REX Y. DU and WAGNER A. KAMAKURA*

Household life cycle has been widely used as a determinant of consumer behavior and a basis for market segmentation. However, there is considerable disagreement about how life stages should be defined and how households progress through these stages. Existing studies use a priori definitions, which are tested on a cross-sectional survey of households collected at a single point in time and thus cannot reveal the real dynamics of the household life cycle. The Panel Study of Income Dynamics provides longitudinal data on household composition in the United States for a period of 34 years; the authors use this to identify empirically the most typical stages and paths that U.S. households have followed since 1968. They develop a hidden Markov model in which the stages of the household life cycle are taken as latent, unobservable states that are uncovered from the manifest household demographic profiles over the 34 years, assuming that households evolve through these latent stages following a first-order Markov process. The authors apply their results to classify members of another panel (Consumer Expenditure Survey) into life stages, which enables them to study the impact of the household life cycle on households' budgetary allocations, providing a comprehensive analysis of lifestyles (through expenditure patterns) over the household life cycle.

Household Life Cycles and Lifestyles in the United States

Since its introduction in the marketing literature in the 1950s (e.g., Lansing and Morgan 1955), the concept of household/family life cycle has been widely studied as a determinant of various types of consumer behavior, such as borrowing, home ownership, purchase of durables (Lansing and Kish 1957), entertainment (Hisrich and Peters 1974), consumption of energy (Fritzche 1981), and a wide range of other goods and services (e.g., Schaninger and Danko 1993; Wagner and Hanna 1983; Wells and Gubar 1966; Wilkes 1995; for a review, see Redondo-Bellon, Royo-Vela, and Aldas-Manzano 2001).

Although there is a consensus on the usefulness of the household life-cycle concept in marketing, there is considerable disagreement about the two interrelated steps in the operationalization of the life-cycle concept: (1) the classification of life stages (i.e., the main types of households in a particular population) and (2) the identification of life paths (i.e., the sequences that households follow through the various life stages). For example, Wells and Gubar (1966) prescribe a household life cycle of ten sequential stages based on age, marital, and employment statuses of the household head and age of the youngest child. Murphy and Staples (1979) propose another life-cycle model that contains 14 stages linked by multiple paths and is based on marital status and presence of children within each of three groups defined by age of the household head. Wilkes (1995) proposes a 15-stage hybrid of Wells and Gubar's (1966) and Gilly and Enis's (1982) typologies.

In the extant life-cycle models, households are classified into life stages on the basis of various a priori definitions that reflect the authors' own beliefs of the typical compositions and evolutions of households in a population. Defining life stages a priori also limits researchers to few demographic markers, typically on two or three levels each (Wilkes 1995). Though postulating specific life paths, the extant life-cycle models are all based on snapshots of a cross-section of households at a point in time and therefore cannot provide any empirical validation of how households actually move from one life stage to another. Thus, these models fail to quantify the transition probabilities between life stages, making these models incapable of predicting future stages given current stages.

In this study, we take advantage of the Panel Study of Income Dynamics (PSID), a nationally representative longi-

*Rex Y. Du is Assistant Professor of Marketing, Terry College of Business, University of Georgia (e-mail: rexdu@terry.uga.edu). Wagner A. Kamakura is Professor of Marketing, Fuqua School of Business, Duke University (e-mail: kamakura@duke.edu).

ISSN: 0022-2437 (print), 1547-7193 (electronic)

tudinal panel of approximately 8000 U.S. households that have been tracked annually from 1968 through the present (for more details, see Hill 1991). With such panel data, which contain state sequences of a rich set of demographic variables that define household statuses, a natural question is whether there is a Markov process that governs the transition probabilities from the states at time t – 1 to the states at time t. Assuming that the successive states of the observed multidimensional demographic variables are indirectly linked through an unobserved Markov chain, we fit a hidden Markov model (HMM; Aldous and Pemantle 1996; Elliott, Aggoun, and Moore 1995) to our data, identifying the typical life stages and life paths observed among U.S. households over the past three decades.

In summary, this study contributes to life-cycle modeling in several important ways: First, we identify life stages by empirically seeking the best way to represent the observed household demographic profiles, thus avoiding previous controversy over life-stage definitions. Second, our identification of life paths is based on direct evidence from a large sample of households over a long period of time. Finally, our approach takes advantage of the correlation among all the observed household characteristics and therefore is not limited to the few demographic markers often used in the literature.

In the rest of the article, we introduce the proposed HMM. We then apply the model to the PSID data, discuss the estimation results, and test its predictive validity. We combine the resulting life-stage classification scheme with the Consumer Expenditure Survey (CEX) to study the relationship between household life stages and consumption patterns (or lifestyles). We then illustrate how our life-cycle model can be used to project a household's expenditures over time. The article concludes with brief discussions of our findings and directions for further research.

AN HMM FOR HOUSEHOLD LIFE CYCLES

Our objective is to use household composition data from the PSID to identify the main types of households (or life stages) and the typical sequences in which households move through these stages (or life paths). For each year, we use the following variables:

- Marital status of the household head (married, never married, widowed, divorced, separated);
- Age of the household head (years);
- Employment status of the household head (working, unemployed, retired, disabled, homemaker, student);
- Employment status of the spouse, if any (working, unemployed, retired, disabled, homemaker, student);
- Number of other adults (none, 1, 2, 3, 4 or more);
- Presence of children younger than 7 years of age;
- Presence of children 7–14 years of age;
- Presence of children 15–18 years of age;
- Presence of children in college; and
- Household size (1, 2, 3, 4, 5, 6, 7, 8, 9 or more).

Permutation of these ten variables in theory could result in tens of thousands of types of households. Fortunately, these demographic characteristics are correlated cross-sectionally as well as longitudinally. To capture these correlations simultaneously, we use an HMM to identify latent classes that represent the most typical household compositions and evolution paths.

