Empirical Economics (1994) 19:451–472

EMPIRICAL ECONOMICS

Estimating the Geographic Diffusion of the Videocassette Recorder Market

YU-MIN CHEN[1]

Planning & Research Department, JC Penney Co., Inc. P.O. Box 10001, Dallas, TX 75301, USA

HIROKAZU TAKADA[1]

Department of Marketing, Box 508, The School of Business and Public Administration, Baruch College, The City University of New York, 17 Lexington Avenue, New York NY 10010, USA

Abstract: The majority of studies concerning diffusion or product growth of consumer durables have treated the U.S. market as a whole and have applied the diffusion model on the assumption that the market exhibits a homogeneous response in its diffusion process. If the market is heterogeneous, however, an aggregate model entails a misspecification problem which could adversely affect the applicability and efficiency of the model. A modeling framework is developed for analyzing the diffusion process in a possibly heterogeneous market. Empirical analysis using data on the videocassette recorder (VCR) market reveals that the modeling framework captures to a fair extent heterogeneous diffusion processes across different regions in the U.S. market. Managerial implications are derived and discussed.

JEL Classification System-Numbers: C10, C53, M31

1 Introduction

Analysis of the diffusion process or product growth of consumer durable goods has been a major concern for marketing researchers and practitioners. Voluminous research results have been presented as indicated by the extensive review of the literature on diffusion studies (Mahajan, Muller, and Bass 1990). Since Bass (1969) first proposed an econometric model to explain the diffusion process, the model has been applied to a wide variety of industries, adequately capturing the diffusion processes of durable goods and predicting their sales and product life cycles accurately (see Mahajan, Muller, and Bass 1990).

Most studies to date using the Bass model have, however, treated the U.S. market as a whole (see Mahajan and Wind 1986 for references therein). The premise in using an aggregate model is that the regional markets under study are homogeneous in the sense that the same diffusion process is applicable

[1] We greatly benefited from Professor Baldev Raj, Editor, and two anonymous reviewers. We thank Professor Aman Ullah, University of California, Riverside, for his valuable suggestions for the manuscript, and Mary M. Long, Ph.D. candidate, Baruch College, for her editorial assistance. Authors are listed in alphabetical order; each contributed equally to this research.

0377–7332/94/3/451–472 $2.50

to every region. It is possible, however, that the diffusion process emerges differently in each area. There are several reasons for postulating differences in the diffusion process across regions, including varying product availability, differing price levels and local advertising competition, strength of the sales force, and socioeconomic and cultural differences among regions (Moriarty 1975; Wittink 1977). It is important for a firm to have a thorough understanding of these regional characteristics in order to tailor its marketing strategy (e.g., advertising campaign, pricing, distribution strategy, and sales force allocation) in response to regional needs and wants. The coefficients provided by the Bass model are valuable sources of information for analyzing these regional variations in the diffusion process.

One of the approaches employed to capture the diffusion process across regions is to treat the process as a spatial diffusion. This type of analysis is often found in studies by geographers, for example, Brown, Malecki and Spector (1976) and Hagerstrand (1965). The former study employed two approaches: the communication approach based on an identification of the neighborhood effect (Brown 1981; Gore and Lavaraj 1987; Hagerstrand 1967), and the market and infrastructure approach stressing the importance of the establishment of diffusion agencies such as retail outlets in order to enhance the availability of the innovation to potential adopters. These conceptual models are helpful in understanding how a diffusion process evolves and what kind of strategy can stimulate the process within a homogeneous market. However, these models do not explicitly address issues related to diffusion across heterogeneous regions.

Mahajan and Peterson (1978) addressed diffusion from a joint space and time perspective and developed a time-space substitution model. They extended the Bass model by assuming that the neighborhood effect diminishes with increased distance from the market of innovation origination, thus decreasing the size of market potential across new markets. Although the model is capable of capturing a new product rollout, it assumes that the diffusion process emanates from one region, termed the innovation region, and then spreads to the rest of the regions. As argued in this paper though, there is a need to extend this model to situations where substitution or diffusion emanates from more than one region since consumer durables are frequently introduced in different areas simultaneously or with only negligible time lags due to the efficient distribution and communication networks available in the modern market.

The major objective of this study is to propose a modeling framework for analyzing the diffusion of consumer durables in different geographic regions. This framework will be developed based on the Bass diffusion model. An empirical analysis is conducted to illustrate the application of the proposed framework using actual data from the videocassette recorder (VCR) market. We then estimate the diffusion process or product growth of the VCR in the U.S. market by geographic region.

VCR's have been gaining popularity in households. Their sales have been steadily increasing since they were first introduced in the U.S. market in 1976. The functions of VCR's have expanded from taping TV programs for viewing at a later convenient time to taping TV programs for the purpose of making a personal collection of videos and to watching movies on rental videocassettes. Recent widespread use of video cameras or camcoders for home use has stimulated further adoption of VCR's by households. Thus, a study on the diffusion of VCR's is useful for manufacturers and dealers in order to formulate their marketing strategies in different geographic markets.

In order to determine whether or not the VCR diffusion process is homogeneous across regions, one may collect sales data of VCR's in each region for analysis. This argument is based on the premise that the data contain the relevant information about the underlying diffusion process. Specifically, we model heterogeneity across regions as a general case and homogeneity as a special case. We then analyze the data to test the hypothesis: Are regression parameters equal across regions or are they varying? If the hypothesis of being equal is accepted, the diffusion process of VCR's across regions is homogeneous, and an aggregate model will adequately represent the structure. Otherwise, the diffusion process is heterogeneous and raises the following questions: "To what extent does the diffusion process vary across regions?" and "What is an appropriate analytical method?" If the diffusion process is completely heterogeneous, it must be analyzed and modeled separately in order to understand the diffusion mechanism in each region. One of the drawbacks of this approach is that it fails to explain heterogeneous variation across regions as a whole. Furthermore, in practice, we hardly know *a priori* to what extent the diffusion process is heterogeneous across regions. This calls for a more comprehensive modeling framework providing a meaningful and accurate representation of varying diffusion processes.

