Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables

Bart J. Bronnenberg, Vijay Mahajan,

Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables

Bart J. Bronnenberg • Vijay Mahajan

John E. Anderson Graduate School of Management, University of California, Los Angeles, 110 Westwood Plaza,
Suite B 4.18, Los Angeles, California 90095
Marketing Department, CBA 7.202, McCombs School of Business, The University of Texas at Austin,
Austin, Texas 78712-1176
bbronnen@anderson.ucla.edu • Vmahajan@mail.utexas.edu

Abstract

Marketing scholars and practitioners frequently infer market responses from cross-sectional or pooled cross-section by time data. Such cases occur especially when historical data are either absent or are not representative of the current market situation. We argue that inferring market responses using cross-sections of multimarket data may in some cases be misleading because these data also reflect unobserved actions by retailers. For example, because the (opportunity) costs of doing so do not outweigh the gains, retailers are predisposed against promoting small share brands. As a consequence, local prices and promotion variables depend on local market shares—the higher the local share, the higher the local observed promotion intensity. We refer to this reverse causation as an endogeneity. Ignoring it will inflate response estimates, because both the promotion effects on share as well as the reverse effects are in the same direction.

In this paper, we propose a solution to this inference problem using the fact that retailers have trade territories consisting of multiple contiguous markets. This implies that the unobserved actions of retailers cause a measurable spatial dependence among the marketing variables. The intuition behind our approach is that by accounting for this spatial dependence, we account for the effects of the retailer's behavior. In this context, our study hopes to make the following contributions at the core of which lies the above intuition.

First, we separate the market response effect from the reverse retailer effect by computing responses to price and promotion net of any spatial—and therefore retailer—influence.

Second, underlying this approach is a new variance-decomposition model for data with a panel structure. This model allows to test for endogeneity of prices and promo-
tion variables in the cross-sectional dimension of the data. This test aims to complement the one developed by Villas-Boas and Winer (1999), who test for endogeneity along the temporal dimension.

Third, to illustrate the approach, we use Information Resources Inc. (IRI) market share data for brands in two mature and relatively undifferentiated product categories across 64 IRI markets. Whereas we only use data with very short time horizons to estimate price and promotion responses with the spatial model, we do have data over long time windows. We use the latter to validate the approach. Specifically, within-market estimates of price and promotion response are not subject to the same endogeneity because we hold the set of retailers constant. Therefore, comparing within- and across-market estimates of price and promotion responses is a natural way to validate the approach. Consistent with our argument, ignoring the reverse causation in the cross-sectional data leads to inferences of price and promotion elasticities that are farther away from zero than the elasticities obtained from within-market analysis. In contrast, cross-sectional spatial estimates and time-series estimates show convergent validity.

From a practical point of view, this means it is possible to obtain reasonable within-market estimates of price and promotion elasticities from (predominantly) cross-sectional data. This may benefit marketing managers. The manager who would act on the inflated elasticities will over-allocate marketing resources to promotions because she ignores retailers' censorship of promotions on the basis of already existing high share. We explore other approaches to correct for the inference bias, and discuss further managerial issues and future research.
(*Spatial Analysis; Promotional Price Response; Promotion Strategy; Endogeneity Biases; Variance-Decomposition*)

1. Introduction

Consider a marketing manager of a nationally distributed brand who wishes to infer price and promotion response for a new product (line), possibly introduced at different times across local markets. Alternatively, consider a marketing manager who wants to evaluate a recently adopted new marketing or promotion campaign for an existing product. Finally, consider a marketing manager who has recently been faced with new competitors so that historical data are of limited use to her. In all these cases, the data that are representative of the new situation are more prominently sampled across markets than across time.

In light of these examples, it is not surprising that the use of cross-sectional data—either in isolated form or as a dimension of panel data with a short time horizon—is widespread among both marketing practitioners and academicians. Practitioners, for example, routinely infer the relation between sales and prices from a *purely* cross-sectional procedure called *sales velocity analysis* (Bucklin and Gupta 1999).

A key issue with the use of cross-sectional data in a multimarket context is that it is not clear whether marketing variables cause observed measures of demand such as market shares or vice versa. For instance, all else equal, retailers in the United States do not like to promote brands that are in low demand because (1) they cannot recoup the operational costs of running the promotion, and (2) their objective is volume generation given that the manufacturer usually bears most of the price reduction. This suggests that observed (promoted) prices and observed market shares in multimarket data are endogenous.

In this context, the goals of this paper are (1) to investigate the extent of this type of endogeneity and (2) to provide a way to correct for its consequences in the measurement of responses to promoted prices and other promotion variables. We accomplish these two goals by making both the ''demand'' side, e.g., market shares, as well as marketing variables, such as price and promotion, a function of an unobserved variable that will have

the interpretation of unobserved retailer influence. To isolate this influence from the data, we make use of the property that retailer territories consist of a set of contiguous markets. Retailer promotion behavior, therefore, manifests itself as spatial dependence in the marketing variables and the market shares. Our contribution to the marketing literature is intended to be threefold.

First, from a substantive viewpoint, we find strong evidence of the dependence of prices and promotion levels on a common unobserved spatial demand state. Specifically, prices covary negatively and promotion levels covary positively with unobserved components in demand. In line with our argument, we find that when we ignore this dependence, elasticities of prices and promotions are farther away from zero than elasticities from within-market analysis. However, when we account for the spatial nature of the data, the cross-market and within-market analyses show convergent validity.

Second, from an empirical standpoint, we find that market shares in our data are very different across regions and that those differences have a spatial structure. For instance, the spatial component of the market-by-time data accounts for 66% to 95% of the variance in brand-shares.

Finally, from a methodological angle, we propose a model for cross-section-by-time processes. We incorporate the spatial structure of data into econometric models using simple techniques. Indeed, whereas the temporal behavior of markets has been studied in depth by marketing academics and practitioners, the spatial aspect of multimarket data has been heretofore largely ignored.

The remainder of this study is organized as follows. Section 2 discusses the substantive background. Section 3 concentrates on a specific formalization of the spatial dimension. The model development occurs in §4, where we incorporate spatiotemporal dependence in a variance decomposition model and show some theoretical results regarding the consequences of ignoring the reverse causation if it actually exists. Section 5 deals with estimation issues. Section 6 contains the empirical application. The paper concludes, in §7, with managerial implications and directions for future research.

2. Spatial Dependencies in Multimarket Data

2.1. Spatial Dependence in Observed Market Shares and Prices

In marketing data, a general cause of spatial dependence is that economic agents, such as retailers and food distributors, are spatially organized. Indeed, trade territories of these channel members typically cover a set of contiguous markets.

Retailers are interested in the local performance of a brand in their own sales territory, unlike manufacturers of national brands who have more global goals. Thus, when retailers affect the observed prices through selective promotion policies, they do so locally. As a consequence, structural price differences across markets can exist for a given brand if it leads in one market but lags in another. In contrast, manufacturers are limited by the Robinson-Patman Act to price-discriminate across retailers (and therefore space).

There are many reasons why a nationally distributed brand might lead in one market but lag in another, even in mature and undifferentiated product categories. This demand for heterogeneity can be caused by sheer inertia of initial market conditions (see, e.g., Arthur 1994), local order-of-entry effects, or can simply reflect regional consumer tastes.

If promotions only occur in markets where a brand was in high demand to start with, then, along the cross-sectional dimension, market shares and prices (or promotions) covary negatively (positively) even in absence of a demand effect. We show later that pooling such data cross-sectionally implies exaggerated promotion and price effects.

2.2. Separating the Simultaneity

We propose to separate the actual market response effects from the reverse causation using a simultaneous equation model in which the equations describing the demand and the marketing variables share a common unobserved demand state. This unobserved demand state has a spatial variance component structure. Loosely speaking, what this achieves is that the market share and price data are corrected for a common source of variance that has a local structure. This effectively filters out the retailers' actions from the data because the set of retailers is locally constant.

In other work in marketing, e.g., Villas-Boas and Winer (1999), the endogeneity problem is resolved through the use of an instrumental variables (IV) approach. Indeed, time series lagged prices and promotion variables are often good candidates for instruments.[1] However, in our case, the endogeneity applies to cross-sections. In this context, Nevo (2000) suggests using price data from other markets as instruments (see also Hausman 1996). This approach works well if promotion strategies are set independently across markets, but there is still shared information in prices across markets. However, because retailers influence the promotion variables, it may not be true in general that promotions are independent across markets where the same retailer operates. We revisit the issue of instruments later.