Hidden Markov models first became popular in electroacoustics, especially in speech recognition (for a review, see Juang and Rabiner 1991), and later received attention in the social sciences. Applications of HMMs in marketing include that of Brangule-Vlagsma, Pieters, and Wedel (2002), who conduct dynamic value segmentation using a time series of ordinal data; Liechty, Pieters, and Wedel (2003), who identify instances of local versus global visual attention when readers are exposed to print advertisements; and Montgomery and colleagues (2004), who identify unobservable goals that drive Web-browsing behavior. Among these applications, our proposed model is more closely related to that of Brangule-Vlagsma, Pieters, and Wedel (2002), though our data involve a time series that is much longer than the three periods these researchers consider, which prevents us from using traditional estimation methods. In addition, a methodological innovation of our HMM enables it to deal with discrete and continuous variables simultaneously.

We denote the K demographic variables observed for the N panel households over the sampling period of T + 1 years as follows:

$$(1)\qquad Y_{nt} = (y_{nt1}, y_{nt2}, y_{nt3}, \ldots, y_{ntK}),$$

$$\text{where } n = 1, 2, \ldots, N \text{ and } t = 0, 1, 2, \ldots, T.$$

We use these observed demographic variables to classify each household n into M latent states at each year t, through the indicator variable x_{nt}:

$$(2)\qquad x_{nt} = i, \text{ where } i = 1, 2, \ldots, M.$$

We assume that households move among the M latent states from one year to the next according to a first-order Markov process, defined by A, the transition probabilities,

$$(3)\qquad a_{ij} = P[x_{nt} = j \mid x_{nt-1} = i], \qquad a_{ij} \geq 0, \ \sum_{j=1}^{M} a_{ij} = 1,$$

and by Π, the probabilities that a household enters the sample at each state,

$$(4)\qquad \pi_i = P[x_{no} = i], \qquad \pi_i \geq 0, \ \sum_{i=1}^{M} \pi_i = 1.$$

The conditional probabilities of the observed demographic variables for household n at year t, given that the household is in state i, are represented by the following:

$$(5)\qquad b_i(Y_{nt}) = P[Y_{nt} \mid x_{nt} = i] = \prod_{k=1}^{K} b_i(y_{ntk}),$$

where

$$b_i(y_{ntk}) = \prod_{l=1}^{L} \left[\theta_{ik}^{(l)}\right]^{y_{ntk(l)}}, \ 0 \leq \theta_{ik}^{(l)} \leq 1,$$

if item k is multichotomous with L categories, or

$$b_i(y_{ntk}) = \frac{1}{\sqrt{2\pi\theta_{ik}^{(2)}}} \exp\left[\frac{-(y_{ntk} - \theta_{ik}^{(1)})^2}{2\theta_{ik}^{(2)}}\right],$$

if item k is continuous.

Equations 1–5 define our multinomial HMM. In essence, it assumes that (1) the observed demographic variables are indicators of the latent state that each household occupies at any given time and (2) transitions between the latent states follow a first-order Markov process. We need to estimate the transition probabilities (A), initial state probabilities (Π), and conditional probabilities (B).

Because we have included household head age, which by definition can only increase with time, as a state indicator, we impose an additional property on the underlying Markov process: There is a natural sequential relationship among the latent states; that is, households cannot "turn back the clock" by moving from a later state to an earlier state. Mathematically, this property implies that A, the transition probability matrix, must be upper diagonal. In other words, we assume that the life stages are ordered in a particular sequence and that households can only move (or skip) forward, which may seem overly restrictive at first. For example, this assumption might appear to prevent a married person to divorce and then later remarry. In reality, however, the model allows for this sequence of events by identifying a younger "married"; a "divorced"; and a subsequently joined, older "married" stage, if the data call for such a pattern. (Details about the estimation of our HMM with the expectation–maximization algorithm appear in an Appendix, which is available at http://www.fuqua.duke.edu/faculty/alpha/kamakura.htm.)

HOUSEHOLD LIFE CYCLES IN THE UNITED STATES

We apply our HMM to a sample of PSID households observed between 1968 and 2001. Because changes in household composition tend to be minor from one year to the next, we use every fourth year of data (i.e., Years 1, 5, 9, and so forth). Otherwise, the resulting transition probability matrix (A) would be dominated by the diagonal because of households' tendencies to remain in the same state in subsequent years, thereby concealing the changes in life stages over the long run. We subsequently investigate the robustness of our results for this treatment of data.

In addition, because most of the PSID households have entered and left the panel during its existence, we select only those for which we had at least three observations, resulting in a final estimation sample of 6887 households. The main reason for dropping "transient" panel members (i.e., those that provide data on no more than a single transition between two observations) is to avoid cases for which the data are extremely right or left censored, which otherwise might bias our estimates of the life-cycle transitions. Finally, we apply the PSID demographic weights in our model estimation, so that the results better reflect the population of U.S. households.[1]

We estimated the proposed life-cycle model, specifying up to 15 life stages. On the basis of the Bayesian information criterion (BIC), we chose the solution with 13 stages. Table 1 displays the BICs for different numbers of life stages.[2] Table 2 reports the parameter estimates of the conditional probabilities (B) for the 13-stage HMM model. On the basis of these results, we can roughly describe these life stages as follows:

Table 1

FIT MEASURES FOR HMM WITH DIFFERENT NUMBER OF LIFE STAGES

Number of Stages	*Number of Parameters*	*BIC*
11	439	573,901
12	485	568,380
13	532	564,017[a]
14	580	568,567
15	629	568,510

[a]Smallest BIC and, thus, the best fit.

1. Co/So = single or young couple with no child,
2. C1 = small household (couple) with children <7 years of age,
3. C2 = large household (couple) with children <15 years of age,
4. C3 = large household (couple) with older children,
5. C4 = small household (couple) with children <15 years of age,
6. S1 = single/divorced with no child,
7. C5 = small household with older children,
8. C6 = empty nest couple,
9. S2 = single/divorced with children <15 years of age,
10. S3 = divorced/widow with older children,
11. S5 = widowed empty nest,
12. S4 = divorced/single empty nest, and
13. C7 = retired/old couple with adult dependents.