The major contribution of this study is that the proposed framework explicitly accounts for heterogeneity across different submarkets, i.e., geographic regions, of a product market. Failure to account for heterogeneity contaminates the diffusion parameter estimates and accordingly may lead to ineffective new product marketing strategy. Previous research has primarily focused on the analysis at an aggregate level assuming that all regions of a product market are homogeneous, and looked at the diffusion of innovations solely as a function of time. This offers a limited perspective on the actual underlying process that is taking place since it fails to include other dimensions such as space, socio-economic strata, user groups and national boundaries that may also effect the diffusion process. It is clear that by analyzing different geographical markets separately, one can determine the attractiveness of each market and then formulate an appropriate product introduction strategy.[2]

This paper is organized as follows: A brief discussion of the Bass diffusion model is presented in the next section; subsequent sections describe the data, empirical analysis, estimation results and managerial implications.

---

[2] The authors wish to thank one of the anonymous referees for the comments on spatial diffusion of innovation.

2 Bass Model

Let T be the random time that a purchase occurs. The probability that an initial purchase will be made at time $T = t$ given that no purchase has been made before t, can be written as:

$$\Pr(t) = f(t)/[1 - F(t)] \tag{1}$$

where $f(\bullet)$ and $F(\bullet)$ are the probability density function and distribution function, respectively.

Bass (1969) presented the basic assumption that the probability that an initial purhcase will be made at time t given that no purchase has yet been made is a linear function of the number of previous buyers. He derived the likelihood of purchase at time $T = t$ as follows:

$$\Pr(t) = p + qF(t) \tag{2}$$

where p and q are defined as the coefficients of innovation and imitation, respectively. Although Bass did not specify it in his work, implicit in equation (2) is that $0 \leq \Pr(t) \leq 1$ to be logically consistent.[3] From equations (1) and (2):

$$f(t) = [p + qF(t)][1 - F(t)] \ . \tag{3}$$

The accumulated sales up to time t is

$$Y(t) = mF(t) \tag{4}$$

where m is the total number of purchases during the period for which the density function was constructed. Therefore, the sales at time t, $S(t)$, is

$$S(t) = mf(t) \ . \tag{5}$$

Substituting (4) and (5) into (3), we have

$$S(t) = pm + (q - p)Y(t) - \frac{q}{m}Y(t)^2 \ . \tag{6}$$

Equation (6) is the diffusion equation of a given consumer durable. The time at which the sales rate reaches its peak, t^*, can be found by differentiating S and is given by

$$t^* = 1/(p + q)\ln(q/p) \ . \tag{7}$$

If $q > p$, then the peak sales at t^*, $S(t^*)$, is given by

[3] An anonymous referee pointed out these restrictions to make equation (2) logically consistent. The referee also raised such issues as translation of these restrictions into restrictions on p and q and implications of these restrictions on estimation of equation (13). These issues are related to the Bass model itself, and Bass did not address them in his paper. Analysis of these issues is beyond the scope of this study, and much further work is needed to thoroughly investigate them in future studies.

$$S(t^*) = m(p+q)^2/4q \ . \tag{8}$$

The unknown parameters to be estimated with data are p, q, and m. Since the equation consists of nonlinear parameters and data are discrete time series, Bass (1969) has suggested the following alternative by rewriting equation (6) as the discrete analog:

$$S(t) = a + bY(t-1) + c[Y(t-1)]^2 \ . \tag{9}$$

Equation (9) is linear in parameters a, b, and c. The linear regression technique can be applied to obtain the estimates of a, b, and c. The estimates of the structural parameters, p, q, and m, can be obtained through the following relationships:

$$p = a/m \ , \tag{10}$$

$$q = -mc \ , \tag{11}$$

and,

$$m = (-b - \sqrt{b^2 - 4ac})/2c \ . \tag{12}$$

The basic model can be extended to the cross-sectional and time series model. The subsequent model for region i is:

$$S_i(t) = a_i + b_i Y_i(t-1) + c_i[Y_i(t-1)]^2 + V_{it} \ . \tag{13}$$

$$i = 1, \ldots, n \ , \quad \text{and} \quad t = 1, \ldots, T \ .$$

The subscript i is introduced to represent the ith region. V_{it} is the disturbance term to specify the Bass model of equation (9) as a regression equation for estimation.[4] If the regions exhibit a homogeneous response in the diffusion process, the subscript "i" for the parameters would not be necessary. Homogeneity is reflected by setting $a_i = a$, $b_i = b$, $c_i = c$, and $V_{it} = V_t$, all $i = 1, \ldots, n$. Equation (13) is then rewritten as:

$$S_i(t) = a + bY_i(t-1) + c[Y_i(t-1)]^2 + V_t \ . \tag{14}$$

$$i = 1, \ldots, n \ , \quad \text{and} \quad t = 1, \ldots, T \ .$$

Regional time series data can be pooled to estimate parameters. We define $S_i = [S_i(1), \ldots, S_i(T)]'$, $Y_i = [Y_i(0), \ldots, Y_i(T-1)]'$, $Y_i^2 = [Y_i^2(0), \ldots, Y_i^2(T-1)]'$, $V_i = [V_{i1}, \ldots, V_{iT}]'$, $Z_i = [l, Y_i, Y_i^2]$, a $T \times 3$ matrix with l being a $T \times 1$ vector consisting entirely of 1, i.e., $l = [1, \ldots, 1]'$, and $\beta_i = [a_i, b_i, c_i]'$. Equation (13) can then be written in vector-matrix notation as:

$$S_i = Z_i \beta_i + V_i \ , \qquad i = 1, \ldots, n \ . \tag{15}$$

It is assumed that $EV_i = 0$, and $EV_iV_j = \sigma_{ii}I$, $i = j$; $= 0$, otherwise.

[4] V_{it} is not added to equation (9), which could contradict equation (13). It should be interpreted as a disturbance term of the estimation equation.