3. Representing Space

3.1. Contiguity Classes

Whereas sampling of time-space processes at regular intervals in time has become a standard of many automated data collection procedures, it is generally not possible to sample the cross-sections of a time-space process on a regular lattice (e.g., domestic U.S. retail markets form irregularly spaced sample units). When confronted with irregularly spaced data, a useful concept is a contiguity class, which is simply the set of direct-neighbor locations (Anselin 1988).

One particularly useful definition of contiguity is obtained by creating imaginary borders around the geographical sample units so that every point in space belongs to its least distant sample point. This creates so-called Voronoi polygons. Figure 1 illustrates Voronoi contiguity for the sample that we had at our disposal.

---

[1]However, this is not the case in general. For instance, lagged prices are not good instruments when reference prices exist and are historically formed.

Figure 1   Definition of Contiguity for the IRI Markets
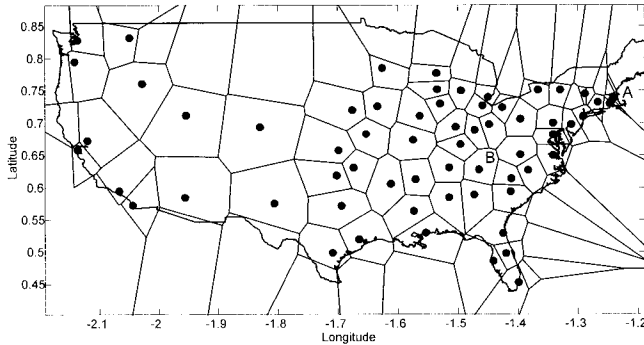
3.2.   Spatial Lags

Contiguity relations among the data can be incorporated into econometric models through a spatial lag. Since no a priori restrictions exist on the direction of the spatial lag, it is customary to define the spatial lag of an observation y_i at location $i = [1, \ldots, N]$ as a weighted sum of all the observations in its contiguous set $\{J_i\}$. Thus, define the spatial lag operator on market i, $\mathbf{L}_i^s$ by

$$\mathbf{L}_i^s y_i = \sum_{\forall \, i' \in \{J_i\}} w_{ii'} y_{i'}, \qquad i' \neq i. \tag{1}$$

The contiguous set of observations $\{J_i\}$, excludes y_i itself, and hence, so does the spatial lag of y_i. Stacking all cross-sectional observations in a vector $\mathbf{y}$, we can write the spatial lag operator as $\mathbf{L}^s \mathbf{y} = \mathbf{W} \cdot \mathbf{y}$, where $\mathbf{W}$ is an $N \times N$ weight matrix with a zero-diagonal. The matrix is said to be standardized if its rows sum to one. If the weights are equal, and the spatial lag operator $\mathbf{W}$ is standardized, then spatial lag is equal to the average of its contiguous observations (Anselin 1988). In this paper, we use the latter concept of a spatial lag in conjunction with Voronoi contiguity. Both concepts are easy to compute and require minimal extra data. That is to say, to compute $\mathbf{W}$, only market coordinates are needed and a program that computes contiguities among data.

The weights alternatively can be modeled as a general distance function that explicitly accounts for direction dependence. In yet another operationalization, the cross-sectional weights can be modeled as the generalized distance derived from a multidimensional scaling analysis. In marketing several examples of

the latter in the context of discrete choice are Elrod (1988), Erdem (1996), Erdem and Winer (1999), and Kamakura and Russell (1984).

4.   Model Development

4.1.   The Variance Components Model

We consider a general linear system of market shares and marketing variables. In this system, we allow for cross-market, within-market, and cross-variable correlation in the data. We use market share data because they can be compared across markets of different size, and because they are a good measure of the relative size of a brand to local retailers.[2]

Locations are indexed by $i = 1, \ldots, N$, time is indexed by $t = 1, \ldots, T$, and the marketing variables are indexed by $j = 1, \ldots, P$. The system contains cross-sections of variables and can be introduced as follows:

$$\begin{aligned}
\mathbf{y}_t &= \alpha_0 \boldsymbol{\iota}_N + \sum_{j=1}^{P} \beta_j \cdot \mathbf{x}_{jt} + \mathbf{e}_{yt} \\
\mathbf{x}_{1t} &= \alpha_1 \boldsymbol{\iota}_N + \mathbf{e}_{x1t} \\
&\vdots \\
\mathbf{x}_{Pt} &= \alpha_P \boldsymbol{\iota}_N + \mathbf{e}_{xPt},
\end{aligned} \tag{2}$$

in which $\boldsymbol{\iota}_N$ is a column vector of ones, $\mathbf{y}_t$ is a cross-section of market shares, $\mathbf{x}_{jt}$ is a cross-section of the jth marketing variable, and the $\mathbf{e}$'s are cross-sections of random components, all of size $N \times 1$. In this system, the α's are the global intercepts of market shares and marketing mix parameters, while the β's are the effects of the marketing mix and are the quantities of interest. The effects β are, for now, homogenous. In the empirical section, we also allow for heterogeneity in β. The equations for the marketing variables are simple but obtain meaning by how they are related to the market share equations through their respective stochastic components. Specifically, it is proposed that the stochastic terms $\mathbf{e}$ follow a general spatiotemporal process with the following characteristics:

[2]Alternatively, market-level or account-level "sales velocity," i.e., sales normalized by all commodity volume (ACV), can also be compared across markets of different size.

$$\mathbf{e}_{yt} = \boldsymbol{\mu} + \boldsymbol{\xi}_t$$

$$\boldsymbol{\mu} = \lambda \cdot \mathbf{W} \cdot \boldsymbol{\mu} + \boldsymbol{v}$$

$$\boldsymbol{\xi}_t = \rho_0 \cdot \boldsymbol{\xi}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\mathbf{e}_{xjt} = \gamma_j \boldsymbol{\mu} + \boldsymbol{\delta}_{jt} \quad \text{and}$$

$$\boldsymbol{\delta}_{jt} = \rho_j \cdot \boldsymbol{\delta}_{jt-1} + \theta_j \boldsymbol{\tau}_t + \boldsymbol{v}_{jt}. \tag{3}$$

The vectors $\mathbf{e}_{yt}$, $\mathbf{e}_{xjt}$, $\boldsymbol{\xi}_t$, $\boldsymbol{\epsilon}_t$, $\boldsymbol{\mu}$, $\boldsymbol{v}$, $\boldsymbol{\delta}_{jt}$, $\boldsymbol{v}_{jt}$, $\boldsymbol{\tau}_t$ are all of size $[N \times 1]$.

The vector $\boldsymbol{\mu}$ contains a cross-section of time-invariant random intercepts with mean zero. They are related across space according to a known weighting matrix $\mathbf{W}$. The random intercepts can be interpreted as the baseline market-shares net of the effect of promoted prices and other promotion variables. They fulfil the role of the ''unobserved market shares.''

The vectors $\boldsymbol{\xi}_t$ and $\boldsymbol{\delta}_{jt}$ both contain cross-sections of stochastic terms that are related through time but not through space.

The vector $\boldsymbol{\tau}_t$ contains the common shocks to marketing variables. These shocks are neither related through time nor space.

The coefficient λ is the spatial autoregressive effect, the ρ's are the temporal autoregressive effects (specific to market shares and marketing variables), and the θ_j's are the factor coefficients (of which one has to be set to unity).

All innovation terms in the model are spherical, i.e., $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2 \cdot \mathbf{I}_N)$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma_\epsilon^2 \cdot \mathbf{I}_N)$, $\boldsymbol{v}_{jt} \sim \mathcal{N}(0, \sigma_{vj}^2 \cdot \mathbf{I}_N)$, and $\boldsymbol{\tau}_t \sim \mathcal{N}(0, \sigma_\tau^2 \cdot \mathbf{I}_N)$, and are independent across time, space, marketing variables, and each other.

From this model, the ''independent'' marketing variables are only truly independent of the $\mathbf{e}_y$ if the vector $[\gamma_1 \cdots \gamma_P]$ is zero. If not, then stochastic terms in the first equation of System (2) will be related to the marketing variables. For instance, if retailers ''censor'' promotion variables on the basis of the market shares, then on average promotional variables will be higher when the share is high. This behavior is effectively captured through the γ's. We can therefore directly test for endogeneity in the jth marketing variable by testing against the null hypothesis that $\gamma_j = 0$.

The model and its account for both sides of the system are inspired by Chamberlain and Griliches (1975), who also account for the endogeneity problem

through an unobservable variable with a variance component structure. We now express the variance covariance structure of the equations through space (first) and then through the combination of time and marketing variables.