This sequence of life stages might seem counterintuitive, given that the estimated model assumes that households move through these stages in that particular order. However, our HMM does not imply that every household goes through each of these 13 stages. First, households enter the PSID panel at different life stages. Thus, the initial state probabilities (π_i) represent the (unconditional) likelihood that a household would enter the PSID panel at state i, and therefore it would serve as an adjustment factor for the percentage of sample households that enter the panel at a given life stage. Second, some households leave the panel before the most recent data collection period. Finally, and most important, the estimated transition probabilities (A) in Table 3 show that transitions are observed only between relatively few pairs of stages, suggesting that most households skip some of the 13 life stages and that different households follow different paths through these stages.

Given the estimates of all the parameters (A, B, and Π) and the observed demographics, we calculated the posterior probabilities that a household belongs to one of the 13 life stages in any given year, and we classified the household's life stage in that year into the state of the largest posterior probability.[3] The most commonly observed transitions from one stage to another, according to our calibrated HMM model, appear in Figure 1. Each arrow in Figure 1 is drawn in proportion to the percentage of households that move

[1]The PSID weights are designed to enable unbiased estimation of descriptive statistics for the U.S. population of individuals and households (for a detailed description of these weights, see Hofferth et al. 1998, pp. 22–41).

[2]Other fit measures, such as Akaike information criterion and consistent Akaike information criterion, are consistent with BIC.

[3]In most cases, because the posterior probability for each state is close to either 1 or 0 for a particular household, the classifications are clear-cut.

Table 2
ESTIMATED DEMOGRAPHIC PROFILES OF THE 13 LIFE STAGES (B)

Demographics	*Level*	*Co/So*	*C1*	*C2*	*C3*	*C4*	*S1*	*C5*	*C6*	*S2*	*S3*	*S5*	*S4*	*C7*
Marital status of household head (%)	Married	**53**	**98**	**99**	**99**	**98**	1	**98**	**99**	2	3	1	1	**92**
	Never married	**38**	0	0	0	0	**62**	0	0	**30**	10	0	**20**	0
	Widowed	0	0	0	0	0	1	0	0	7	**33**	**98**	16	6
	Divorced	6	1	0	0	1	**27**	1	0	**41**	**42**	0	**57**	1
	Separated	3	1	0	0	0	9	0	0	21	11	1	6	1
Age of household head	M	26	30	37	45	39	34	51	62	34	53	75	60	70
	SD	4	5	4	5	5	8	6	11	7	9	9	11	7
Employment status of household head (%)	Working now	**90**	**94**	**95**	**92**	**96**	**87**	**89**	**49**	**59**	**63**	9	**46**	14
	Homemaker	0	0	0	0	0	0	0	0	17	11	**23**	4	1
	Retired	0	0	0	1	0	0	4	**45**	4	13	**57**	**35**	**79**
	Other	10	6	5	7	4	12	7	6	19	14	11	14	7
Employment status of spouse (%)	Working now	**34**	**50**	**51**	**61**	**72**	1	**65**	**37**	2	1	0	0	16
	Homemaker	4	**43**	**45**	**34**	**23**	0	**29**	**35**	1	0	0	0	**45**
	Retired	0	0	0	0	0	0	1	24	0	0	0	0	**27**
	Not applicable	**59**	2	0	1	1	**98**	1	0	**97**	**99**	**100**	**99**	5
	Other	3	5	4	4	4	0	5	4	0	0	0	0	6
Other adults (%)	None	**92**	**98**	**96**	**45**	**98**	**96**	**33**	**100**	**86**	**22**	**91**	**99**	**39**
	1	7	2	3	**34**	2	4	**52**	0	12	**56**	8	1	**49**
	2+	0	0	1	**21**	0	0	**16**	0	3	**22**	1	0	12
Children in college (%)		2	0	0	**17**	0	0	**22**	0	1	**13**	0	0	4
Family size (%)	1	**37**	1	0	0	0	**93**	0	0	2	4	**90**	**98**	2
	2	**60**	0	0	0	1	7	2	**100**	**29**	**56**	9	2	**29**
	3	2	**41**	0	0	**21**	0	**51**	0	**34**	**22**	1	0	**50**
	4	0	**45**	1	3	**75**	0	**37**	0	18	10	0	0	12
	5	0	12	**58**	**45**	2	0	9	0	10	5	0	0	3
	6+	0	1	**41**	**52**	0	0	1	0	8	3	0	0	5
Children <7 years of age (%)		1	**70**	**51**	12	11	0	3	0	**32**	4	0	0	4
Children 7–14 years of age (%)		0	11	**89**	**65**	**62**	0	9	0	**48**	10	0	0	6
Children 15–18 years of age (%)		0	0	11	**62**	15	0	24	0	18	16	0	0	4

Notes: Bold numbers indicate the primary levels typical of a household in a life stage.

Table 3

ESTIMATED INITIAL STATE (Π) AND TRANSITION PROBABILITIES (A) FROM THE HMM

π_i	*30%*	*12%*	*6%*	*5%*	*4%*	*10%*	*6%*	*6%*	*8%*	*6%*	*7%*	*2%*	*0%*
a_{ij}	*Co/So (%)*	*C1 (%)*	*C2 (%)*	*C3 (%)*	*C4 (%)*	*S1 (%)*	*C5 (%)*	*C6 (%)*	*S2 (%)*	*S3 (%)*	*S5 (%)*	*S4 (%)*	*C7 (%)*
Co/So	46	**47**	0	0	1	3	0	2	0	0	0	0	0
C1		44	**19**	0	**32**	4	0	0	1	0	0	0	0
C2			42	**51**	2	3	0	0	1	0	0	0	0
C3				44	2	1	**49**	1	0	2	0	1	0
C4					56	3	**38**	1	0	1	0	0	0
S1						77	**6**	**6**	**6**	3	0	3	0
C5							55	**37**	0	1	0	1	6
C6								93	0	0	4	1	2
S2									75	**22**	0	3	0
S3										70	**10**	**18**	1
S5											98	0	2
S4												98	2
C7													100

Notes: Bold numbers indicate transitions between life stages that are commonly observed in the PSID sample.

