

Two-Stage Partial Observability Models of Innovation Adoption

Christophe Van den Bulte
University of Pennsylvania

Gary L. Lilien
The Pennsylvania State University

October 9, 2001

Acknowledgements

We benefited from comments by David Bell, Marnik Dekimpe, Jehoshua Eliashberg, Peter Fader, Teck Ho, Wendy Moe, and audience members at the 1999 INSNA Sunbelt Conference. We thank Thomas Valente for providing us with the *Medical Innovation* data set. Financial support from the Richard D. Irwin Foundation and Penn State's Institute for the Study of Business Markets is gratefully acknowledged.

Correspondence address: Christophe Van den Bulte, The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104-6371.
Tel: 215-898-6532; fax: 215-898-2534; e-mail: vdbulte@wharton.upenn.edu.

Two-Stage Partial Observability Models of Innovation Adoption

Abstract

Many theories on the adoption and diffusion of innovations posit that adoption is the outcome of a decision process. A two-stage conception of that process involving an awareness stage followed by a stage combining evaluation and adoption seems particularly useful. Several empirical studies indicate that the effect of mass media and change agents is relatively concentrated in creating awareness, while the effect of personal influence from earlier adopters (i.e., social contagion) is more concentrated in bringing about a positive evaluation and hence adoption. A frustrating problem, however, is that most data record only the final outcome of the process, i.e. the time of adoption. We bridge this gap in richness between theory and data by developing new event history models, which we call partial observability models of innovation adoption. An application to the classic *Medical Innovation* data illustrates that these finer-grained models can not only provide better descriptive fit but, more importantly, can also detect weak social contagion patterns that traditional event history models cannot detect in the presence of strong mass media effects.

In *A Matter of Taste*, Lieberman (2000) offers an intriguing anecdote. As he was writing this book on the dynamics in the popularity of first names, public interest was turned toward a sexual liaison between President Bill Clinton and a White House intern, Monica Lewinsky. This episode received extremely wide media coverage. A fellow sociologist asked Lieberman what he thought would be the consequences for the name *Monica*: would the name become more or less popular because of the scandal? Assuming that the name was dormant, i.e., that few girls were being named *Monica* in the years immediately preceding the scandal, Lieberman offers the following reasoning:

“Here we start with a condition in which parents are ignoring the name Monica. Because of this highly publicized event, it is more or less impossible for them to ignore the name in the same way. If many parents find the name even less attractive because of the scandal (the contamination effect), we can expect the dormancy to continue. But here’s the catch. If a small proportion of parents use the name because of the [media] focus on Monica Lewinsky or because the name *Monica* seems enhanced, then the name will gain in popularity even if the same events contaminate the name for the vast majority of parents” (Lieberman 2000, p. 264)

Lieberman argues that the press coverage of the Lewinsky affair has a dual effect: On the one hand, it makes more parents aware of the name *Monica*, increasing the odds that parents would consider giving it to their baby girl. On the other hand, the connotation of scandal makes it less likely that parents who are aware of the name evaluate it positively and use it. Hence, a single event, press coverage of a scandal, can have opposite effects on the probability of being aware of a name and on that of evaluating it positively. The net effect, Lieberman notes, is hard to predict

and may even be nil if the two effects balance each other out. Hence, a researcher who did not make the distinction between both stages of parental decision making, awareness and evaluation given awareness, and collapsed them into a single stage, might very well conclude the scandal did not matter, even when it had sizable but opposite effects on awareness and evaluation.