4.2. Spatial Structure
The spatial component in the panel data is modeled using the $[N \times N]$ spatial contiguity matrix $\mathbf{W}$ introduced in the previous section. Specifically, the random intercepts are assumed to be related as follows (Anselin 1988):

$$\boldsymbol{\mu} = \lambda \cdot \mathbf{W} \cdot \boldsymbol{\mu} + \boldsymbol{v}. \tag{4}$$

Equation (4) is identified because the diagonal of $\mathbf{W}$ is zero. We recall that the ith row of the spatial weighting matrix $\mathbf{W}$ is zero everywhere except for those i' contiguous to i. Therefore, Equation (4) implies that if the spatial autoregression parameter $\lambda \approx 0$, then the random intercept i is unrelated to the intercepts of the spatially contiguous units i', whereas if $\lambda \approx 1$, then a given random intercept is the average of the intercepts in the markets that are located around i.

From Equation (4), $[\mathbf{I}_N - \lambda \cdot \mathbf{W}] \cdot \boldsymbol{\mu} = \boldsymbol{v}$, or equivalently,

$$\boldsymbol{\mu} = [\mathbf{I}_N - \lambda \cdot \mathbf{W}]^{-1} \cdot \boldsymbol{v} \equiv \mathbf{B} \cdot \boldsymbol{v}. \tag{5}$$

Therefore, given that $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2 \cdot \mathbf{I}_N)$, the random intercepts have the following cross-sectional covariance structure for all t (Anselin 1988):

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = E(\mathbf{B}\boldsymbol{v}\boldsymbol{v}'\mathbf{B}') = \sigma_v^2 \cdot \mathbf{B}\mathbf{B}' \equiv \Gamma_{(N \times N)}. \tag{6}$$

So that the inversion in Equation (5) exists, the values of λ will need to fall between $[\eta_{\min}^{-1}, \eta_{\max}^{-1}]$, where $\eta_{\min}$ and $\eta_{\max}$ are the minimum and maximum eigenvalue of the matrix $\mathbf{W}$. Typically, these values are near -1 and $+1$ (see, e.g., LeSage 1998).

4.3. Intertemporal and Intervariable Structure
In the case where $T > 1$, the intertemporal component in the panel data is modeled using a first-order autoregressive formulation of the errors. We derive the temporal structure of the time series ξ_{it} and δ_{ijt} for $t = 1, \ldots, T$. Array the stochastic terms in time series, so that $[\boldsymbol{\xi}_i' \boldsymbol{\delta}_{i1}' \cdots \boldsymbol{\delta}_{iP}']'$ is a $(P + 1)T \times 1$ column

vector of all observations in market i arranged by variables and time. It is shown in Appendix A that the $T \cdot (P + 1) \times T \cdot (P + 1)$ covariance matrix of these $(P + 1)$ time series for each market i is equal to

$$\Psi = \begin{pmatrix} \sigma_\epsilon^2 \mathbf{M}_T^{00} & 0 & 0 & 0 \\ 0 & (\theta_1^2 \sigma_\tau^2 + \sigma_{\nu 1}^2) \mathbf{M}_T^{11} & \cdots & \theta_1 \theta_P \sigma_\tau^2 \mathbf{M}_T^{1P} \\ 0 & \vdots & \ddots & \vdots \\ 0 & \theta_P \theta_1 \sigma_\tau^2 \mathbf{M}_T^{P1} & \cdots & (\theta_P^2 \sigma_\tau^2 + \sigma_{\nu P}^2) \mathbf{M}_T^{PP} \end{pmatrix},$$

$$\forall \, i, \tag{7}$$

in which $M_T^{jj'}$ is an asymmetric $T \times T$ matrix, defined as

$$\mathbf{M}_T^{jj'} = \frac{1}{1 - \rho_j \rho_{j'}} \begin{pmatrix} 1 & \rho_{j'} & \cdots & \rho_{j'}^{T-1} \\ \rho_j & 1 & \cdots & \rho_{j'}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-1} & \rho_j^{T-2} & \cdots & 1 \end{pmatrix}. \tag{8}$$

The asymmetry stems from the differences in the ρ_j across j, where $j = 0$ denotes the stochastic structure of market shares and $j = 1, \ldots, P$ denotes that of the marketing variables. When the autoregression parameters ρ_j are equal, the usual symmetric result obtains (see Judge et al. 1985, p. 284). If so desired, it is standard to include higher order time effects.

4.4. The Total Covariance Structure

We can now consider the total covariance structure of the model. Stack the entire system by cross-section, time, and variables to form one single column vector of stochastic components (of size $NT(P + 1) \times 1$), i.e.,

$$\mathbf{e} = [\mathbf{e}_{y1}' \cdots \mathbf{e}_{yT}' \; \mathbf{e}_{x11}' \cdots \mathbf{e}_{x1T}' \cdots \mathbf{e}_{xP1}' \cdots \mathbf{e}_{xPT}']'. \tag{9}$$

Define the vector $\mathbf{a} = [1 \; \gamma_1 \cdots \gamma_P]'$. Given the assumption of independence of the cross-sectional and the intertemporal structure, it follows that the vector $\mathbf{e}$ has a variance-covariance matrix $\Omega_{NT(P+1) \times NT(P+1)}$ equal to

$$\Omega = \Psi \otimes \mathbf{I}_N + (\mathbf{a} \otimes \iota_T)(\mathbf{a} \otimes \iota_T)' \otimes \Gamma, \tag{10}$$

where ι_T is a $[T \times 1]$ vector of ones.

4.5. Ignoring Spatial Dependence

Suppose the process in Equation (2) is the data-generating mechanism but we omit the spatially depen-

dent μ from it. In this case, we have $P + 1$ independent equations, and inferring the vector β can be done by applying ordinary least squares (OLS) to the first equation. To illustrate the biased nature of the OLS estimates if μ is ignored incorrectly, recall that the least squares estimate of the parameters is equal to $\hat{\beta}_{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{e})$. Hence, $\text{plim} \, \hat{\beta}_{ols} \neq \beta$ if $\text{plim}(\mathbf{X}'\mathbf{e}) \neq 0$. Using the decomposition $\mathbf{e}_y = \iota_T \otimes \mu + \xi_y$, current interest is with the asymptotic properties of $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\iota_T \otimes \mu)$ because of the dependence of the X's and the μ's.

To make the argument more concise, we make two simplifying assumptions. The first assumption is that $\mathbf{X}$ consists of only one variable (say price), while the second is that its cross-sectional mean is zero. Neither of these assumptions changes the basic result. Decompose the X_{it} using the within-market means $\bar{X}_i$ into $X_{it} = \bar{X}_i + (X_{it} - \bar{X}_i)$. In notation used above, the cross-market variance in $\bar{X}_i$ is $\gamma^2 \sigma_\mu^2$ and the within-market variance in the time series $(X_{it} - \bar{X}_i)$ is σ_δ^2 in all markets i (here σ_δ^2 denotes the variance of the time component). Then $\text{plim}(1/NT)(\mathbf{X}'\mathbf{X}) = \gamma^2 \sigma_\mu^2 + \sigma_\delta^2$.

Further, because $\Sigma_t (X_{it} - \bar{X}_i) \cdot \mu_i = 0$, $\forall i$, the term $\mathbf{X}'(\iota_T \otimes \mu)$ only involves the market averages $\bar{X}_i$, which are in expectation equal to $\gamma \mu_i$. The covariation between the market averages $\bar{X}_i$ and the random intercepts μ_i is therefore $\gamma \sigma_\mu^2$. Using $\text{plim}(1/N \cdot T)E(\mathbf{X}'(\iota_T \otimes \mu)) = \gamma \sigma_\mu^2$, and Slutsky's theorem (see Greenberg and Webster 1991, p. 8), we obtain, as the probability limit for the bias,

$$\text{plim}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\iota_T \otimes \mu)$$

$$= \frac{\gamma \sigma_\mu^2}{\gamma^2 \sigma_\mu^2 + \sigma_\delta^2} = \begin{cases} \dfrac{1}{\gamma} \cdot \text{var}_X(\text{spatial}) & \text{if } \gamma \neq 0 \\ 0 & \text{if } \gamma = 0, \end{cases} \tag{11}$$

in which $\text{var}_X(\text{spatial})$ is the fraction of the total price variation that is common with the intercepts. From Equation (11), the bias is stronger when the fraction of price variation due to dependence on the random intercepts is higher and the variance in the X's is lower.

Because of selective discounting by the retailer, we expect γ to be negative for prices and positive for promotions. In other words, we expect that with

higher market shares, we tend to observe also lower prices and higher promotion levels. This implies that ignoring spatial dependence will inflate both the inferred price and the promotion response.

4.6. Alternative Approaches

An alternative approach to solving the inference problem considered here is the IV approach. We already noted that spatial lags are not likely to be good instruments because of the possibility that these are correlated with the error terms in the model. It is known that $\text{plim}(\beta_{IV} - \beta)$ is very sensitive to violations of independence (Bound et al. 1995, Staiger and Stock 1997). It should therefore not be surprising that the IV estimator with spatially lagged prices for instruments can be prone to even larger biases than the OLS estimator.