through the respective pair of life stages as a proportion of the population of households observed in the PSID data.[4]

A key advantage of our HMM model over the extant life-cycle models is that it offers not only a life-stage classification scheme (based on the initial state and conditional probabilities) but also a means to project a household's expected life paths (based on the transition probabilities). To the extent that life stages are a determinant of consumption, our life-cycle model can be used to project a household's future expenditures. To evaluate our life-cycle model's longitudinal predictive validity, we benchmark our model against one that uses 13 household head age dummies as the direct indicators of life stages. We select this benchmark model because household head age is the only immediately projectable household characteristic and has been commonly used in studies of life-cycle income and consumption in the economics literature (e.g., Deaton 1992; Gourinchas and Parker 2002). We limit our analysis to two dependent variables for which we have enough longitudinal data in the PSID panel: income and in-home food consumption. We first calibrate regression models with the first-year data of each household, explaining each of the two dependent variables as a linear function of stable household head characteristics (i.e., birth cohort, gender, ethnicity, and education), and of the life-stage indicators (either our 13-stage classification or 13 household head age dummies). We then combine the calibrated regression models with the projected life paths (based either on our life-cycle model or on household head age alone) to predict the dependent variables four and eight years ahead.[5] Table 4 compares the goodness-of-fit and predictive fit of the proposed model with the benchmark model. The results show that the life-stage indicators based on our life-cycle model outperform the household head age dummies in predicting future and current income and in-home food consumption.

HOUSEHOLD LIFE CYCLES AND LIFESTYLES IN THE UNITED STATES

The purpose of this section is to "drill down" further on the relationship between household life stages and consumption. Because patterns of expenditures across the various accounts of a household's budget provide hard evidence of differences in lifestyles, we hope that our analysis sheds light on the linkages between life cycles and lifestyles in the United States. We believe that this is important because life cycles and lifestyles are often used together as bases for market segmentation (e.g., the life cycle/lifestyle grid popularized by Spectra; see www.spectramarketing.com).

Because the PSID is focused on income, it provides limited data on consumption. Therefore, we take advantage of the CEX, a large-scale study conducted by the Bureau of Labor Statistics that provides information on expenditures, incomes, and other demographic characteristics of U.S. households. The National Bureau of Economic Research (http://www.nber.org/data/ces_cbo.html) provides annual extracts on consumption for each survey respondent from a sample that ranges between 1300 and 3300 households each year. Unfortunately, in contrast to PSID's static panel, the CEX panel is a rotating panel, thus precluding a longitudinal analysis at the household level. However, the CEX provides all the information necessary to classify each household into our 13 life stages, which enables us to analyze the relationship between consumption profiles (which we call lifestyles) and life stages over a period of 19 years (1980 through 1998), using a total cross section of 52,061 U.S. households.

The first step in our analysis is to classify the CEX sample into the 13 life stages we identified previously, using household composition data and the parameter estimates of the HMM model (because the CEX sample is not longitudinal, households are classified into the life stages through a standard latent-class classification). The classification of the CEX sample into the 13 life stages over the 19 available years suggests that the proportion of households at each life stage was fairly stable over time, except for a slight increase

[4]In Figure 1, we show only transitions observed for at least 1% of the sample households. Because most panel members enter or exit the panel during its duration and because this may happen at any life stage, the number of households observed exiting a particular life stage may be unequal to the number observed entering that same stage. For example, Figure 1 shows that 4.9% of all PSID households moved from stage S_2 to S_3, even though a smaller proportion entered stage S_2.

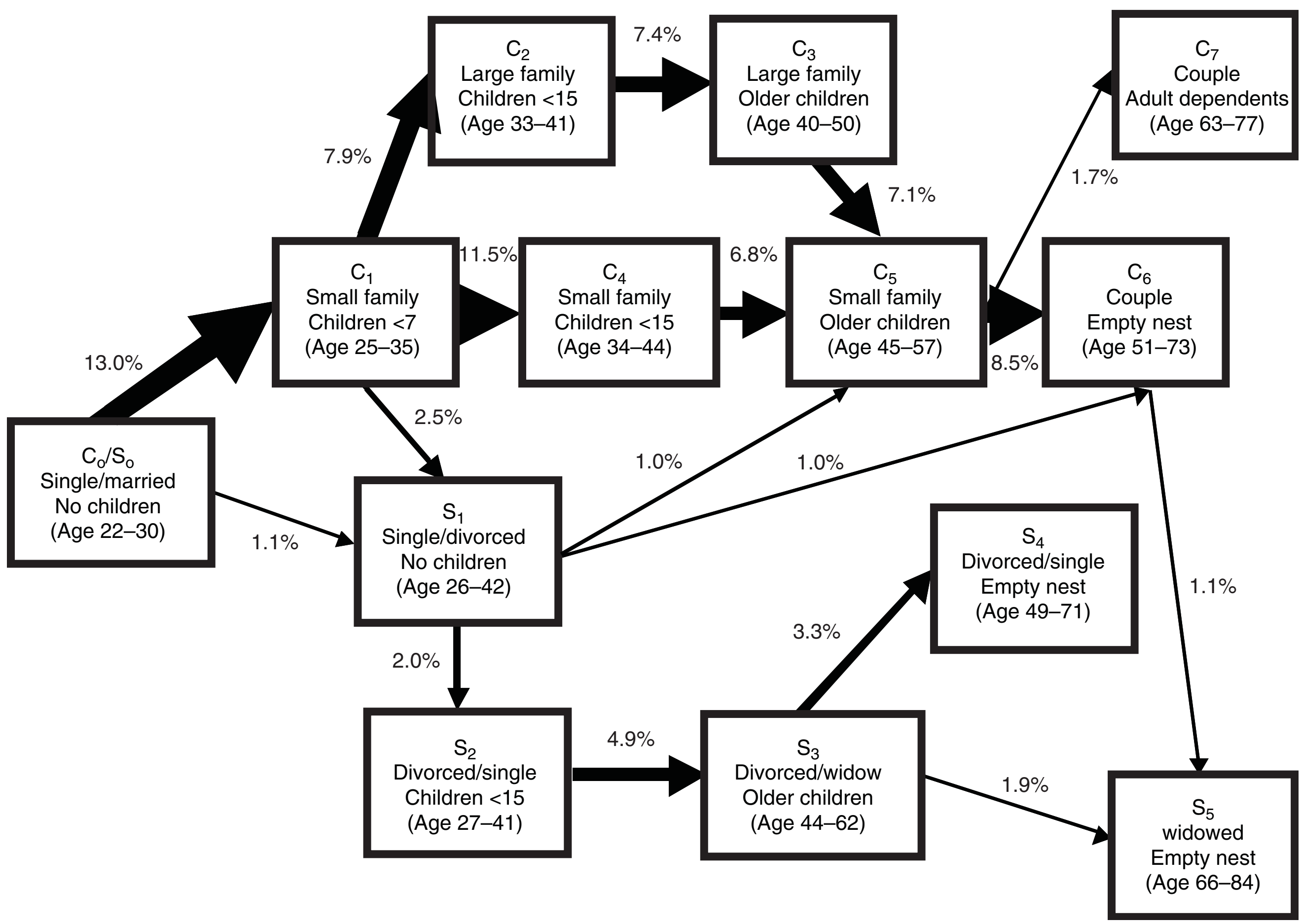
[5]The two dependent variables are not available for all households for eight years or more, reducing the sample used in this predictive validity test to approximately 3000 households. In terms of the average demographic profile, this predictive test sample is not significantly different from the estimation sample.

