R. Pieters. 2010. Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Sci.* **29**(5) 783–804.

van der Lans, R., J. A. Cote, C. A. Cole, S. M. Leong, A. Smidts, P. W. Henderson, C. Bluemelhuber et al. 2009. Cross-national logo evaluation analysis: An individual-level analysis. *Marketing Sci.* **28**(5) 968–985.

Walters, R. G. 1989. An empirical investigation into retailer response to manufacturer trade promotions. *J. Retailing* **65**(2) 253–272.

Wedel, M., J. Zhang. 2004. Analyzing brand competition across subcategories. *J. Marketing Res.* **41**(4) 448–456.