In addition, whereas accounting for the spatial dependence in the data will make the estimates of $\boldsymbol{\beta}$ more efficient (Case 1991), the use of instrumental variable estimation may be less efficient, depending on the choice for instruments (see Judge et al. 1985).

5. Estimation

The model defined by (2) and (3) can be estimated through maximum likelihood under the assumption of normality. Maximum likelihood estimation achieves the desirable properties of consistency, asymptotic efficiency, and asymptotic normality (see Anselin 1988, p. 60 ff.). The likelihood of the observations given the parameters is

$$\ell(\mathbf{y} \,|\, \lambda, \boldsymbol{\rho}, \boldsymbol{\theta}, \sigma_\epsilon, \sigma_\upsilon, \boldsymbol{\sigma}_\nu, \boldsymbol{\beta}, \boldsymbol{\alpha})$$

$$\propto |\Omega|^{-1/2} \cdot \exp\left(-\frac{1}{2}\mathbf{e}'\Omega^{-1}\mathbf{e}\right), \qquad (12)$$

with $\mathbf{e}$ defined in Equation (3) and Ω in Equation (10) previously. The number of parameters is $5 + 6 \times P$ and is linear in the number of equations.

The computation of the likelihood of the observations requires evaluation of $|\Omega|$ and Ω^{-1}. The inversion problem Ω^{-1} is of dimension $NT(P + 1)$ and will therefore quickly run into computational problems. However, using Magnus (1982) and the special nature

of Ω, this problem can be reduced to one of dimension $\max\{N, T(P + 1)\}$, which allows us to obtain maximum likelihood estimates even for very large problems. For instance, we will illustrate the model using 4 weeks of data, 64 markets, and 3 marketing variables (plus the market data). The total number of observations in the system is therefore $4 \times 64 \times (3 + 1) = 1{,}024$. At the current state of computing, it is burdensome to deal with an inversion of a $1{,}024 \times 1{,}024$ matrix in a maximum likelihood algorithm. Magnus's results, however, allow for a partitioning of the matrix that reduces the inversion problem to one of size $\max\{64, 4 \times (3 + 1)\} = 64$. Appendix B gives the log-likelihood function.

6. Empirical Analysis

6.1. Data

The data used in this analysis contain sales, price, feature, and display data for two product categories, Mexican hot sauce and tortilla chips, collected from 64 different domestic U.S. markets. Although only short windows of time relative to the number of markets are used in the estimations, there are 104 weeks of data to draw from, which allows for careful benchmarking of the estimation results obtained from spatial, OLS, and IV approaches.

The data are collected using the INFOSCAN sample of 3000 national stores. The raw data are at the brand/market/week level. In both data sets, the top five national brands were extracted. Brands with largely missing temporal data were disregarded in a given market. The major brands in the Mexican hot sauce category are Pace, Tostitos, and Old El Paso, whereas the major brands in the tortilla chips category are Doritos, Tostitos, and Santitas. These brands account for more than 70% of observed category volume in either category and are the focus of the empirical investigation.

Market shares of the brands in all 64 markets were computed on the basis of volume sold of the top five brands. Price and promotion data were computed relative to their market averages. Distribution data were also available but are not used in the empirical anal-

Table 1 Averages of the Data Across Time and Markets

Category	Brand	Share*	Price**	Feature	Display
Tortilla chips	Doritos	0.42	1.13	2.04	1.34
	Tostitos	0.31	1.10	1.18	1.42
	Baked Tostitos	0.09	1.36	0.43	1.27
	Santitas	0.10	0.67	0.92	0.43
	Eagle	0.08	0.73	0.40	0.51
Mexican hot sauce	Pace	0.37	0.96	1.21	0.81
	Old El Paso	0.27	0.92	0.41	0.43
	Tostitos	0.18	1.27	1.68	2.67
	Chichi's	0.16	0.91	0.92	0.37
	Las Palmas	0.08	0.81	0.34	0.06

*Because of limited availability in some markets, shares do not add to unity.
**The price, feature, and display data are defined relative to their averages in each week/market.

ysis because most brands have full distribution and lack empirical variation. Table 1 shows the averages of the data across time and markets conditional on availability.

The 64 U.S. markets are identified either by a city, a pair of cities, or a (part of a) state. Latitude and longitude coordinates were collected for these markets using an Internet mapping service (http://www. indo.com/distance). In cases where a market was identified with a pair of cities or a state, an interior point was taken as the relevant coordinate.

Figure 1 shows the contiguity patterns based on Voronoi polygons for the cross-sectional sample of 64 cities. It is worth noting the relative sparse sampling of the western part of the United States, which loosely follows population density. The number of contiguous markets ranges from 1 (Boston—located at A) to 8 (Knoxville—located at B). In general, the Voronoi polygons represent the relative isolation of the coastal areas well.

6.2. Spatiotemporal Market Structure
In the context of our motivating example of unobserved retailer behavior, the two categories under analysis have very different characteristics that are worth discussing at the outset of the empirical analysis.

The Mexican hot sauce category is cross-sectionally very heterogenous. The national market leader is Pace, but this brand leads in only 46% of the markets,

and it is first or second in only 66% of the cases. Old El Paso, which is a local market leader in 39% of the markets, is among the two leaders in 70% of the markets.

The tortilla chips category is cross-sectionally comparatively homogeneous. For instance, Doritos is the market leader in 89% of the markets and is either first or second in all markets. Likewise, Tostitos is either first or second in 81% of the markets. The remaining three brands are very rarely second largest.

This disparity between categories implies that the Mexican hot sauce category is more likely than the tortilla chips category to display cross-sectional differences in pricing and merchandizing if retailers apply market leadership or co-leadership as a necessary condition to obtain promotions. It is therefore expected that the problems in cross-sectional inference are less severe in the tortilla chips category than in the Mexican hot sauce category. Nonetheless, it is shown that for both categories the same principal result holds.

6.3. Testing for Spatial Dependence
Several procedures exist to statistically test for the presence of spatial dependence in cross-sectional data, against the null hypothesis of spatial independence (see Anselin 1988 for a full discussion). We present one of these: Moran's I statistic. If the rows in the spatial weight matrix $\mathbf{W}$ add to unity, Moran's I statistic for a single cross-section of N data points is equal to

$$I_{yt} = \frac{\mathbf{e}'_{yt} \cdot \mathbf{W} \cdot \mathbf{e}_{yt}}{\mathbf{e}'_{yt} \cdot \mathbf{e}_{yt}} \quad \text{and} \quad I_{xjt} = \frac{\mathbf{e}'_{xjt} \cdot \mathbf{W} \cdot \mathbf{e}_{xjt}}{\mathbf{e}'_{xjt} \cdot \mathbf{e}_{xjt}}. \quad (13)$$

The asymptotic distribution of Moran's I statistic is normal with means and variance detailed in Anselin (1988). The interpretation of this statistic is that larger values indicate that the terms $\mathbf{e}$ are positively correlated with their spatial lags $\mathbf{W} \cdot \mathbf{e}$.

The test for spatial dependence of regression residuals was carried out on all available cross-sections of data. So, for each variable, the tests for spatial dependence were carried out per brand 104 times, for each subsequent cross-section of 64 markets. The tests are carried out on the residuals of the

Table 2 Moran's *I* Test for Spatial Dependence in the Residuals

	Mexican Hot Sauce			Tortilla Chips		
	Pace	Tostitos	Old El Paso	Doritos	Tostitos	Santitas
Share residuals						
Moran's *I*	0.504	0.345	0.557	0.453	0.499	0.372
z_I	6.589	4.685	7.102	5.933	6.583	4.544
Significance*	1.000	1.000	1.000	1.000	1.000	1.000
Price residuals						
Moran's *I*	0.443	0.610	0.360	0.327	0.414	0.489
z_I	5.502	7.514	4.515	4.118	5.153	5.416
Significance*	1.000	1.000	1.000	0.971	1.000	1.000
Feature residuals						
Moran's *I*	0.193	0.192	0.128	0.195	0.237	0.290
z_I	2.502	2.492	1.725	2.524	3.034	3.303
Significance*	0.769	0.702	0.548	0.654	0.817	0.875
Display residuals						
Moran's *I*	0.292	0.177	0.046	0.199	0.259	0.239
z_I	3.689	2.317	0.748	2.581	3.291	2.759
Significance*	0.894	0.721	0.231	0.721	0.875	0.837

*Significance is computed as the fraction of the 104 weeks that the test statistic was significant, i.e., $1/T \sum_{t=1}^{T} D_{p(z_l)<0.05}$, where $D_{p(z_l)<0.05}$ is an indicator dummy of significance in sample t.

market share and marketing variables equations in the System (2). All analyses were run on log transforms of the share data and log transforms of the price and promotion data (to circumvent the log-of-zero problem, 0.01 was added to the original variables). Table 2 lists the tests results for all major brands and variables.