in the proportion of households in stage C6 (empty nest couple), perhaps as a result of aging baby boomers. When the households in the CEX sample are classified into the life stages, we can assess how the household life cycle affects consumption, after accounting for other factors, such as income, gender, ethnicity, and birth cohort. Although age (a component in our life-stage classification) is collinear with birth cohort, having expenditure data spanning almost two decades enables us to account for both life-stage and cohort effects on consumption.

The expenditure categories that the CEX tracks are not positive for all households, resulting in truncated data (e.g., not all households purchase a new car in a given year). Therefore, we resort to a Type-2 Tobit regression model to distinguish the incidence and the quantity components (Amemiya 1985, p. 385), using income, stable household head characteristics (gender, ethnicity, and birth cohort), and the life-stage indicators as the explanatory variables. We apply the Type-2 Tobit regression model to 35 major categories of expenses and assets obtained from the CEX.

Figure 1

HOUSEHOLD LIFE CYCLES IN THE UNITED STATES



Notes: The percentages reported are partially influenced by the attritions and expansions of the PSID panel and therefore are best interpreted in relative terms.

Table 4

PREDICTIVE VALIDITY TESTS FOR TWO LIFE-CYCLE MODELS

			Prediction (%)	
Dependent Variable	*Model*	*Current Fit (%)*	*Four Years*	*Eight Years*
Income	Life stages	156	189	158
	Age dummies	158	259	189
In-home food consumption	Life stages	59	67	61
	Age dummies	64	72	64

Notes: The measure of goodness-of-fit and predictive fit is the mean absolute percentage error between actual and fitted/predicted values.

Table 5 shows the average percentage of households in each life stage that have positive expenditure or ownership in a category, holding all the other explanatory variables at the population mean. We present only categories in which we found the most substantial differences in incidences across the life stages. For example, the average percentage of households with positive health insurance expenditure is lower in single and younger life stages, the lowest occurring in Co/So (single or young couple with no child), S1 (single/divorced with no child), and S2 (single/divorced with children <15) and the highest occurring in C6 (empty nest couple), C7 (retired/old couple with adult dependents), and S5 (widowed empty nest). In terms of lower-education expenditure, however, there is a different life-cycle pattern in incidence: The average percentage of households with positive lower-education expenditure is the highest in C1 (young couple with children <7), C2 (large household with children <15), and C4 (small household with children <15) and the lowest in Co/So (single/young couple with no child), C6 (empty nest couple), and S5 (widowed empty nest).

For categories in which incidence is high across all households, life-stage differences are largely manifest in the amount spent or owned, which will be reflected in the estimates from the Type-2 Tobit regression on the conditional quantity component. For ease of interpretation, we transformed the regression coefficients for life stages into deviations from the population means (see Table 6) so that they reflect how each life stage compares with the population average, after accounting for income, household head birth cohort, gender, and ethnicity.

We present only categories with average incidence approximately or greater than 80%. For each of these categories, Table 6 focuses on the life stages with either the two highest or the two lowest average household expenditures, showing substantial life-cycle differences. In general, the lowest levels of expenditure are observed for older households with single, divorced, or widowed heads (S5 and S4), and the highest are observed for large households with children (C2 and C3). For example, the top two tax- and rent-paying life stages are Co/So (single or young couple with no child) and S1 (single/divorced with no child), and the bottom two stages are S5 (widowed empty nest) and C7 (retired/old couple with adult dependents). Not surprisingly, top spenders on health care are households in C6 (empty nest couple) and C7 (retied couple with adult dependents), whereas Co/So and S1 spend the least. Households in S1 (single/divorced with no child) are the largest spenders on eating out, dry cleaning, and alcohol and tobacco products; they do not use much domestic help, and the bills for personal care and utilities are relatively small. In contrast, young couples with children <7 (C1) spend a lot more on domestic help, do not consume much alcohol or tobacco, and spend the least on eating out.