3 Data

The data analyzed in this study were obtained from *Electrical Merchandising* and the annual *Electronic Market Data Handbook* of the Electronic Industry Association (EIA) which started reporting regional breakdowns in 1981. The data from 1981 through 1987 are based on Bureau of Census geographical regions where the United States is divided into four major quadrants, and further subdivided into nine regions. In this study, we employ data on the nine regions rather than on the four quadrants so that geographic variation within each of the regions is smaller, and consequently the diffusion process in each region will appear more homogeneous. Although the data at the SMSA or state level would be desirable, they are not published for this product class. It is worth noting that four out of the nine regions account for almost 70% of the market. The order of these regions in terms of size from largest to smallest is: the Mid-Atlantic region which includes the New York City market, the Pacific region which includes the Southern California market, the East North Central region which includes Chicago, and the South Atlantic region which encompasses the southern Atlantic seaboard. The data consist of sales to dealers and distributors of VCR's compiled by the Electronic Industries Association, including sales of both Beta and VHS formats. Dealers cover a wide variety of retail establishments such as appliance dealers, department stores, video specialty stores, discount stores, catalog showrooms, and advanced consumer electronics specialist departments in traditional stores.

4 Model Specification and Estimation

There are two divergent and commonly used ways of specifying a model with time series data across regions: Estimating each region separately, or using an aggregate model. The former assumes that the diffusion processes of VCR's across regions are independent; the latter assumes that the processes are identical. These two cases are the extreme cases. However, these are by no means the only two specifications. It is possible that the diffusion processes are not identical but correlated with each other. The seemingly unrelated regression (SUR) approach proposed by Zellner (1962) is suitable for the situation involving some omitted variables that are common to all regions. The randomly varying coefficient model, in which the coefficients are decomposed into systematic and random components (Raj and Ullah 1981), is another approach to model correlated diffusion processes. As stated previously, we contend that the form of specification is testable, and a suitable specification is the one which is supported by the data. Therefore, the model will be specified in alternative forms, and statistical hypotheses tests will be conducted to determine an appropriate specification.

Table 1 summarizes estimation results of the OLS procedure applied to each region separately. The fit of the model to the data appears to be good as indicated by the extremely high adjusted R^2 values, all of which are over 0.98, and the proper expected signs for all coefficients. The parameter values of m, p, and q are derived by applying the formulas in (10)–(12) to the OLS estimates. The estimation result of the aggregate model of equation (14) is also reported in the table. It is worth noting that the averages of the nine OLS estimates of the coefficients p and q appear to show values close to the corresponding estimates of the aggregate model. This result is plausible because the aggregate model is nothing more than an averaging process.

Table 2 presents estimation results of the SUR estimation. The estimates were computed based on a smooth improved SUR estimator presented by Ullah and Racine (1992). It can be seen that the SUR estimates are close to the separate OLS estimates.

Because the Bass model has nonlinear parameters, Srinivasan and Mason (1986) and Jain and Rao (1990) proposed nonlinear least squared (NLS) estima-

Table 1. OLS estimation results

			OLS Estimates			Derived Parameters		
SIC Regions	F	R^2	a	b	c	p	q	m
New England	1304	0.998	18.08 (9.38)	1.15 (0.04)	−0.0004 (0.00003)	0.0041	0.7671	4371
Mid Atlantic	328	0.991	71.74 (57.91)	1.04 (0.07)	−0.0001 (0.00001)	0.0051	0.7136	13983
South Atlantic	460	0.994	47.73 (38.82)	1.10 (0.06)	−0.0002 (0.00002)	0.0045	0.7420	10628
East North Central	287	0.990	33.12 (55.85)	1.15 (0.08)	−0.0002 (0.00002)	0.0028	0.7631	11927
East South Central	1161	0.997	21.07 (7.35)	1.04 (0.04)	−0.0005 (0.00003)	0.0062	0.7160	3376
West North Central	171	0.982	2.85 (26.10)	1.22 (0.11)	−0.0005 (0.00006)	0.0006	0.7944	4148
West South Central	350	0.992	39.11 (29.23)	1.07 (0.07)	−0.0002 (0.00003)	0.0054	0.7271	7180
Mountain	368	0.992	19.44 (12.01)	1.01 (0.06)	−0.0005 (0.00005)	0.0063	0.6981	3106
Pacific	422	0.993	116.43 (43.02)	0.90 (0.05)	−0.0001 (0.00001)	0.0092	0.6481	12650
Mean of nine regions			369.57*	1.08	−0.0003	0.0049	0.7299	71369*
Aggregate model	455	0.993	365.89 (253.79)	1.05 (0.06)	−0.00002 (0.000002)	0.0051	0.7203	71474

Note: Numbers in parentheses represent standard errors.

* The number indicates a sum of values in nine regions for a comparison to an aggregate model.

tion procedures recognizing potential shortcomings associated with the OLS estimation procedure. The OLS estimation, for example, does not provide standard errors for the estimates p, q, and m since they are nonlinear functions of parameters of the discrete analog model. A time-interval bias is also evident since discrete time series data are used for estimating a continuous-time model. Nonetheless, the OLS estimation procedure has been widely used because of its easy implementation and, more importantly, it allows researchers to expand the basic model with a certain flexibility. While choice of the nonlinear estimation procedures is an empirical issue (Mahajan et al, 1986), Takada (1989) showed that the Jain and Rao algorithm provided better fit for the data studied than the Srinivasan and Mason algorithm. Thus, we employ the NLS estimation based on the Jain and Rao algorithm. The estimation results are summarized in Table 3. The OLS and SUR estimates are also included in Table 3 for the purpose of comparison with the NLS estimates. Out of 57 OLS and SUR parameter estimates, 52 parameter estimates are within asymptotic 95% confidence interval of the NLS estimates. Two OLS estimates and three SUR estimates lie slightly outside the confidence interval. We thus conclude that the OLS and SUR estimates are reasonably close to the NLS estimates. A closer look at the parameter estimates reveals that the NLS estimates of the q coefficient are larger across all the regions and the United States as a whole than the OLS and SUR estimates,

Table 2. SUR estimation results

	SUR Estimates			Derived Parameters		
SIC Regions	a	b	c	p	q	m
New England	18.83 (5.99)	1.15 (0.019)	−0.0004 (0.000013)	0.0043	0.7648	4381
Mid Atlantic	72.15 (36.19)	1.04 (0.033)	−0.0001 (0.000006)	0.0052	0.7135	13922
South Atlantic	52.72 (24.69)	1.01 (0.028)	−0.0002 (0.000007)	0.0055	0.6993	9653
East North Central	37.98 (35.81)	1.14 (0.038)	−0.0001 (0.000008)	0.0032	0.7591	11919
East South Central	22.04 (35.81)	1.04 (0.038)	−0.0005 (0.000008)	0.0065	0.7131	3383
West North Central	5.86 (16.58)	1.20 (0.045)	−0.0004 (0.000026)	0.0014	0.7873	4168
West South Central	42.93 (18.48)	1.05 (0.030)	−0.0002 (0.000011)	0.0059	0.7218	7218
Mountain	20.31 (7.594)	1.00 (0.0274)	−0.0005 (0.000022)	0.0065	0.6960	3107
Pacific	117.51 (26.46)	0.90 (0.021)	−0.0001 (0.000004)	0.0093	0.6482	12590

Note: Numbers in parentheses represent standard errors.