Appropriately distinguishing between the awareness and evaluation decision stages can also be important to separate the effect of two different forces, each operating at different stage, such as advertising and social contagion. Van den Bulte and Lilien (2001) raise this issue in their re-analysis of *Medical Innovation*—the landmark study by Coleman, Katz and Menzel (1966) on the role of social contagion in the adoption of innovations. Using a situational analysis based on historical materials and current diffusion theory, Van den Bulte and Lilien find no reason to expect social contagion to have been a very important factor in physicians' decision to adopt the tetracycline drug. Instead, they suggest that aggressive marketing efforts may have been a key driver of the adoption decision. Using event history analysis, they find that evidence of social contagion disappears once they control for marketing effort, and conclude that they are unable to reject the null hypothesis of no social contagion. However, they stop short from claiming that social contagion did not matter, citing self-reports from physicians tabulated in the original *Medical Innovation* publications, and a number limitations in the data and analysis. One of those limitations is of interest here: the inability of their hazard models to distinguish between the awareness stage and the evaluation/adoption stage (conditional on awareness) of the adoption process. Modeling the effect of marketing effort and social contagion without distinguishing between awareness and evaluation might produce misleading results. When marketing effort is quite important in creating awareness, and social contagion is only moderately important in persuading actors to adopt the innovation, but both explanatory variables are forced into a single-

stage model, the weaker social contagion effect could be washed out by the stronger marketing effort effect, erroneously suggesting that social contagion is not at work.

Both the *Monica* and the *Medical Innovation* examples emphasize that distinguishing awareness from evaluation can be critical for empirical studies to arrive at a correct understanding of the causal mechanisms underlying the adoption of innovations and other social practices. Yet, few empirical studies investigating the mechanisms underlying innovation adoption make this distinction. The dominant approach is to directly relate the event or time of adoption to explanatory variables in a single-stage model. This is not because researchers do not believe in the staged nature of the adoption decision. There is consensus across several disciplines that individuals and organizations go through several stages before adopting an innovation or not, and that this staged process can be represented as a sequence of first becoming aware of the alternative(s) available for adoption followed by evaluating whether or which to adopt (e.g., Bonus 1973; Lin and Burt 1975; Rogers 1995; Weenig and Midden 1991; Zaltman and Stiff 1973). The problem lies not with theory but with data. Specifically, it is very hard, if not impossible, to obtain good data on time of awareness. Becoming aware is not an overt behavior. Unlike adoption, it does not leave a paper trail of purchase orders sent to commercial vendors, prescriptions collected by pharmacies, notifications submitted to regulatory bodies, or financial statements and other announcements made to shareholders. Asking respondents for retrospective accounts does not produce reliable data because becoming aware of an innovation is hardly memorable (Snyder 1991). Multi-wave data-collection designs in which respondents are periodically measured are expensive and not always feasible, since directly asking a potential adopter whether he or she is aware at a given point in time is likely to influence the awareness-generation process, yielding erroneous research conclusions.

So, researchers wary of the need to distinguish between awareness and evaluation in the innovation adoption process face a disconnect between theory, positing at least two distinct two stages, and data, recording only the final outcome of the process, i.e. intention to adopt or, more often, actual adoption. In this paper, we present two event history models that are specified such that they reflect the theoretical assumption of adoption being a two-stage process, but can be estimated using only traditional data about the time of adoption. Following Abowd and Farber (1982) we call these partial observability models to reflect the idea that they allow one to specify and empirically assess how variables affect each stage of the adoption process, even though one only observes the final outcome, i.e. the time of adoption.

We proceed as follows. First, we briefly review the literature on decision stages in the innovation adoption process, noting earlier findings that mass media communications and marketing effort tend to affect the awareness stage, while social contagion tends to have a larger effect on the evaluation stage. Next, we present several micro-level models of adoption. We start with the standard case of a single-stage process, extend it to the case with two stages, both of which are fully observed, and, finally, present two partial observability models, differing in their assumptions about whether or not people can return to a state of unawareness after having become aware. Subsequently, we return to the *Medical Innovation* data. We find not only that the two-stage models fit better than does a traditional single-stage model, but also that, when marketing effort is allowed to affect awareness and social contagion is allowed to affect evaluation, both effects are statistically significant. In a single-stage model, in contrast, social contagion is not significant. While only illustrative, this application to the *Medical Innovation* data suggests that the ability to separate the two stages, posited by theory, can lead to substantially different conclusions. We conclude by briefly discussing how partial observability

models should be useful tools not only for students of social contagion in innovation diffusion but also for students of legitimation and life course transitions.