AN ILLUSTRATION OF OUR METHOD TO PROJECT LIFE-CYCLE CONSUMPTION NEEDS

The HMM we calibrated on the nationally representative PSID data can provide marketers with a valuable tool for life-cycle segmentation. The life stages identified empirically by our model represent the most common types of U.S. households over the past three decades. Based on widely available demographic variables, our life-stage classification scheme can be readily applied (without any additional model estimation) to classify any other sample of households into the same typology of life stages, following the well-known latent-class model (Wedel and Kamakura 2000). As for the prior probability that a household is at any of the 13 life stages, we suggest the percentages of households in each life stage in the PSID sample: τ = 10.4%, 12.6%, 4.0%, 3.9%, 6.4%, 11.9%, 8.1%, 12.2%, 7.4%, 6.3%, 9.2%, 5.4%, 2.1%.[6] When data are obtained for household head age (y_{n1}) and the other nine nominal demographic variables that our model uses (y_{nk}, k = 2, …, 10), the posterior probability that this household (n) belongs to one of the 13 life stages, for example, i′, can be easily computed following the Bayes rule:

$$P(n \in i' | \Theta, Y_n) = \frac{\tau_{i'}\left\{\frac{1}{\sqrt{2\pi\theta_{i',1}^{(2)}}}\exp\left[\frac{-(y_{n1}-\theta_{i',1}^{(1)})^2}{2\theta_{i',1}^{(2)}}\right]\right\}\prod_{k=2}^{10}\theta_{i'k^*}}{\sum_{i=1}^{13}\tau_i\left\{\frac{1}{\sqrt{2\pi\theta_{i,1}^{(2)}}}\exp\left[\frac{-(y_{n1}-\theta_{i,1}^{(1)})^2}{2\theta_{i,1}^{(2)}}\right]\right\}\prod_{k=2}^{10}\theta_{ik^*}}, \quad (6)$$

where θs are the parameter estimates we report in Table 2 and k* denotes the particular level occupied by the household in a nominal demographic variable.

Furthermore, marketers can develop regression models similar to the one we used in our analysis of the CEX data to predict the annual expenditures of a household in a wide range of categories as a function of life stages, income, and other demographics, such as household head ethnicity, education, and gender. For example, to target customer acquisition campaigns more effectively, a marketer can first classify prospects into life stages and then determine the expected annual consumption needs of each prospect, using the demographic information available in rented mailing lists or geodemographic databases, or the marketer can apply our results to his or her current customer database to determine the total consumption needs for each customer (and thus the firm’s share of his or her wallet).

As an illustration, consider a hypothetical consumer, John, who is born after the 1950s, is a 28-year-old white male, is married, and has one child younger than seven. Both John and his wife are working. Together, they earn an annual income of $52,292. Following the latent-class classification procedure we previously outlined, we can calculate the posterior probability of John’s household belonging to each of the 13 life stages. As it turns out, the chance of John’s household being in stage C1 is 99.8%. In other words, our model suggests that a household such as John’s can be classified with near certainty as in C1 (small household [couple] with young children).

Next, suppose that we are interested in predicting John’s household’s current annual expenditure on durable goods (excluding cars). To do so, we use three sets of predictors: stable demographics (John’s birth cohort, gender, and ethnicity), income, and the life-stage classification of John’s

[6]Note that τ should be distinguished from Π (the initial state probabilities), the estimates of which we report on the top of Table 3, representing the likelihood that a household will enter the PSID panel at a particular life stage.

Table 5
AVERAGE PERCENTAGE OF HOUSEHOLDS WITH POSITIVE EXPENDITURE/OWNERSHIP BY LIFE STAGE

	Life Stages												
Expenditures/Assets	*Co/So (%)*	*C1 (%)*	*C2 (%)*	*C3 (%)*	*C4 (%)*	*S1 (%)*	*C5 (%)*	*C6 (%)*	*S2 (%)*	*S3 (%)*	*S5 (%)*	*S4 (%)*	*C7 (%)*
Health insurance	52	58	58	64	61	45	69	84	**39**	72	**98**	74	**98**
Life insurance	60	72	73	66	**78**	47	73	66	50	59	**45**	46	68
New car	33	33	33	47	30	20	**43**	21	27	32	**7**	13	25
Airfare	32	20	**19**	25	22	**42**	29	34	21	30	31	33	30
Books and maps	74	75	82	77	**85**	71	69	56	71	59	**33**	48	55
Higher education	37	28	44	**58**	51	27	48	15	34	30	**6**	13	21
Lower education	**1**	47	**40**	21	29	2	8	**1**	28	6	**1**	**1**	5
Car payment	**60**	54	50	55	52	35	57	28	35	45	**8**	19	33
Interest payment	**67**	**66**	65	64	67	59	58	34	45	50	**16**	31	36
Mortgage payment	49	74	77	79	**84**	49	78	80	**49**	69	69	59	**81**
Home maintenance	45	67	68	67	73	**42**	67	76	**42**	61	66	54	**79**
Pension contribution	23	24	23	20	31	**34**	32	23	18	25	**3**	26	10
Asset: home	**33**	61	70	71	77	40	80	**86**	40	74	79	66	**90**

Notes: We report only categories with the most substantial differences across life stages. Extreme values are in bold.