Table 3. Parameter estimates of the three estimation procedures

	Parameter Estimates								
	p			q			m		
SIC Regions	*OLS*	*NLS*+	*SUR*	*OLS*	*NLS*	*SUR*	*OLS*	*NLS*	*SUR*
New England	0.0041	0.0031	0.0043	0.7671	0.8592	0.7648	4371	3979	4381
Mid Atlantic	0.0051	0.0044	0.0052	0.7136	0.8169	0.7135	13983	12423	13922
South Atlantic	0.0045	0.0040	0.0055	0.7420	0.8467	0.6993*	10628	9402	9653
East North Central	0.0028	0.0034	0.0032	0.7631	0.8913	0.7591	11927	10316	11919
East South Central	0.0062*	0.0045	0.0065*	0.7160	0.7947	0.7131	3376	3093	3383
West North Central	0.0006	0.0027	0.0014	0.7944	0.9566	0.7873	4148*	3474	4168*
West South Central	0.0054	0.0044	0.0059	0.7271	0.8253	0.7218	7180	6384	7218
Mountain	0.0063	0.0052	0.0065	0.6981	0.8002	0.6960	3106	2742	3107
Pacific	0.0092	0.0071	0.0093	0.6481	0.7358	0.6482	12650	11311	12590
Aggregate	0.0051	0.0044	n/a	0.7203	0.8247	n/a	71474	63265	n/a

+ All of the NLS parameter estimates (p, q, and m) are statistically significant at the 0.01 level.

* The OLS and SUR parameter estimates lie outside the asymptotic 95% confidence interval of the NLS parameter estimates.

whereas the NLS estimates of the m coefficient are consistently smaller than the OLS and SUR estimates. On the other hand, the p estimates do not exhibit any clear pattern across the three different estimates.

The SUR model can be viewed as a model where all regions have different coefficient vectors which are regarded as fixed. The random coefficient regression (RCR) model relaxes this restrictive assumption. We develop specifications based on the RCR model in the next section.

4.1 Proposed Specification

Carefully examining the Bass model of equation (6), one will find that marketing mix variables, such as advertising, price, and promotion activities are not explicitly incorporated in the model. This does not imply that marketing activities have a negligible impact on sales. The sales response models capture varying degrees of impact of marketing mix on firms' sales and profits (Kotler 1988). Their effectiveness is dependent upon regional, socioeconomic and cultural characteristics (Moriarty 1975; Wittink 1977).

Considering the effects of regional differences, we model

$$\beta_i = \beta + \delta_i \tag{16}$$

where β is a regional invariant parameter and δ_i represents region-specific factors unobservable in the Bass model. It is assumed that δ_i's are i.i.d. random variables with $E\delta_i = 0$ and the variance-covariance matrix Δ. Consequently, β_i is stochastic and varies randomly across regions with $E\beta_i = \beta$ and the variance-covariance matrix Δ (that is, $E(\beta_i - \beta)(\beta_i - \beta)' = \Delta$). Substituting equation (16) into (15), we have

$$S_i = Z_i\beta + Z_i\delta_i + V_i\ , \qquad i = 1, \ldots, n\ . \tag{17}$$

It is assumed that δ_i and V_i are independent. Equation (16) is in the framework of RCR models.

The premise of applying the RCR model is that the effects (regions in the current study) are drawn randomly from a large population of effects. While it might be reasonable to assume random coefficients in a time series context when the data reflect a random sample of all time periods, this argument will only work in a cross-sectional context if we have a sample from a large population of possible parameters. Our study does not have this condition because we include all the regions, covering the entire U.S. market. The randomness in our model, however, emanates from the unobservable variables not being modeled in the equation such as regional variations in marketing activities, socioeconomic and cultural characteristics. This interpretation was viewed by Maddala (1977) as a more appropriate approach for econometric models.[5]

5 Hypothesis Testing

The test of homogeneity can be represented by the hypotheses:

$$H_0\colon \Delta = 0 \ , \quad \text{given } E\beta_i = \beta \ , \qquad i = 1, \ldots, n \ ,$$

versus

$$H_1\colon \Delta \neq 0 \ , \quad \text{given } E\beta_i = \beta \ , \qquad i = 1, \ldots, n \ .$$

If the null hypothesis is accepted, there are no statistically significant effects which will cause varying diffusion processes across regions. In other words, β_i parameters are nonstochastic and equal to a common value β; that is, the regions exhibit a homogeneous response to the diffusion process. Otherwise, it may be concluded that the regions are heterogeneous.

Swamy (1971) has proposed an estimator of Δ, denoted as $\hat{\Delta}$, as:

$$\hat{\Delta} = \frac{R}{n-1} - \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_{ii}(Z_i'Z_i)^{-1} \tag{18}$$

where n is the number of regions. R and $\hat{\sigma}_{ii}$ are defined as:

$$R = \sum_{i=1}^{n} \hat{\beta}_i \hat{\beta}_i' - \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i \sum_{i=1}^{n} \hat{\beta}_i' \tag{19}$$

and

$$\hat{\sigma}_{ii} = \frac{\hat{u}_i'\hat{u}_i}{T-k} \tag{20}$$

where $\hat{\beta}_i$ is the OLS estimate of β_i in equation (15), $\hat{u}_i$ is the OLS residual, and $T-k$ are the degrees of freedom in the regression equation.