The results of the tests using the Moran I and likelihood ratio statistics indicate that the residuals of the log market shares of the brands are all spatially dependent. For instance, the first row in Table 2 indicates that across the 104 weekly cross-sections of 64 markets for Pace, the average Moran I across the 104 replications of the test was equal to 0.504 for market shares. More generally, the I statistic is in the range of 0.35–0.56 for the share data of the major brands, and the z statistic is in the 4.5–7.1 range. The tests were significant at the 0.05 level for all brands across all replications.

From Table 2, we see further that prices are also spatially dependent, whereas the feature and display

data, which are more volatile, are spatially dependent in the majority of replications of the test (short of the display data for Old El Paso). This exception aside, the test results support the conclusion that both the residuals in market shares as well as in marketing mix variables of national brands across national markets are spatially dependent.

6.4. Model Test and Empirical Estimates

6.4.1. Controlled Data Experiment. We test the model in a controlled data experiment for two reasons. First, we want to get an idea of the extent of biases when OLS estimators are used on the type of data considered here. Second, we want to illustrate that the data-generating process can be empirically identified from short time samples of data. We generated data for market shares and one marketing variable, price, by drawing random vectors $\boldsymbol{v}$, $\boldsymbol{\epsilon}_t$, and $\boldsymbol{v}_t$, and by performing the transformations in Equation (5) to construct $\boldsymbol{\mu}$ and in Equation (3) to construct $\boldsymbol{\xi}_t$ and $\boldsymbol{\delta}_t$. For the spatial component, the contiguity matrix $\mathbf{W}$ was computed from the actual Voronoi contiguity classes in the IRI markets. Along the time dimension, data were generated for 4 weeks. The market share and price data were generated to depend both on $\boldsymbol{\mu}$ through the γ parameter, just as in Model (3). Data were generated with the same scenario 25 times for each of 8 different sets of parameter values. All parameter values were chosen to cover their actual empirical range. Given the bias in Equation (11), the scenarios concentrate on variation in the data-generating parameters on λ, σ_v, and the dependence γ between prices and market shares. The other parameters were as follows: $\alpha_0 = -1.0$, $\alpha_1 = 0.0$, $\rho_0 = 0.4$, $\rho_1 = 0.4$, $\sigma_\epsilon = 0.1$, $\sigma_v = 0.03$, and $\beta = -2$. The estimation results are in Table 3.

Two results are important. First, as argued before, OLS estimates of the β coefficient are biased away from zero. What is striking in the OLS estimates is the extent of the bias, even in cases where only small components of the demand data are common to the price data. However, when this common structure is properly accounted for in full model, there is no bias in the estimator for the price effect β. Second, our

Table 3 Estimation Using Synthetic Data

Data Generating Parameters	$\lambda = 0.65$				$\lambda = 0.90$			
	$\sigma_v = 0.15$		$\sigma_v = 0.30$		$\sigma_v = 0.15$		$\sigma_v = 0.30$	
	$\gamma = -0.10$	$\gamma = -0.05$	$\gamma = -0.10$	$\gamma = -0.05$	$\gamma = -0.10$	$\gamma = -0.05$	$\gamma = -0.10$	$\gamma = -0.05$
Maximum likelihood								
λ	0.61	0.61	0.64	0.63	0.88	0.87	0.88	0.89
σ_v	0.15	0.15	0.30	0.30	0.15	0.15	0.30	0.30
σ_ϵ	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
σ_ν	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
β_{ML}	−1.99	−1.96	−1.95	−1.95	−1.95	−2.04	−2.02	−1.99
α_0	−1.00	−1.00	−1.02	−1.01	−1.03	−0.99	−0.93	−0.97
α_1	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00
γ	−0.11	−0.05	−0.10	−0.05	−0.10	−0.05	−0.10	−0.05
ρ_0	0.38	0.40	0.41	0.38	0.41	0.41	0.39	0.39
ρ_1	0.79	0.80	0.79	0.80	0.80	0.80	0.79	0.80
OLS								
β_{OLS}	−3.30	−2.68	−5.68	−4.59	−5.22	−3.84	−8.71	−9.16
α_0	−1.00	−1.00	−1.01	−1.01	−1.02	−1.00	−0.96	−1.00

variance decomposition model correctly reproduces the structure used to generate the data. This is to say, the model's parameters can be identified from short time samples.

6.4.2. Operationalizations. The actual empirical analysis is set up as follows. For each brand, we estimate a log-log model of market shares against the relative price, feature, and display variables. There exist 104 cross-sections in 2 years of data, and so it is possible to run the SpatioTemporal Model (3) repeatedly. Doing so will yield a total of $104 - T$ sets of estimates of the parameters of interest for each model. Having multiple sets of parameter estimates is helpful in assessing whether the presence of the endogeneity problem is incidental or structural.

To obtain the maximum likelihood estimates, we concentrated first on finding good starting values for the parameters by estimating the model in parts. The full model, containing 23 parameters, was run subsequently over the moving windows of time. We used some logical limits on the parameter space (e.g., positivity constraints for the variances) and imposed very wide feasibility bounds on the parameter space (e.g., price elasticities between −10 and +10). These bounds were not binding in the estimations

and were used to avoid bad jumps in the initial steps of the algorithm where the algorithm still has imperfect estimates of the gradients.[3] The experience with the estimations is that the model converges rapidly.

6.4.3. Test Results for Endogeneity. For space reasons, we limit our discussion to estimations using windows of 4 weeks ($T = 4$). Table 4 shows the average estimates across the 100 repeated estimations. Most of model parameters are consistently significant at the 0.05 level. To report on significance across replications, Table 4 notes those parameters that do not have average absolute t statistics in excess of 2.

The estimates in Table 4 show the following results. First, the spatial autocorrelation parameter λ is large and positive for all brands. Second, the vector $\gamma = [\gamma_1\ \gamma_2\ \gamma_3]'$ contains the parameters that measure the

[3]In a few cases, we did find that the bounds were binding. For instance, the lower bound for temporal variation in price was binding in the few cases where price remains nearly constant over 4 weeks (as one would expect). For the Dorito brand, we found that in 3 out of 100 cases the bounds on γ_1 were binding. This is because estimates of σ_v^2 for this brand can become very small, essentially making the estimates of the γs irrelevant (again as expected—see also the subsequent discussion of the Dorito brand).

Table 4 Estimation Results Based on 4 Weeks of Data[a, b]

	Mexican Hot Sauce			Tortilla Chips		
	Pace	Tostitos	Old El Paso	Doritos	Tostitos	Santitas
λ	0.863	0.803	0.800	0.828	0.856	0.723
σ_v	0.261	0.160	0.239	0.074	0.136	0.311
σ_ϵ	0.097	0.093	0.104	0.082	0.100	0.153
$\sigma_{v,price}$	0.020	0.013	0.027	0.052	0.035	0.038
$\sigma_{v,feature}$	1.847	1.944	1.815	1.604	1.833	1.746
$\sigma_{v,display}$	0.778	0.116	0.907	0.277	0.249	0.665
σ_τ	1.020	0.487	0.965	1.339	1.338	1.015
β_{price}	−1.692	−1.645	−1.809	−1.252	−2.034	−1.891
$\beta_{feature}$	0.006	0.004*	0.011	0.009	0.008	0.014
$\beta_{display}$	0.024	0.284	0.021	0.107	0.149	0.069
α_0	−1.193	−1.574	−1.513	−0.640	−0.918	−2.891
α_1	−0.042	0.241	−0.079	0.141	0.107	−0.378
α_2	−1.596	−2.966	−0.775	−0.669	−1.642	−2.958
α_3	−0.859	0.934	−1.793	0.178	0.288	−1.506
γ_1	−0.052	−0.223	0.008*	−0.011*	0.057*	−0.088
γ_2	1.081	0.954*	0.431*	1.485*	0.080*	1.726
γ_3	1.212	−0.042*	0.861	−0.597*	0.712	1.633
ρ_0	0.380	0.499	0.394	0.451	0.404	0.626
ρ_1	0.889	0.858	0.824	0.629	0.727	0.634
ρ_2	0.259	0.328	0.226	0.238	0.329	0.305
ρ_3	0.754	0.804	0.760	0.733	0.737	0.761
θ_1	−0.027*	−0.070	−0.042*	−0.041	−0.028	−0.062
θ_2	1.000	1.000	1.000	1.000	1.000	1.000
θ_3	0.279	0.379	0.244*	0.093	0.076	0.331

[a]Parameters are averaged for 100 successive estimations on moving windows of 4 weeks.
[b]All estimates have average absolute *t* statistics >2 unless superscripted by an asterisk.