INNOVATION ADOPTION AS A PROCESS: THEORY AND RESEARCH

We define innovation adoption as the acceptance or first use of an innovation (i.e., something new), and the adoption process as an actor's decision-making process that results in adoption (or not). We define diffusion as the increase in the extent to which an innovation has been adopted among a set of actors. Hence, adoption occurs at the level of the individual actor (micro-level), although it may be affected by social contagion, whereas diffusion occurs at the level of the population (macro-level).¹ Note that we define diffusion as a growth process, not as a contagion or imitation process: contagion is a causal mechanism that may, but need not, drive actors to adopt and the innovation to diffuse. These definitions of adoption and diffusion closely follow Zaltman and Stiff (1973).

Stages in the innovation adoption process

Diffusion researchers have long recognized that adoptions are not instantaneous events, but the outcome of a decision process consisting of a sequence of actions and decisions (Rogers 1995; Robertson 1971; Zaltman and Stiff 1973). Many different sequences have been proposed, often heavily influenced by the nature of the innovation investigated in the empirical studies.

¹ This difference in level of aggregation is reflected in the data. For adoption, the data ideally indicate each actor's time of adoption or, equivalently, whether or not the actor has adopted for each time of observation. This can be represented in a set of binary indicators y_{it} that vary both across actors i and time t . For diffusion, the data are aggregated across actors, resulting in a time series of count data $n_t = \sum_i y_{it}$ or a time series of penetration rates $f_t = n_t / N$ where N is the number of actors in the population. Adoption data are typically analyzed using event history techniques while diffusion data are typically analyzed using regression techniques (Poisson regression, nonlinear least squares regression, etc.).

Probably the best-known sequence is the one proposed by Rogers (1962, pp. 81-86), mostly based on work in rural contexts. This conceptualization consists of five stages:

Awareness. At this stage, the actor gains knowledge of the existence of the innovation, but may not fully comprehend the innovation nor be motivated to seek further information.

Interest. At this stage, the actor becomes interested in the innovation but has not yet judged its utility in terms of the actor's own situation and seeks additional information about it.

Evaluation. At this stage, the actor assesses the appeal of the innovation to his or her own present and anticipated future situation, and decides whether to try it or not.

Trial occurs when the actor puts the innovation into use on a small scale to further determine its utility in the actor's own situation.

(Sustained) adoption. This stage is the continued full use of the innovation.

As we mentioned, this sequence is based on agricultural innovations, and some stages are not relevant in other contexts. For instance, when an actor becomes aware through exposure to a very persuasive advertisement or through word of mouth from a very close and enthusiastic peer, interest is achieved simultaneously with awareness and is hence not a separate stage. As another example, small scale trial is often possible for farmers by planting a new seed variant only on one small plot of land, but is often impossible for "lumpy" innovations such as durable manufacturing equipment or a new organizational form. Many researchers are not interested in repeated use, but only in the first time an innovation was used or implemented ("trial"), regardless of sustained use. Rather than collapsing stages, some researchers may prefer to subdivide some of the five stages. For instance, researchers interested in identifying the psychological underpinnings of the adoption decision process or in designing communication campaigns may split the evaluation stage into attitude formation followed by the formation of a

specific behavioral intention. Making this distinction is important to understand what prevents people who intend to adopt from actually doing so.