We have conducted hypothesis testing with the data:

$$H_0\colon \Delta = 0 \ , \quad \text{given } E\beta_i = \beta \ , \qquad i = 1, \ldots, 9 \ .$$

The variance-covariance matrix, $\hat{\Delta}$ is estimated to be

$$\hat{\Delta} = \begin{bmatrix} -91.860 & -0.892 & 0.004 \\ -0.892 & 0.004 & -0.000003 \\ 0.004 & -0.000003 & 0.00000003 \end{bmatrix} . \tag{21}$$

[5] We have benefited from the Editor's comments (Professor Raj) on the issue of the proper context for the use of the random coefficient model.

It appears that the first diagonal term of the matrix which corresponds to the variance of the parameter a_i is estimated to be negative. As a variance must assume a nonnegative value, the negative value implies that the proposed specification may not be appropriate. Scheffé (1959) has suggested a way to circumvent this problem by setting the variance equal to zero whenever a negative estimate is obtained.[6] Following this suggestion, we infer that the first parameter a_i is not a random variable. The variances of b_i and c_i are nonnegative as indicated by the second and third diagonal elements of the matrix. We thus assume that the a_i's are nonstochastic, i.e., they are 'fixed-effect' parameters of regions, and that $[b_i, c_i]' = \alpha_i$ varies randomly across regions with mean $E\alpha_i = \alpha = [\bar{b}, \bar{c}]'$ and the variance-covariance matrix Δ_2. Based on this assumption, we write

$$\alpha_i = \alpha + \gamma_i \tag{22}$$

where $E\gamma_i = 0$ and $E\gamma_i\gamma_i' = \Delta_2$, $i = j$; $= 0$, otherwise. The revised and updated model can be written as:

$$S_i = la_i + X_i\alpha + X_i\gamma_i + V_i \qquad i = 1, \ldots, 9 \tag{23}$$

where $X_i = [Y_i, Y_i^2]$ is a 7×2 matrix, and l is a 7×1 vector consisting entirely of 1. It is assumed that γ_i and V_i are independent. Swamy (1971) has proposed an estimator of Δ_2, denoted by $\hat{\Delta}_2$, as:

$$\hat{\Delta}_2 = \frac{R_2}{n-1} - \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{ii}(X_i'MX_i)^{-1} \tag{24}$$

where n is the number of regions. R_2 is defined as:

$$R_2 = \sum_{i=1}^{n} \hat{\alpha}_i\hat{\alpha}_i' - \frac{1}{n}\sum_{i=1}^{n} \hat{\alpha}_i \sum_{i=1}^{n} \hat{\alpha}_i' \tag{25}$$

where

$$\hat{\alpha}_i = (X_i'MX_i)^{-1}X_i'MS_i \; . \tag{26}$$

$\hat{\sigma}_{ii}$ and M are defined as:

$$\hat{\sigma}_{ii} = \frac{\hat{u}_i'\hat{u}_i}{T-k} \; , \tag{27}$$

and

$$M = I_T - l(l'l)^{-1}l' \tag{28}$$

where $\hat{u}_i$ is the OLS residual for region i, I_T is a $T \times T$ identity matrix, and $T - k$ are the degrees of freedom of the regression equation.

The estimation result of the variance-covariance matrix, $\hat{\Delta}_2$, is:

---

6 An anonymous referee points out an alternative way to approximate $\tilde{\Delta}$ (see Rao 1973, p. 63).

$$\hat{\Delta}_2 = \begin{bmatrix} 0.0039 & -0.000003 \\ -0.000003 & 0.00000003 \end{bmatrix} \tag{29}$$

which indicates that the diagonal terms are now positive. We now test the hypothesis that the a_i's are equal across regions and that $[b_i, c_i]' = \alpha_i$ is not affected by unobserved random factors. The alternative hypothesis is that the a_i's are different and that the α_i's are random. The hypothesis is thus stated as:

$$H_0\colon \Delta_2 = 0 \ , \quad \text{and} \quad a_1 = a_2 = \cdots = a_9 = \bar{a} \ , \quad \text{given } E\alpha_i = \alpha \ , \quad \text{for all } i \ ,$$

versus

$$H_1\colon \Delta_2 \neq 0 \ , \quad \text{and} \quad a_i\text{'s are not equal, given } E\alpha_i = \alpha \ , \quad \text{for all } i \ .$$

A likelihood ratio test proposed by Swamy (1971) shows that under the null hypothesis

$$-2 \log \lambda = T \sum_{i=1}^{n} \log \hat{S}_{ii} - (T - k_2) \sum_{i=1}^{n} \log \hat{\sigma}_{ii} - \sum_{i=1}^{n} \log|X_i'X_i| - n \log|n^{-1}R_2| \tag{30}$$

is asymptotically χ^2 distributed with $\{\frac{1}{2}k_2(k_2 + 1) + k_1(n - 1)\}$ degrees of freedom. Here, $\hat{S}_{ii}$ is the estimated variance of regression for region i under the null hypothesis, k_2 is the dimension of the vector of parameters α_i, and k_1 is the number of parameters which are specified as non-random in each equation. For the current data, equation (30) is computed to be 129.91, which is greater than the critical value of a χ^2 distribution with 11 degrees of freedom under a 5% significance level. This leads us to reject the null hypothesis. Therefore, it is an appropriate specification that a_i's are nonstochastic but different for the nine regions and that $[b_i, c_i]'$ is random. Thus, this implies that separate OLS and SURE are not proper specifications for the current data.

6 Parameter Estimation

Since we can reasonably conclude that equation (23) is properly specified, we will estimate the parameters. Defining $a = [a_1, \ldots, a_9]'$, a 9×1 vector of parameters, the estimator given by Swamy (1971) is

$$\begin{bmatrix} \hat{a} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} D(l)'H^{-1}D(l) D(l)'H^{-1}X \\ XH^{-1}D(l) \quad X'H^{-1}X \end{bmatrix}^{-1} \begin{bmatrix} D(l)'H^{-1}S \\ X'H^{-1}S \end{bmatrix} \tag{31}$$

where $\hat{a}$ and $\hat{\alpha}$ are the estimators of a and α respectively, and

$$S = [S_1', S_2', \ldots, S_9']'$$

$$D(l) = \begin{bmatrix} l & 0 & \dots & 0 \\ 0 & l & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & l \end{bmatrix} = \mathrm{diag}(l, \dots, l)$$

$$X = [X_1', X_2', \dots, X_9']'$$

and matrix H is:

$$\begin{bmatrix} X_1\hat{\Delta}_2X_1' + \hat{\sigma}_{11}I_T & 0 & \dots & 0 \\ 0 & X_2\hat{\Delta}_2X_2' + \hat{\sigma}_{22}I_T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_9\hat{\Delta}_2X_9' + \hat{\sigma}_{99}I_T \end{bmatrix}.$$