Table 6

DEVIATION FROM POPULATION AVERAGE BY LIFE STAGES AFTER ACCOUNTING FOR INCOME (IN DOLLARS) AND STABLE DEMOGRAPHICS

	Life Stages												
Expenditures/Assets	*Co/So*	*C1*	*C2*	*C3*	*C4*	*S1*	*C5*	*C6*	*S2*	*S3*	*S5*	*S4*	*C7*
Taxes	**2,204**					**3,683**					(2,309)		(3,040)
Rent	**2,724**					**2,884**					(1,553)		(2,917)
Food in the house			**2,197**	**3,074**							(2,158)	(2,231)	
Food out		(435)				**388**	**367**				(423)		
Alcohol and tobacco		(144)				**173**				**214**	(211)		
Clothing			**570**	**965**							(998)	(825)	
Dry cleaning	**108**					**128**					(87)		(72)
Personal care				**108**		(130)	**84**					(71)	
Durables	**521**	**435**									(628)	(621)	
Utilities			**510**	**761**		(818)						(705)	
Telephone				**287**			**199**				(231)	(188)	
Domestic help	(751)	**982**	**466**			(521)							
Health	(524)					(681)		**490**	(523)				**1,036**
Health insurance	(665)					(683)		**652**					**823**
Life insurance			**819**		**1,134**	(1,190)					(1,340)		
New car			(5,386)	(12,159)							**19,666**	**12,014**	
Car maintenance				**1,350**			**966**				(1,113)	(885)	
Mass transport		(127)	(130)			**87**							**73**
Automobile insurance		(247)	(237)	**287**			**402**						
Airfare	**224**		(128)					**230**		(136)			
Books and maps				**94**			**81**	(50)			(82)		
Magazines and news		**221**	**231**								(136)	(125)	
Recreation				**525**	**816**						(886)	(528)	
Higher education				**3,006**			**2,965**				(3,880)	(2,492)	
Lower education	**963**	(1,558)	(1,396)								**1,983**		
Charity			**729**					**788**	(1,417)	(732)			
Car payment	(2,963)	(2,238)									**7,163**	**3,654**	
Interest payment	**2,102**	**2,026**									(5,361)	(2,418)	
Mortgage payment	(2,902)		**1,503**		**1,775**				(2,427)				
Home maintenance			**1,130**			(1,487)		**1,221**	(2,098)				**1,485**
Pension contribution				(679)				**1,318**			(1,024)	**1,252**	
Asset: home						**17,999**	(15,303)		**14,317**	(7,569)			
Asset: checking				(5,130)					(3,850)		**4,948**	**3,873**	
Asset: savings			(9,853)	(13,216)							**21,005**		**16,872**
Asset: stocks/bonds	(34,735)			(31,561)							**40,369**	**34,671**	

Notes: Bold numbers indicate the two stages with the highest volume in the category, and numbers in parentheses indicate two stages with the lowest categories.

household. Plugging these predictors into the Tobit regression model for durable goods that we calibrated on the CEX data, we find that (1) the likelihood of John's household having any positive annual expenditure on durable goods is approximately 96% and (2) conditional on having any positive expenditure, the expected value is approximately $1,254. Similarly, we can estimate the expenditures of John's household in other CEX categories, thus forming a broad picture of its current lifestyle.

Aside from classifying John's household into the life stages on the basis of its current composition, our HMM can make projections about how it will move through the life stages over time, which in turn can be used to predict its future consumptions. For example, we may be interested in estimating the household's annual expenditure on durable goods 4 (as well as 8, 12, 16, or 20) years into the future.[7] To achieve this goal, we first need to determine the probability that John will be in each of the 13 life stages in four years. This can be done by multiplying the vector of current probabilities of being in each life stage by the estimated transition probability matrix A that we report in Table 3. Subsequently, we plug these estimated life-stage probabilities into the income regression model that we calibrated on the PSID data (as described in testing the predictive validity of our model), which gives us estimates of John's household income 4 years into the future. Finally, with the estimated life-stage probabilities and income, we can estimate John's annual expenditure on durables, using the Type-2 Tobit regression model we described in the analysis of the CEX data. Table 7 reports the predictions about John's future life-stage probabilities, income, probability of having positive expenditure on durables, and the conditional means of expenditure on durable goods, all of which are based on his current demographic profile.

CONCLUDING REMARKS

We attempted to identify empirically the most common types of households and life paths observed between 1968 and 2001 in the United States. To check the robustness of our findings, we subsequently report the results from tests of (1) the representativenss of the estimation sample and (2) the validity of the first-order assumption in our HMM model.

Recall that when we estimated our model, to avoid any potential biases in the Markov probabilities due to left or right data censoring, we considered only the PSID panel members who provided at least three observations. To check the representativeness of our selected estimation sample, we compared its demographic profile with those of the full PSID sample and the CEX sample. The results showed that though household heads in our estimation sample were, on average, approximately 1.4 years older and that a slightly smaller percentage were never married than was the case in the other samples, in general, our selected sample is in line with the other two samples in terms of the other key variables we used to identify our HMM. In case of significant sample differences in future applications, the estimates for initial state probabilities (Π) need to be applied with caution, whereas the estimates for transition probabilities (A) and conditional probabilities (B) may be less dependent on the sampling frame.

Another issue regarding the representativenss of our estimation sample stems from our use of every fourth observation of each household (for reasons we noted previously). To check the sensitivity of our results to this data treatment, we estimated our model using every third and fifth observation, respectively. We then compared the estimated latent stages and transition probability matrices with those based on every fourth observation. The profiles of the estimated latent stages were essentially the same across the three sets of data. However, note that the corresponding transition probability matrices cannot be compared directly, because they refer to the transition probabilities over every three-, four-, and five-year period, respectively. To make them comparable, we first defined the four-year transition probability matrix as A_4 and the three- and five-year transition probability matrices as A_3 and A_5. Then, we compared (A_3)^(10/3) and (A_5)^(10/5) with (A_4)^(10/4) because all three represent a ten-year-ahead transition probability matrix. To ascertain how close these probability matrices are to one another, we used the cross-entropy probability test (Theil 1972), with (A_4)^(10/4) as the benchmark. The entropy (i.e., perfect fit) for (A_4)^(10/4) was 14.3, whereas the cross-entropies for (A_3)^(10/3) and (A_5)^(10/5) were 14.4 and 14.5, respectively. The cross-entropy under the null hypothesis of equal transition probabilities was 22.6. Thus, the cross-entropy test suggests that our model estimates are robust to our treatment of data.

In developing our model, we assumed that households move through the life stages according to a first-order Markov process (similar to previous applications of the HMM in marketing).[8] To check the first-order assumption, our first-order HMM can be compared with a zero-order model and a second-order HMM. We estimated a zero-order 13-state model, which resulted in a log-likelihood of –316,690. In contrast, the first-order 13-state model has a log-likelihood of –278,406. The difference between the log-likelihoods (38,284) is significant by any likelihood-based goodness-of-fit test, providing strong evidence that the zero-order model should be rejected in favor of the first-order model. We leave the second-order extension for further research. One alternative to higher-order Markov processes would be to model the probability of staying at a given stage as a continuous hazard model (e.g., with an exponentially decreasing survival rate), whereas the transition probabilities, given that the household is changing stages, is modeled by a simple first-order Markov process. Another promising alternative is to consider a parsimonious, yet flexible formulation for higher-order Markov processes (Raftery 1985).