Table 5 Fraction of Spatial Variance in the Data

	Mexican Hot Sauce			Tortilla Chips		
	Pace	Tostitos	Old El Paso	Doritos	Tostitos	Santitas
Market shares	0.936[a]	0.737	0.901	0.662	0.837	0.623
Prices	0.185	0.405	0.017	0.047	0.056	0.238
Feature	0.076	0.028	0.016	0.029	0.013	0.089
Display	0.239	0.015	0.063	0.049	0.197	0.258

[a]Values are averages for 100 successive estimations on moving windows of 4 weeks.

dence between prices and the unobserved component of shares. For Old El Paso and Tostitos tortilla chips, the display data are endogenous.

There is no clear evidence of endogeneity in any variable for Doritos. This brand is the *only* brand in our data that leads or is second in all of the U.S. markets (note also the low standard deviation σ_v in the spatial component for this brand). This fact accords nicely with the stylized observation that retailers promote brands on the basis of extant market (co)leadership.

We conclude that there is substantial support for the fact that observed price and promotion data depend on the market shares.

6.4.4. Results for the Factor Model. The θ's are all of the expected sign. One of these parameters has to be fixed to an arbitrary value. We set $\theta_2 = 1$. The remaining parameters suggest that prices covary negatively with the promotion variables, which in turn covary positively among themselves.

6.4.5. Variance Components. Table 5 shows the fraction of the unobserved variation in the data explained by the spatial component of the model as opposed to the time component. The fraction of variance explained by the spatial component in variable $j = 0, 1, \ldots, P$ is defined as $T \cdot \text{trace}(\gamma_j^2 \Gamma)/[T \cdot \text{trace}(\gamma_j^2 \Gamma) + N \cdot \text{trace}(\Psi_{jj})]$, where the last term is the jth $T \times T$ submatrix on the diagonal of Ψ, and $\gamma_0 = 1$.

From the table, it is observed that market shares are almost completely spatial in terms of variation. This is a surprising finding in and of itself, given the low degree of product differentiation across brands,

degree to which prices and promotions depend on baseline shares (intercepts). For instance, for Pace this vector is equal on average to [−0.052, 1.081, 1.212], and it is consistently signed and significant in each of the 100 windows (104 weeks minus 4 weeks for estimation). This means that high baseline shares consistently go hand in hand with low prices, high feature, and high display levels, exactly as proposed in the introduction. This pattern generalizes to other brands.

Namely, when significant, the vector γ contains negative values for the association of share and price and positive values for the association of share and promotion variables. For Old El Paso, Doritos, and Tostitos tortilla chips, there is no evidence for depen-

because it implies that market shares vary more across markets than within markets even in the same competitive set. Market structures in the categories under analysis are therefore local with a spatial component.

Overall, the marketing variables share this unobserved spatial component. Prices for Pace, Tostitos hot sauce, and Santitas tortilla chips have sizable spatial components. The spatial components of display are also large for these brands. Prices for brands with less cross-sectional variance in market share also have smaller spatial components (Old El Paso, Doritos, and Tostitos tortilla chips).

6.5. Comparison of Response Estimates and Validation

Besides determining the degree to which the share and marketing mix data are jointly dependent and providing a test for endogeneity of the latter, a key aspect of our investigation involves the consequences of this endogeneity in estimating marketing mix effects. To determine this aspect of our investigation, two comparisons need to be made. First, we compare the estimates from the spatiotemporal model (SPATIAL) with those from OLS and IV estimations. Second, in addition to the validation results on synthetic data, it is desirable to have an empirical benchmark to validate the estimation results with the spatial model.

With respect to a choice for instruments, factor prices cannot be used because they lack of cross-sectional variation. However, Nevo (2000), in a setting somewhat similar to ours, uses as instruments the mean of marketing variables taken across markets that belong to the same region (excluding the market for which the instrument is constructed). In our notation, Nevo (2000) uses $(\mathbf{I}_T \otimes \mathbf{W})\mathbf{X}$, in which $\mathbf{W}$ is the contiguity matrix defined in §3, and $\mathbf{X}$ is a $NT \times (P + 1)$ matrix of marketing mix variables organized by time and cross-sections.

With respect to an appropriate benchmark, we use the fact that a within-market analysis cannot suffer from the same bias as discussed above, because the retailers are kept constant. Hence, given that we have 104 weeks of data, it is natural to take the average of

Table 6   Comparison of Estimation Results Using 4 Weeks of Data

	Mexican Hot Sauce			Tortilla Chips		
	β_{price}	$\beta_{feature}$	$\beta_{display}$	β_{price}	$\beta_{feature}$	$\beta_{display}$
	Pace			Doritos		
Temporal	−1.286	0.006	0.027	−1.120	0.008	0.183
Spatial	−1.692[a]	0.006	0.024	−1.252	0.009	0.107
OLS	−4.124[a]	0.036	0.183	−1.123	0.019	0.042
IV	0.897[a]	0.101	0.734	−1.005	0.070	−0.322
	Tostitos			Tostitos		
Temporal	−2.149	0.003	0.201	−1.797	0.008	0.259
Spatial	−1.645	0.004	0.284	−2.034	0.008	0.149
OLS	−3.932	0.022	0.286	−1.400	0.011	0.469
IV	−5.137	0.072	−0.244	−1.420	0.000	1.229
	Old El Paso			Santitas		
Temporal	−1.407	0.010	0.024	−1.854	0.011	0.103
Spatial	−1.809	0.011	0.021	−1.891	0.014	0.069
OLS	−1.574	0.022	0.081	−2.012	0.052	0.224
IV	−3.214	−0.008	0.334	2.666	0.059	0.743

[a]Parameters are averaged for 100 successive estimations on moving windows of 4 weeks.

the within-market marketing mix effects as the benchmark for the overall price and promotion effects. This benchmark is referred to as TEMP.

Table 6 compares SPATIAL, OLS, and IV results to TEMP. The estimates from SPATIAL and those of TEMP are mostly very close. In other words, even in absence of long temporal data, the model correctly produces the within-market estimates of marketing mix effects. In contrast, the OLS estimates of price and promotion effects are severely biased away from zero. The IV estimates are also biased. In addition, they are very unstable across the rolling windows of estimation. To show that the SPATIAL estimates are fairly stable, even with only 4 weeks of data, Figure 2 visualizes the price estimates by estimation method across 100 moving data windows for two brands. The top panel applies to Pace hot sauce and the bottom panel to Tostitos hot sauce. Both graphs show the TEMP benchmark, the spatial and the OLS estimates (the IV estimates are omitted in this graph to avoid cluttering). The SPATIAL estimates of price effects for

these two brands are very close to the TEMP benchmark, despite the fact that almost all of the variance in the market share data is cross-sectional (see Table 5). As can also be seen, the OLS estimates are biased away from TEMP.

6.6. Accounting for Response Heterogeneity

Although not the main goal of this paper, it is possible to incorporate heterogeneity in marketing effects across markets into the model. We illustrate this by accounting for random price effects. The approach adopted here is the same as in Hsiao (1986) and can therefore be introduced very briefly.

Assume that instead of price effects β, the data are generated by market-specific random effects $\beta_i \sim \mathcal{N}(\beta, \sigma_\beta^2)$. Arrange all NT observations on price in an $NT \times N$ matrix,

Figure 2 Price Effects Across 100 Repeated Estimations in 4-Week Windows

$$\check{\mathbf{X}} = \begin{bmatrix} X_{111} & \cdots & 0 & X_{11t} & \cdots & 0 & X_{11T} & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & X_{N11} & 0 & \cdots & X_{N1t} & 0 & \cdots & X_{N1T} \end{bmatrix}', \tag{14}$$

where x_{i1t} is the price in market i at time t. Then, the total variance-covariance structure of the previously considered System (2), augmented with the stochasticity introduced on the market shares because of random price effects, can be shown to equal
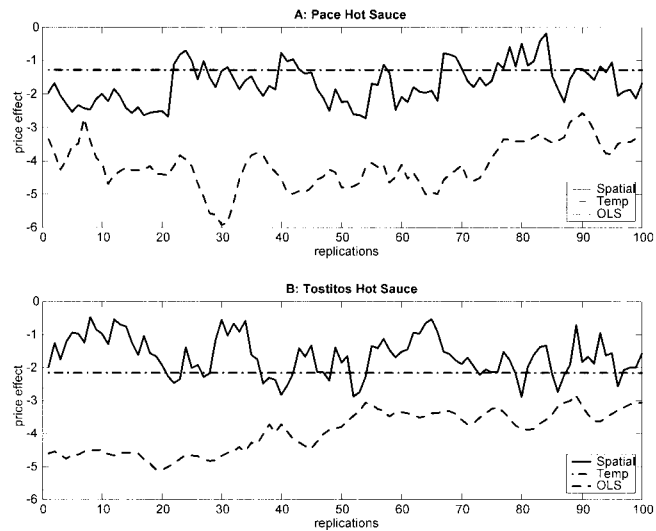
$$\Omega = \Psi \otimes \mathbf{I}_N + (\mathbf{a} \otimes \boldsymbol{\iota}_T)(\mathbf{a} \otimes \boldsymbol{\iota}_T)' \otimes \Gamma$$
$$+ \mathbf{bb}' \otimes \sigma_\beta^2 \check{\mathbf{X}}\check{\mathbf{X}}', \tag{15}$$

with $\mathbf{b} = [1\ 0\ 0\ 0]'$. This model can again be estimated through maximum likelihood.[4]