The estimation results are summarized in Table 4, which provides the estimates of a and $\alpha = [\bar{b}, \bar{c}]'$, the mean vector of $[b_i, c_i]'$. The estimates have the expected signs. Three of the $\hat{a}_i$'s, as well as, $\bar{b}$ and $\bar{c}$ are significant. In order to study how widely the values of b_i and c_i vary across regions on either side of their mean values, we compute a measure called the coefficient of variation. The coefficient of variation is calculated as $(\hat{\sigma}_{b_i}/|\bar{b}|) \times 100$ and $(\hat{\sigma}_{c_i}/|\bar{c}|) \times 100$, respectively, where $\hat{\sigma}_{b_i}$ and $\hat{\sigma}_{c_i}$ are the estimated standard deviations of b_i and c_i, which are

Table 4. RCR estimation results

SIC Regions	Coefficients a_1	$\bar{b}$	$\bar{c}$
New England	22.45 (8.79)		
Mid Atlantic	70.51 (49.64)		
South Atlantic	59.36 (33.63)		
East North Central	65.15 (46.89)		
East South Central	20.29 (6.91)	1.058 (0.029)	−0.0003 (0.00005)
West North Central	22.61 (20.86)		
West South Central	41.20 (25.20)		
Mountain	15.21 (10.53)		
Pacific	80.80 (38.52)		

Note: Numbers in parentheses represent standard errors.

given in the diagonal terms of the variance-covariance matrix $\hat{\Delta}$. The calculated values of the coefficient of variation of b_i and c_i are 5.9% and 58.9% respectively, which indicate that there is a slight variation in b_i and a substantial variation in c_i across regions. In light of this result, it is reasonable to state that b_i and c_i are not the same for all regions.

Based on the findings that $Eb_i = \bar{b}$ and $Ec_i = \bar{c}$, as well as the findings that p_i and m_i are not random because a_i is not a random variable, we are able to obtain the following equations which relate the structural parameters (p, q, and m) to the regression coefficients (a, b, and c):

$$a_i = m_i p_i \ , \tag{32}$$

$$\bar{b} = Eq_i - p_i \ , \quad \text{and} \tag{33}$$

$$\bar{c} = -Eq_i/m_i \ . \tag{34}$$

One should note, however, that the relationship of $a_i = m_i p_i$ implies that both m_i and p_i could affect a_i. In order to properly measure the effect of these on the diffusion process, they need to be separated and isolated from each other. Hence, instead of m_i, we introduce m which is defined as the expected number of purchases over the period for which the structure of the model is constructed so that the effect of p_i on the diffusion process alone can be measured. We thus rewrite the above relationships in the following way:

$$a_i = m p_i \ , \tag{35}$$

$$\bar{b} = Eq_i - p_i \ , \quad \text{and} \tag{36}$$

$$\bar{c} = -Eq_i/m \ . \tag{37}$$

Essentially, the intercept term, a_i, reflects the amount of purchases by innovators in region i. Other things being equal, the higher the value of the coefficient of innovation, the larger the value of a_i. If $\sum_i a_i$ represents the amount of purchases by the innovators in the market as a whole, then the ratio of $a_i/\sum_i a_i$ measures the relative amount of purchases by the innovators in region i. The market size for region i, m_i, will then be derived by $m \bullet a_i/\sum_i a_i$. The overall market size, m, is readily available by applying the Bass model to the aggregate data. Applying the ratio to the overall potential market size yields the market potential in each region, m_i, as reported in Table 5. The use of the aggregate model in this context requires some clarifications and will be investigated in detail in the subsequent discussion section. The estimates of p_i and Eq_i obtained by replacing the estimates of a_i, $\bar{b}$, and $\bar{c}$ in equations (35)–(37) are also reported in Table 5.

It can be seen that there is a wide variation in p_i (0.003–0.014), but the mean of q_i, Eq_i, does not show a wide variation. On the basis of the estimated mean and variance of q_i, we can approximately evaluate the probability that the normally distributed q_i takes values within the interval $Eq_i \pm 2\sigma_{q_i}$, i.e.,

Table 5. Parameter values derived from the RCR estimates

SIC Regions	p	q Lower Bound	q Eq	q Upper Bound	m
New England	0.004	0.638	0.723	0.809	4036
Mid Atlantic	0.012	0.647	0.726	0.817	12676
South Atlantic	0.010	0.644	0.725	0.815	10671
East North Central	0.011	0.646	0.726	0.816	11712
East South Central	0.004	0.638	0.723	0.808	3648
West North Central	0.004	0.638	0.723	0.809	4065
West South Central	0.007	0.642	0.724	0.812	7407
Mountain	0.003	0.637	0.722	0.807	2734
Pacific	0.014	0.648	0.727	0.818	14526

$\Pr(E_{q_i} - 2\sigma_{q_i} < q_i < Eq_i + 2\sigma_{q_i}) = 0.95$, where σ_{q_i} is the standard deviation of q_i. It should be noted that the computed interval is the exact probability interval and is not the same as the usual concept of confidence intervals. As it is known, it is not correct to state that q_i lies in the computed confidence interval with the probability of 0.95 without regarding q_i as a random variable. Since q_i is a random variable, we state that q_i lies in the computed interval with a probability of 0.95, Since Eq_i and σ_{q_i} are unknown, their estimates can be used to obtain the estimate interval. Table 5 gives the estimated intervals for the nine regions.

7 Discussion

The empirical results indicate that the nine regions exhibit varying diffusion processes. The RCR model depicts the amalgamated structure of parameters where intercepts are not equal across regions while other parameters are stochastic. In the Bass model formulation, these parametric properties lead to a structural relationship where the coefficients of innovation and potential market size are nonstochastic and are non-uniform across regions, while the imitation coefficient is stochastic. Therefore, an aggregate model which assumes all the parameters are uniform is not appropriate. The different RCR estimates of p_i's, which show much wider variations than the separate OLS estimates, reveal varying adoption rates by innovators across regions.