[7]We make the projection every four years because the estimated first-order Markov process operates at a four-year level.

[8]We note that a Markov process of any order can be represented as a first-order one by redefining the state space appropriately. Given that the states are not defined a priori in an HMM, the first-order assumption is less restrictive than it may seem because the data determine both the state space and the transition matrix. We believe that the first-order assumption of our HMM is reasonable for the purpose of identifying latent life stages, given that we determine the number of latent states in our model empirically. In addition, the parsimony of the parameter space and the ease in interpreting the results add to the appeal of the first-order assumption in practice.

Table 7

PROJECTIONS OF JOHN'S LIFE STAGES, INCOME, AND EXPENDITURES ON DURABLES

	Life Stages														Durables	
Year	*So/Co*	*C1 (%)*	*C2 (%)*	*C3 (%)*	*C4 (%)*	*S1 (%)*	*C5 (%)*	*C6 (%)*	*S2 (%)*	*S3 (%)*	*S5 (%)*	*S4 (%)*	*C7 (%)*	*Income ($)*	*Incidence (%)*	*Expen-diture ($)*
Current		99.8			.2									52,292	96.4	1,254
+4		44.0	19.3	.1	31.8	4.0	.1	.2	.6					50,193	95.4	1,126
+8		19.4	16.6	9.9	32.3	6.4	12.5	.9	1.1	.5		.2		46,756	94.3	966
+12		8.5	10.8	12.9	24.9	7.4	24.6	6.3	1.4	1.3	.1	.9	.8	48,006	93.6	912
+16		3.8	6.2	11.2	17.2	7.3	30.0	15.7	1.6	2.1	.5	2.0	2.4	52,377	93.3	934
+20		1.7	3.3	8.1	11.2	6.6	29.1	26.3	1.7	2.7	1.4	3.3	4.5	58,006	93.3	992

REFERENCES

Aldous, D. and R. Pemantle, eds. (1996), *Random Discrete Structures*: *The IMA Volumes in Mathematics and Its Applications*, Vol. 76. New York: Springer.

Amemiya, Takeshi (1985), *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Brangule-Vlagsma, Kristine, Rik M. Pieters, and Michel Wedel (2002), "The Dynamics of Value Segments: Modeling Framework and Empirical Illustration," *International Journal of Research in Marketing*, 19 (3), 267–86.

Deaton, Angus (1992), *Understanding Consumption*. Oxford: Oxford University Press.

Elliott, R.J., L. Aggoun, and J.B. Moore (1995), *Hidden Markov Models: Estimation and Control*. New York: Springer-Verlag.

Fritzche, David J. (1981), "An Analysis of Energy Consumption Patterns by Stage of Family Life Cycle," *Journal of Marketing Research*, 18 (May), 227–32.

Gilly, Mary C. and Ben M. Enis (1982), "Recycling the Family Life Cycle: A Proposal for Re-Definition," in *Advances in Consumer Research*, Vol. 9, Andrew A. Mitchell, ed. Ann Arbor, MI: Association for Consumer Research, 271–76.

Gourinchas, Pierre O. and Jonathan A. Parker (2002), "Consumption Over the Life Cycle," *Econometrica*, 70 (1), 47–89.

Hill, Martha S. (1991), *The Panel Study of Income Dynamics: A User's Guide*, Vol. 2. Newbury Park, CA: Sage Publications.

Hisrich, Robert D. and Michael P. Peters (1974), "Selecting the Superior Segmentation Correlate," *Journal of Marketing*, 38 (July), 60–63.

Hofferth, S., F.P. Stafford, W.J. Yeung, G.J. Duncan, M.S. Hill, J. Lepkowski, and J.N. Morgan (1998), *A Panel Study of Income Dynamics: Procedures and Codebooks*, (accessed October 18, 2005), [available at http://psidonline.isr.umich.edu/data/Documentation/Fam/1993/93guide.pdf].

Juang, B.H. and L.R. Rabiner (1991), "Hidden Markov Models for Speech Recognition," *Technometrics*, 33 (August), 251–72.

Lansing, John B. and Leslie Kish (1957), "Family Life Cycle as an Independent Variable," *American Sociological Review*, 22 (October), 512–19.

——— and J.N. Morgan (1955), "Consumer Finances over the Life Cycle," in *Consumer Behavior*, Vol. 2, Lincoln H. Clark, ed. New York: New York University Press, 36–51.

Liechty, John, Rik Pieters, and Michel Wedel (2003), "Global and Local Covert Visual Attention: Evidence from a Bayesian Hidden Markov Model," *Psychometrika*, 68 (December), 519–41.

Montgomery, Alan, Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579–86.

Murphy, Patrick E. and William A. Staples (1979), "A Modernized Family Life Cycle," *Journal of Consumer Research*, 6 (June), 12–22.

Raftery, A.E. (1985), "A Model for High-Order Markov Chains," *Journal of the Royal Statistical Society, Series B*, 47 (3), 528–39.

Redondo-Bellon, I., M. Royo-Vela, and J. Aldas-Manzano (2001), "A Family Life Cycle Model Adapted to the Spanish Environment," *European Journal of Marketing*, 35 (June), 612–38.

Schaninger, Charles M. and William D. Danko (1993), "A Conceptual and Empirical Comparison of Alternative Household Lifecycle Models," *Journal of Consumer Research*, 19 (March), 580–94.

Theil, Henry (1972), *Statistical Decomposition Analysis*. Amsterdam: North–Holland.

Wagner, Janet and Sherman Hanna (1983), "The Effectiveness of Family Life Cycle Variables in Consumer Expenditure Research," *Journal of Consumer Research*, 10 (September), 281–91.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual Methodological Foundations*, 2d ed. Boston: Kluwer Academic.

Wells, William and George Gubar (1966), "Life Cycle Concept in Marketing Research," *Journal of Marketing Research*, 3 (November), 355–63.

Wilkes, Robert E. (1995), "Household Lifecycle Stages, Transitions and Product Expenditures," *Journal of Consumer Research*, 22 (June), 27–42.