For illustration, we incorporate heterogeneity in three alternative specifications using the data for the Pace brand. The first specification is the full model containing both the cross-sectional (spatial) and temporal structure. The second specification accounts

only for temporal dependence. The third model accounts for neither and is referred to as the independent model. We ran these models for nonoverlapping moving windows of 4 weeks of data. Table 7 lists the means of selected parameters across the 25 windows of 4 weeks of data. We present the estimation results with and without the added heterogeneity estimates.

The estimation results show that the inclusion or exclusion of price heterogeneity does not meaningfully affect the mean price effect across markets. This can be seen by comparing the estimates of price response *within* model type. From the full model, we observe that the estimated standard deviation for the price effects is 1.365, which is consistent with a standard deviation across markets of 1.008 computed directly from the 64 within-market price effects.

Looking *across* model types we still find that price responses are exaggerated greatly when the cross-sectional structure of the model is unaccounted for. When we first compare the price effects of the full model with the temporal model, and then with the

---

[4]To speed up the estimation, one should use results on partitioned matrices (see, e.g., Searle 1982) to compute the determinant and the inverse of Ω from its components. Direct evaluation of these quantities becomes very time consuming when $NT(P + 1)$ becomes large.

Table 7 Estimation Results for Selected Parameters Using Specifications with Heterogeneity

	Full-Model		Temporal Model		Independent Model	
	$\sigma_{\beta price}$ > 0	$\sigma_{\beta price}$ $= 0$	$\sigma_{\beta price}$ > 0	$\sigma_{\beta price}$ $= 0$	$\sigma_{\beta price}$ > 0	$\sigma_{\beta price}$ $= 0$
λ	0.925[a]	0.837	–	–	–	–
σ_{υ}	0.282	0.292	–	–	–	–
ρ_0	0.315	0.242	0.976	0.973	–	–
σ_{ϵ}	0.086	0.093	0.110	0.120	0.326	0.483
$\sigma_{\beta price}$	1.365	–	1.476	–	8.134	–
β_{price}	−1.542	−1.736	−2.244	−2.300	−4.430	−4.113
$\beta_{feature}$	0.006	0.007	0.005	0.006	0.026	0.037
$\beta_{display}$	0.022	0.024	0.033	0.035	0.172	0.187
α_0	−1.172	−1.193	−1.241	−1.242	−1.191	−1.125

[a]Parameters are averaged over 25 successive estimations on unique moving windows of 4 weeks.

independent model, we see that the price effects become more and more biased away from zero. That is to say the main argument in the paper does not vanish when we account for price heterogeneity. Also note that in the temporal model, the autocorrelation ρ_0 in the market share data is unrealistically high, because the temporal structure attempts to substitute for the cross-sectional structure.

7. Discussion and Conclusion

We have considered the case where a marketing analyst needs to infer response estimates from cross-market (Bucklin and Gupta 1999) or cross-market by time data with insufficient time periods to calibrate models on time series only. In such cases, variation in promoted price and other promotion data across markets may reflect unobserved retailer behavior. We have developed an approach that allows researchers to test for this type of endogeneity and to take into account its effects on inference. This approach may benefit managers who need to estimate market responses from short windows of data. It may also benefit those marketing research companies—noted by Bucklin and Gupta (1999)—who use cross-sectional samples of sales velocity to infer price response.

With respect to the nature of the endogeneity, Villas-Boas and Winer (1999) and Besanko et al. (1998)

model the *temporal* effect of performance on prices and promotion while we concentrate on the *cross-sectional* effect. The two approaches are complementary and add to each other. Whereas, Villas-Boas and Winer (1999) and Besanko et al. (1998) interpret their results from the perspective of a manufacturer who sets prices on the basis of performance, our interpretation is that the actions of retailers introduce another layer of price-variation that is easiest observed cross-sectionally.

Empirically, we found evidence that suggests that unobserved components of market shares are related to the promotion variables. We demonstrated, using data experiments, that this reverse causation of market shares on pricing and promotion, even when it is weak, causes large biases in OLS estimators based on cross-sectional or panel data.

To correct for this bias, we used the fact that retail territories are spatially contiguous sets of markets, which in turn cause the unobserved retailer influence on marketing variables to manifest itself as a spatial variance component in the data. As long as the unobserved component in the data is spatial, this framework may even hold in other cases where a similar bias could emerge, such as share-dependent policies by the manufacturer (as opposed to the retailer).

To capture the spatial variance component, we introduced a set of tools to formalize space, tested for spatial dependence in the data, and used contiguity classes to model the data. To the best of our knowledge, this is the first paper in marketing to formalize the spatial dimension of multimarket data using (Voronoi) contiguity classes.

In using our framework for inference with multimarket data, we also applied it to time windows with a length of 8 weeks and even 1 week. Our experience is that the approach works well as long as there is sufficient independent variation across markets in the promotion variables after the common spatial variation is accounted for. Interestingly, even a single-equation spatial regression, i.e., the first equation in System (2), along with the assumptions about $\mathbf{e}_y$ from Equation (3), seems to work quite well with data windows of 1 week, in the sense that the estimates from cross-sections of the data are close to the estimates obtained from time series of the same data.

In terms of more general applicability of spatial analysis with other data, the aim of spatial models is often to define a ''similarity'' structure on the data. In many cases, the ''similarity structure'' is based on distances. However, this is but one definition of ''kinship'' among many. Other definitions can be suitably incorporated into a model representation of cross-sectional data by redefining the similarity matrix **W**. For instance, in survey work involving managers, one may include several questions that identify the similarity of respondents (e.g., schooling, years of experience, etc.). As another example, in the context of the PIMS database, one can model similarity on the basis of SIC codes, the presence of common parent companies across multiple business units, etc. It is also possible to use more than one component of similarity by defining multiple spatial variance components as opposed to one. In sum, the model structure discussed here can be expanded with some generality. We expect that accounting for the similarity structures when pooling cross-sectional data is likely to enhance the quality of our inferences because they help to create the ''holding everything else equal'' conditions that are required by methods of inference.

To clearly illustrate the effects of ignoring the cross-sectional dependence, we made several simplifying and sometimes limiting assumptions. These therefore deserve mentioning. First, we used aggregate data. Christen et al. (1997) show that these aggregate data can lead to biases in nonlinear models. However, in our defense we note that many of our conclusions pertain to promoted price, which varies less with a given market than, for instance, promotional dummy variables, and should therefore be less prone to such a bias. Also this argument does not explain the difference between the cross-sectional and within-market estimates that are based on data with the same level of aggregation. Second, we have simplified the dynamics of market share in response to promotion, whereas Foekens et al. (1999) model more elaborate dynamics of promotion effects. Third, we are not considering competition between brands explicitly, although we are accounting for some competitive influence by using relative marketing variables.

With respect to future research, other applications of spatial analysis can be found in the analysis of private label brands, the analysis of multimarket competition (Jayachandran et al. 1999), the spatial interpolation of local demand from sparse spatial panels, and spatial diffusion. These and other topics need a formalization of the spatial dimension. It is hoped this paper constitutes a first constructive step in this direction.

Appendix A. The Temporal Structure of the Data
First consider the temporal stochastic terms in the market share equation $\epsilon_{it} = \rho_0 \cdot \epsilon_{it-1} + \xi_t$. Assuming $|\rho_0| < 1$, the standard result holds that $E(\epsilon_{it}\epsilon_{i't-s}) = \sigma_\xi^2 \cdot \rho_0^s/(1 - \rho_0^2)$ if $i = i'$ and 0 else.