Since the coefficient of imitation, q_i describes the behavior of the imitator segment, the randomness of this parameter, as revealed by the RCR model, seems to be plausible considering the particular characteristics of this segment in which consumers tend to purchase the product under pressure from adopters, i.e., peer pressure plays a significant role. Indeed, one can regard each region's coefficient of imitation as a random drawing from a common distribution. The varying regional characteristics also make the realized q_i deviate from the mean.

The empirical results reveal that the coefficient of innovation varies more across the nine regions than the coefficient of imitation. This result has useful implications for product introduction strategies. Given that the coefficients of innovation and imitation vary across the regions, the time to peak sales will also be different in these regions. Accordingly, product managers may decide on which market(s) are the best to introduce the products. There may be situations where it is not possible for firms to introduce its product simultaneously in all markets.[7]

These insights into the parameter structure gained from RCR modeling are not readily available from the separate OLS models which assume the regions are independent from each other. The RCR results clearly imply that it would be inappropriate to treat regions independently and separately, and that the appropriate model structure is the one that is supported by the data, i.e., the assumptions made must be consistent with the data.

We will analyze the estimation results in detail through comparisons of the RCR results with those of the aggregate and separate OLS models. From Tables 1 and 5, we see that the RCR estimates of Eq's are close to the aggregate OLS estimate of q, 0.7203, which in turn is close to a simple average of the nine OLS estimates, 0.7299. Thus, the aggregate model will yield an almost exact estimate of the mean of q. This is plausible since the aggregate model results in an averaging process. As for the p values of the RCR estimates, five out of the nine regions exhibit larger estimated values than the aggregate values while four regions exhibit smaller estimated values than the aggregate values. The correlation coefficient of the p values between the RCR estimates and the OLS estimates is not statistically significant ($r = 0.36$ with $p = 0.37$). A simple average of the nine RCR estimates is, however, close to the aggregate model estimate. The comparison of the separate OLS values with those of RCR reveals that six of nine RCR estimates are larger than the separate OLS estimates.

8 Managerial Implications

The empirical results demonstrate a substantial amount of variation across regions in the diffusion process of VCR's in the United States. The coefficients of innovation of the RCR model reveal that the Pacific region shows the highest

---

[7] The managerial implications derived from the varying innovation coefficient were offered by one of the referees. We would like to present them in the manuscript with the referee's permission.

p value among the regions, followed by the Mid Atlantic, the East North Central, and the South Atlantic regions. Each of these regions includes major metropolitan areas such as Los Angeles and San Francisco in the Pacific region, New York City and Philadelphia in the Mid Atlantic region, Chicago in the East North Central region, and Washington D.C. and Atlanta in the South Atlantic region, respectively. The lowest p value, on the other hand, is observed in the Mountain region, followed by the New England, the East South Central, and the West North Central regions. These regions consist of such states as Nevada, Utah, and Idaho. According to Rogers (1983), homogeneous regions exhibit faster diffusion rates than heterogeneous ones because of homophilous communication. The first group of regions may be characterized by the heterogeneous influx of the population, notably of the mobile segment, which leads to a large value of p due to their reliance on non-personal sources of information such as advertising. In these regions, the initial sales of the product could be higher than the second group of regions. On the other hand, the second group of regions may be characterized by a fairly stable population that has not had a large influx of new people to the area. The homogeneous population tends to have a small value of p. In the regions belonging to the first group, one would expect a C-curve cumulative penetration. In the regions of the second group, on the other hand, the expected pattern would be an S-curve.

The expected number of purchases, m, varies across regions partly due to the heterogeneous diffusion processes and partly due to different regional market sizes. If we can filter out the effect of different market sizes, the remaining variations can be attributed to the heterogeneous diffusion processes. In order to eliminate the effect of market size, we have calculated market penetrations, i.e., ratios of the expected number of purchases to the population and to the number of households, respectively. It can be seen from the market penetration ratios indicated by the last two columns in Table 6 that the penetration ratios vary substantially across regions. Two of the metropolitan regions, namely the Pacific and the Mid-Atlantic regions, show the highest penetration, while the Mountain and the West North Central regions show the lowest penetration. This result appears to be analogous to the result of the p value analysis. Hence, the market penetration values comply with the findings that regional heterogeneity does exist.

To further examine the information contained in the market penetration ratios, we have postulated a model to analyze the relationship between the penetration and the variables which reflect regional characteristics and variations. Considering how the VCR is used in households, taping television and cable television programs for later viewing seems to have become a common phenomenon in an increasing number of households in recent years. Other popular uses of the VCR at home include taping programs on TV off the air, watching movies on rental videocassettes, and viewing videocassettes recorded by video cameras or camcoders. The model is postulated to measure the influence of these effects on the potential market size of the VCR market. The independent variables include the number of TV households, households sub-

Table 6. Penetration of market potential

SIC Regions	m (thousands)	Population (thousands) [I]	Households (thousands) [II]	Population per Household	Population Penetration [m/I]	Household Penetration [m/II]
New England	4036	12737	3398	3.75	0.32	1.19
Mid Atlantic	12676	37280	9969	3.74	0.34	1.27
South Atlantic	10671	40916	11336	3.61	0.26	0.94
East North Central	11712	41737	11095	3.76	0.28	1.06
East South Central	3648	15209	4076	3.73	0.24	0.89
West North Central	4065	17576	4749	3.70	0.23	0.86
West South Central	7407	26864	7123	3.77	0.28	1.04
Mountain	2734	13023	3518	3.70	0.21	1.78
Pacific	14526	35737	9226	3.87	0.41	1.57

scribing to cable television, and households with annual income over $25,000 in each of nine regions. The data on TV households and cable TV households were collected from the *Broadcasting Cablecasting Yearbook* in which the U.S. market is divided into 211 areas, each defined as an Area of Dominant Influence (ADI). We have aggregated the ADI's into nine regions according to these variables. All the variables except for the p coefficient values are transformed into penetration terms by dividing the variables by households. After eliminating insignificant variables, the final model is estimated and shown in Table 7.