Although several staged acceptance processes can be posited, several of which are reviewed by Zaltman and Stiff (1973), most multi-stage empirical studies distinguish only two separate and temporally sequenced stages in the decision to adopt (e.g., Bonus 1973; Hauser and Urban 1977; Kalish 1985; Lin 1971; Lin and Burt 1975; Rogers 1995; Weenig and Midden 1991). During the awareness or knowledge stage, the individual or corporate actor learns about the innovation's existence and gains some understanding of how it functions. The next stage is evaluation or persuasion, during which the actor forms an assessment of the innovation that leads to the decision to adopt or not. In this two-stage conceptualization, Rogers' awareness and interest stages compress into a single awareness or knowledge stage. Also, adoption is equated with first-time use or trial, and assumed to occur immediately after positive evaluation.² Bandura's (1977, 1986) social learning theory similarly distinguishes between two separable processes: the acquisition of knowledge concerning the innovation and the adoption of that innovation in practice. Bandura's (1986, pp. 144-145) discussion of the latter stage refers to concepts like utility, incentives, benefits and detriments, suggesting that his two-stage conception is essentially identical to ours. Lieberman (2000, p. 167) distinguishes between awareness and use, and notes that what separates them is whether the object of adoption (e.g., a name) has appeal to the potential adopter (e.g., parents). Hence, the essence of his dichotomy is the same as ours. While obviously a coarser conceptualization than models with five or more steps, the awareness- evaluation-two-stage model is useful in assessing the effect of mass media versus

² The so-called KAP studies in family planning (e.g., Freedman and Takeshita 1969) also merge awareness with interest and trial with adoption, but keep evaluation and trial/adoption as a separate stages. This results in a three-stage model consisting of knowledge (*K*) to attitude (*A*) to practice (*P*).

personal influence on adoption.

Distinguishing between the drivers of awareness and evaluation

Explicitly distinguishing between awareness and evaluation may be critical to understand what drives adoption decisions, because research suggests that different factors affect these two stages differently. Two such factors are marketing efforts by change agents and word of mouth (or, more generally, social network exposure). Many early studies in rural contexts indicate that “impersonal sources are most important at the awareness stage, and personal sources are most important at the evaluation stage in the adoption process” (Rogers 1962, p. 99; see also Rogers 1995, p. 195). Similarly, studies of new drug adoption by physicians report that initial knowledge occurs mainly through commercial sources such as salespeople and direct mailings, whereas personal contacts with colleagues gain importance in later stages (e.g., Coleman et al. 1966; Peay and Peay 1984). Such findings generalize across many settings (e.g., Lin 1971; Lin and Burt 1975; Pareek and Singh 1969; Pelz 1983; Valente and Saba 1998). Given these differences in effectiveness across stages, modeling the effect of mass media and commercial efforts compared to word of mouth and other social contagion processes without distinguishing between awareness and evaluation may produce misleading results, as mentioned earlier.

Some marketing researchers interested in assessing the impact of advertising on new product diffusion have built macro-level models incorporating two or more stages to reflect the sequential nature of the adoption process. The empirical application and validation of such models has proven quite difficult. Two studies use data on the number of people at each stage at each point in time to estimate separate parameters for the awareness and evaluation sub-processes (Mahajan, Muller and Kerin 1984; Urban, Hauser and Roberts 1990). Macro-level data of such great richness in which each stage is fully observed are rarely available outside an

experimental setting. More often, one observes only the final outcome, i.e. the number of adoptions. In such situations of partial observability, researchers may attempt to derive the number of actors at various stages by decomposing the data, but this approach often results in ill-conditioned macro-level models having too many parameters to be estimated adequately from the available data (Silver 1984). Researchers have therefore most often collapsed their macro-level two-stage model into a single-stage specification prior to estimation (Kalish 1985) or have foregone empirical analysis entirely and focused on analytical model results only (Dodson and Muller 1978; Jedidi, Eliashberg and DeSarbo 1989; Muller 1983; Tapiero 1983). Other empirical studies of the impact of advertising on new product diffusion ignore the staged decision process altogether (Bass, Krishnan and Jain 1993; Golder 2000; Horsky and Simon 1983; Kalish and Lilien 1986; Mesak 1996; Simon and Sebastian 1987).