Next, we derive the temporal structure of the stochastic terms $\delta_{ijt} = \rho_j \cdot \delta_{ijt-1} + \theta_j\tau_{it} + \nu_{ijt}$ in the equations describing the marketing variables. Through backward substitution, we can write $\delta_{ijt} = \Sigma_{s=0}^{\infty} \rho_j^s \cdot (\theta_j\tau_{it-s} + \nu_{ijt-s})$. From this form it is standard to compute the cross-structure of the stochastic terms. It can be shown that

$$E(\delta_{ijt}\delta_{i'j't'}) = \begin{cases} (\theta_j^2\sigma_\tau^2 + \sigma_{vj}^2)\dfrac{\rho_j^{|t-t'|}}{1-\rho_j^2} & \text{if } i = i', \quad j = j' \quad \text{and} \quad \forall\, t, t' \\[2ex] \theta_j\theta_{j'}\sigma_\tau^2\dfrac{\rho_j^{t-t'}}{1-\rho_j\rho_{j'}} & \text{if } i = i', \quad j \neq j' \quad \text{and} \quad \forall\, t \geq t' \\[2ex] 0 & \text{if } i \neq i', \quad \text{and} \quad \forall\, j, j', t, t'. \end{cases}$$

$$(A1)$$

Note the asymmetry in the second case, which originates from the fact that the error terms have different autoregressive parameters. Combining these results, gives the variance-covariance matrix in the text.

Appendix B. The Log-Likelihood Function
The likelihood function for the sample of observations depends on $5 + 6P$ parameters is equal to $(2\pi)^{-NT(P+1)/2} \cdot |\mathbf{\Omega}|^{-1/2} \cdot \exp(-\frac{1}{2}e'\mathbf{\Omega}^{-1}e)$, where e and $\mathbf{\Omega}$ are defined in the text. For computational convenience, arrange the cross-sections of error terms to a matrix $V_{N\times T(P+1)}$ defined by $V = [e_{y1} \cdots e_{yT}\, e_{x11} \cdots e_{x1T} \cdots e_{xP1} \cdots e_{xPT}]$.

Magnus (1982) developed two results for a family of matrices to which the structure of Ω belongs. These results can be applied to the covariance structure at hand. The first result concerns the determinant of Ω, which can be shown to equal

$$|\Omega| = |\Psi|^N \cdot |\mathbf{I}_N + C \cdot \Gamma|, \tag{B1}$$

where $C = (\mathbf{a} \otimes \boldsymbol{\iota}_T)' \Psi^{-1} (\mathbf{a} \otimes \boldsymbol{\iota}_T)$. The second result concerns the inverse of the matrix Ω. After some algebraic manipulations on Magnus's results, this inverse is equal to

$$\Omega^{-1} = \Psi^{-1} \otimes \mathbf{I}_N$$

$$- \Psi^{-1}(\mathbf{a} \otimes \boldsymbol{\iota}_T)(\mathbf{a} \otimes \boldsymbol{\iota}_T)'\Psi^{-1} \otimes \frac{1}{C}(\mathbf{I}_N - (\mathbf{I}_N + C \cdot \Gamma)^{-1})$$

$$\equiv \Psi^{-1} \otimes \mathbf{I}_N - \Xi \otimes \mathbf{Z}. \tag{B2}$$

Taking the log of the likelihood function gives (up to an irrelevant constant)

$$\ell\ell(\mathbf{y} \,|\, \lambda, \boldsymbol{\rho}, \boldsymbol{\theta}, \sigma_\epsilon, \sigma_v, \boldsymbol{\sigma}_v, \boldsymbol{\beta}, \boldsymbol{\alpha}) = -0.5 \cdot \log|\Omega| - 0.5 \cdot \mathbf{e}'\Omega^{-1}\mathbf{e}. \tag{B3}$$

This function can be simplified using the following result on traces from matrix algebra (assuming the right-hand side exists):

$$\text{vec}(\mathbf{A})' \cdot (\mathbf{B} \otimes \mathbf{C}) \cdot \text{vec}(\mathbf{D}) = \text{trace } \mathbf{D} \cdot \mathbf{B}' \cdot \mathbf{A}' \cdot \mathbf{C}. \tag{B4}$$

Now, substituting (B1) and (B2) into (B3) and using (B4), the following log-likelihood function is obtained (ignoring an irrelevant constant):

$$\ell\ell(\mathbf{y} \,|\, \lambda, \boldsymbol{\rho}, \boldsymbol{\theta}, \sigma_\epsilon, \sigma_v, \boldsymbol{\sigma}_v, \boldsymbol{\beta}, \boldsymbol{\alpha})$$

$$= -0.5 \cdot N \cdot \log(|\Psi|) - 0.5 \cdot \log|\mathbf{I}_N + C \cdot \Gamma| - 0.5 \cdot \text{trace}(\mathbf{V} \cdot \Psi^{-1'} \cdot \mathbf{V}')$$

$$+ 0.5 \cdot \text{trace}(\mathbf{V} \cdot \Xi' \cdot \mathbf{V}' \cdot \mathbf{Z}). \tag{B5}$$

This log-likelihood uses inversions and determinants for matrices of size $\max\{T(P + 1), N\}$ only.

References

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, Boston.

Arthur, W. Brian. 1994. *Increasing Returns and Path Dependence in the Economy.* Michigan University Press, Ann Arbor, MI.

Besanko, David, Sachin Gupta, Dipak Jain. 1998. Logit demand estimation under competitive pricing behavior: an equilibrium framework. *Management Sci.* **44**(11) 1533–1547.

Bound, John, David A. Jaeger, Regina Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* **90** (June) 443–450.

Bucklin, Randolph E., Sunil Gupta. 1999. Commercial use of UPC scanner data: industry and academic perspectives. *Marketing Sci.* **18**(3) 247–273.

Case, Anne C. 1991. Spatial patterns in household demand. *Econometrica* **59**(4) 953–965.

Chamberlain, Gary, Zvi Griliches. 1975. Unobservables with a var-

iance-components structure: ability, schooling, and the economic success of brothers. *Internat. Econom. Rev.* **16**(2) 422–449.

Christen, Markus, Sachin Gupta, John C. Porter, Richard Staelin, Dick R. Wittink. 1997. Using market-level data to understand promotion effects in a nonlinear model. *J. Marketing Res.* **34** (August) 322–334.

Elrod, Terry. 1988. Choice map. Inferring a product-market map from panel data. *Marketing Sci.* **7** (Winter) 21–40.

Erdem, Tülin. 1996. A dynamic analysis of market structure based on panel data. *Marketing Sci.* **15**(4) 359–378.

——, Russell Winer. 1999. Econometric modeling of competition: a multi-category choice-based mapping approach. *J. Econometrics* **89** 159–175.

Foekens, Eijte W., Peter S.H. Leeflang, Dick R. Wittink. 1999. Varying parameter models to accommodate dynamic promotion effects. *J. Econometrics* **89**(1/2) 249–268.

Greenberg, Edward, Charles E. Webster. 1991. *Advanced Econometrics: A Bridge to the Literature.* Krieger Publishing Company, Malabar, FL.

Hausman, J. 1996. Valuation of new goods under perfect and imperfect competition. T. Bresnahan, R. Gordon, eds. *The Economics of New Goods, Studies in Income and Wealth,* vol. 58, National Bureau for Economic Research, Cambridge, MA.

Hsiao, Cheng. 1986. *Analysis of Panel Data.* Cambridge University Press, New York, NY.

Jayachandran, Satish, Javier Gimeno, P. Rajan Varadarajan. 1999. The theory of multi-market competition: A synthesis and implications for marketing strategy. *J. Marketing* **63**(3) 49–66.

Judge, George C., W.E. Griffiths, R. Carter Hill, Helmut Lütkepohl, Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics.* John Wiley and Sons, New York, NY.

Kamakura, Wagner A., Rajendra K. Srivastava. 1984. Predicting choice shares under conditions of brand interdependence. *J. Marketing Res.* **21** 420–434.

LeSage, James P. 1998. *Spatial Econometrics.* Book-manuscript available online from http://www.econ.utoledo.edu/faculty/lesage/lesage.html.

Magnus, Jan. 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *J. Econometrics.* **19** 239–285.

Nevo, Aviv. 2000. Measuring market power in the ready-to-eat cereal industry. *Econometrica.* Forthcoming.

Niu, Xufeng, George C. Tiao. 1995. Modeling satellite ozone data. *J. Amer. Statist. Assoc.* **90** (September) 969–983.

Searle, Shayle R. 1982. *Matrix Algebra Useful for Statistics.* John Wiley and Sons, New York, NY.

Staiger, Douglas, James H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65** (May) 557–586.

Villas-Boas, J. Miguel, Russell S. Winer. 1999. Endogeneity in brand choice models. *Management Sci.* **45**(10) (October) 1324–1338.