The model has a good fit with an adjusted $R^2 = 0.77$ and $F = 14.30$ ($p = 0.005$). The seemingly positive relationship of the market penetration ratios (m_i/households) to the innovation coefficient as observed above is statistically confirmed by the above model with a significant estimated parameter. The cable penetration, measured by the ratio of the number of households subscribing to cable television to the number of households in the region, is also found to be a good predictor of the market penetration.

Those metropolitan areas which have large innovation coefficients and achieve relatively high market penetration also tend to have high cable television penetration. On the other hand, the rural and suburban areas having low market penetration and small innovation coefficients tended to have low cable television penetration.

In the metropolitan areas, the heterophilous population yields a large p value. The heterophilous population slows down the diffusion process in general. However, the high level of market penetration in metropolitan areas is perhaps attributable to the fact that these urban consumers may have easier access to the products, broader choices and more competitive prices primarily due to a large number of outlets, coupled with heavy advertising and promotion activities in the regions. Within the framework of the five key dimensions proposed

Table 7. Estimation results of market penetration

Dependent Variable		
Potential market size penetration		
Independent Variables	Parameter Estimates	Standardized Estimates
Intercept	0.65 (0.09)	0.01
Cable TV households penetration	1.02) (0.33)	0.39
Innovation coefficients	22.96 (12.24)	0.63
An adjusted R^2	0.77	
F statistic	14.30 ($p = 0.005$)	

Note: Numbers in parentheses represent standard errors.

by Rogers (1983), these factors may enhance the consumer's understanding of the relative advantages and characteristics (or complexity) of the product and increase trialability and observability of brand variety. Busy metropolitan dwellers, especially working couples, might find the VCR highly compatible with their lifestyles.

The homophilous population in rural and suburban areas, on the other hand, indicates small p values and does not lead to a high level of market penetration. This is probably because of the low level of effects in these regions on the five key dimensions of innovations compared to the metropolitan population.

While using the U.S. parameters for national sales forecast will be useful for long-range product planning, understanding the diffusion characteristics of the different regions will allow managers to make better short term product plans. The empirical results imply that the location of the test market could have profound effects on the generalizability of the test results to a national market. If a region such as the West North Central is chosen, the initial sales may be smaller than that of the nation as a whole. Conversely, if the test market is in the Pacific region, test results may be inflated and lead the marketer to have an overly optimistic forecast. The manager should strive to either mix test markets or carefully interpret the test results.

9 Conclusion

When modeling regional differences using temporal, cross-sectional data, one of the key issues confronting researchers is what is an appropriate specification of parameter variations across cross-section units. In this paper, we argue that this issue can be solved as a testable hypothesis. If there is no prior information, one should explore the data and let the data determine the right specification instead of making an assumption *a priori*.

This study has shown that the Bass model, frequently used for forecasting purposes, is useful for analyzing the diffusion process in different geographic regions. The empirical analysis of the VCR market leads us to conclude that the market has varied diffusion processes which can be delineated by geographic regions. Although we have analyzed only one product category, the proposed modeling framework can be extended to other new technological products such as compact disc players, laser disc players, and cellular phones, to name but a few. Further accumulation of empirical findings across various product categories will enhance an understanding of the different regions and their unique characteristics and will help marketing managers formulate a more efficient and effective strategy for the introduction and maintenance of products in a heterogeneous marketplace.

References

Bass FM (1969) A new product growth model for consumer durables. Management Science 15:215–27
Brown LA (1981) Innovation diffusion: A new perspective. New York: Methuen and Co. Ltd.
Brown LA, Malecki EJ, Spector AN (1976) Adopter categories in a spatial context: Alternative explanations for an empirical regularity. Rural Sociology 41:99–118
Gore AP, Lavaraj VA (1987) Innovation diffusion in a heterogeneous population. Technological Forecasting and Social Change 32:163–7
Hagerstrand T (1965) A Monte Carlo approach to diffusion. Archives Europeennes de Sociologie 6:43–67
Hagerstrand T (1967) Innovation diffusion as a spatial process. Chicago: University of Chicago Press
Jain D, Rao RC (1990) Effects of price on the demand for durables: Modeling, estimation, and findings. Journal of Business and Economic Statistics 8:163–70
Kotler P (1988) Marketing management: Analysis, planning, implementations, and control. New Jersey: Prentice-Hall
Mahajan V, Mason CH, Srinivasan V (1986) An evaluation of estimation procedures for new product diffusion models. In: Innovation diffusion models of new product acceptance, Mahajan V, Wind Y (ed) Massachusetts: Ballinger Publishing Co 203–32
Mahajan V, Muller E, Bass FM (1990) New product diffusion models in marketing: A review and directions for research. Journal of Marketing 54:1–26
Mahajan V, Peterson RA (1978) Innovation diffusion in a dynamic potential adopter population. Management Science 24:1589–97
Mahajan V, Wind Y (1986) Innovation diffusion models of new product acceptance. Massachusetts: Ballinger Publishing Co
Maddala GS (1977) Econometrics. New York: McGraw-Hill
Moriarty M (1975) Cross-sectional, time-series issues in the analysis of marketing decision variables. Journal of Marketing Research 12:142–50
Raj B, Ullah A (1981) Econometrics: A varying coefficients approach. London: Croom Helm
Rao CR (1973) Linear statistical inference and its applications, New York: John Wiley and Sons
Rogers EM (1983) Diffusion of innovations. 3rd ed. New York: The Free Press
Scheffé H (1959) The analysis of variance. New York: John Wiley & Sons
Srinivasan V, Mason C (1986) Nonlinear least squares estimation of new product diffusion models. Marketing Science 5:169–178
Swamy PAVB (1971) Statistical inference in random coefficient regression models. Berlin: Springer-Verlag
Takada H (1989) Does the Bass model perform well internationally? In New-product development and testing, Henry WA, Menasco M, Takada H (ed) Massachusetts: Lexington Books 243–58
Ullah A, Racine J (1992) Smooth improved estimators of econometric parameters. In Readings in econometric theory and practice, Griffith WE et al (ed) in honor of G. Judge, North Holland
Wittink D (1977) Exploring territorial differences in the relationship between marketing variables. Journal of Marketing Research 14:145–55
Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57:348–68

First version received: 06.01.1992
Final version received: 08.09.1993