The problem of not observing each stage separately is compounded by the rather limited amount of information that macro-level diffusion data contain. Even very simple single-stage models have been shown to suffer from important estimation problems stemming from ill-conditioning, i.e. a lack of richness in the data (e.g., Van den Bulte and Lilien 1997). Over the last two decades, diffusion researchers across many disciplines have exploited the benefits of micro-level adoption data using hazard or event history models. Micro-level adoption data, being richer than aggregate time series, can provide a start toward a solution. In the next two sections, we extend the traditional discrete-time, single-stage micro-level hazard model into several two-stage specifications, including two that can be estimated even when one does not observe both stages of the adoption process but only the final outcome, i.e. time of adoption.

SINGLE-STAGE MODELS OF ADOPTION

Micro-level hazard models, also referred to as duration or event history models, allow one to

study how the probability of adoption varies not only over time—the focus of macro-level diffusion models—but also across actors. The ability of hazard models to exploit the richness of micro-level data also allows those models to better document the effects of contextual elements and of the pattern of communication on diffusion (e.g., Strang and Tuma 1993). The models we develop here fall in the category of discrete-time models. We first present the basic single-stage discrete-time hazard model, closely following Allison’s (1982) exposition. Next, we show how this model can be vested in random utility theory. This interpretation provides a sound theoretical underpinning for both the single-stage model and two-stage models.

The basic single stage model

We denote the uncensored time of adoption by actor i as the discrete random variable T_i . We assume that time takes only positive integer values ($t = 1, 2, 3, \dots$) and that we observe a total of n independent actors ($i = 1, \dots, n$) beginning at the time of launch $t = 1$ (i.e., we assume no left-censoring). Each actor is observed until time t_i , at which point either adoption occurs or the actor is censored from the study (e.g., because the actor drops from the panel or because the study ends before all actors have adopted). We assume that the time of censoring is independent of the adoption process. Let the censoring indicator c_i equal 1 if i is uncensored, and zero otherwise. Using $\Pr[\cdot]$ to denote probabilities in general, the log-likelihood for the data may be written as:

$$LL = \sum_i [c_i \ln \{ \Pr [T_i = t_i] \} + (1 - c_i) \ln \{ \Pr [T_i > t_i] \}], \quad [1]$$

and, using P to denote conditional probabilities, the hazard rate may be written as:

$$P_{it} = \Pr [T_i = t_i \mid T_i \geq t_i]. \quad [2]$$

The likelihood function for the data can be written as a function of the hazard rate. Let the

dummy variable y_{it} be 1 if i has adopted at time t , and 0 otherwise. Actors are not observed after adoption and we assume that they do not disadopt so $y_{it} = 1$ implies $y_{is} = 1$ for all $s > t$. Also, let the dummy variable d_{it} be 1 if i has already adopted ($y_{it-1} = 1$) or if i has dropped out of the study by time t , and $d_{it} = 0$ otherwise ($d_{it} = 1$ implies $d_{is} = 1$ for all $s > t$). Using properties of conditional probabilities, one obtains (Allison 1982):

$$LL = \sum_i \sum_t [1 - d_{it}] [y_{it} \ln\{P_{it}\} + (1 - y_{it}) \ln\{1 - P_{it}\}] \quad [3]$$

which is the log likelihood for a binary dependent variable (BDV) model for observations with $d_{it} = 0$. As a result, discrete-time hazard models can be estimated using standard BDV models such as binary logit or probit as follows: organize all the data in a panel with y_{it} as dependent variable, delete all post-adoption observations, and estimate a BDV model (Allison 1982).

A threshold interpretation

The fact that the discrete-time hazard rate is a probability and that discrete-time hazard models can be estimated as BDV models has an appealing consequence: one can interpret the hazard as the conditional probability that some behaviorally relevant threshold has been crossed. This interpretation directly relates the statistical hazard model to theoretical threshold models of adoption and contagion developed in sociology (Granovetter 1978; Granovetter and Soong 1983; Valente 1996) and to random utility theory popular in economics and marketing (e.g., Greene 2000; McFadden 1986). Event history analysis and threshold modeling have sometimes been presented as alternative methods (Strang and Soule 1998; Valente 1995), but one can interpret discrete-time hazard models as threshold models, as follows. We assume that actor i adopts when his evaluation of the innovation, say U_{it} , meets his minimum utility threshold, say ϕ_{it} . Also, we assume that U_{it} is a random variable additively composed of a deterministic element V_{it} and a

stochastic element - ε_{it} , such that $U_{it} = V_{it} - \varepsilon_{it}$, and that ε_{it} is independently distributed over i and t with cumulative distribution function (cdf) F . We then collect all explanatory variables affecting i 's perceived utility in the vector \mathbf{x}_{it} , such that the extent to which the innovation exceeds actor's i threshold at time t can be expressed as $V_{it} - \phi_{it} = \boldsymbol{\alpha}\mathbf{x}_{it}$, where $\boldsymbol{\alpha}$ is a vector of parameters (to be estimated). We can then write:

$$P_{it} = \Pr[y_{it}=1 | y_{it-1} = 0] = \Pr[U_{it} \geq \phi_{it}] = \Pr[V_{it} - \phi_{it} \geq \varepsilon_{it}] = F(\mu\boldsymbol{\alpha}\mathbf{x}_{it}) \quad [4]$$

where μ is a positive scale parameter related to the variance of ε that, for convenience, can be set equal to 1³. If ε is normally distributed, the probit model results. If ε is logistically distributed, the logit model results. Researchers who believe that the true adoption process is continuous in time, even though the data are grouped in discrete time intervals, may want to assume the Type I extreme value distribution. The resulting model, known as the complementary log-log model, uses $F(u) = 1 - \exp[-\exp(u)]$ and produces estimates that are independent of the length of the time interval when the true adoption process behaves according to a continuous-time proportional hazards model (Allison 1982; Prentice and Gloeckler 1978).

Equation (4) shows that the arguments of the cdf F , i.e. the arguments of the hazard model captured in $\boldsymbol{\alpha}\mathbf{x}_{it}$, express the extent to which the utility of the innovation to actor i at time t exceeds i 's threshold. The derivation of equation (4) also shows that discrete-time hazard models can be interpreted as representing the behavior of an actor maximizing a random utility function, where the arguments of the cdf F represent the deterministic utility component. While some

³ The scaling factor μ in a BDV model can be set to an arbitrary positive value. For convenience, one usually assumes that $\mu = 1$. This corresponds to assuming the variance of ε to be 1 in the probit model, $\pi^2/3$ in the logistic model, and $\pi^2/6$ in the complementary log-log model (Agresti 1990).

researchers using continuous-time models have interpreted the hazard as the limiting probability that the innovation meets the reservation utility level, we have seen no formal derivation for this claim (e.g., Saloner and Shepard 1995). In discrete-time, the derivation is straightforward and the statistical model is tightly linked to the choice process it is intended to represent.

TWO-STAGE MODELS OF ADOPTION

In line with our earlier discussion, we now model adoption as the outcome of a two-stage process. First, we assume that actors become aware of the innovation if they get exposed to information and that information passes their perceptual threshold. Second, we assume that, conditional upon being aware, actors evaluate the innovation and adopt within the same time period if the innovation meets their minimum utility threshold. Let a_{it} be a binary indicator of whether actor i is aware at time t and let e_{it} be a binary indicator of whether i evaluates the innovation highly (i.e., higher than the threshold). Hence, the probability that i adopts at time t ($y_{it}=1$) is the probability that i is aware at time t ($a_{it}=1$) and, being aware, evaluates the innovation highly enough to adopt ($e_{it}=1 \mid a_{it}=1$). We can write:

$$\begin{aligned} \Pr[y_{it}=1 \mid y_{it-1}=0] &= \Pr[e_{it}=1, a_{it}=1 \mid y_{it-1}=0] \\ &= \Pr[e_{it}=1 \mid a_{it}=1, y_{it-1}=0] \Pr[a_{it}=1 \mid y_{it-1}=0] \end{aligned} \quad [5]$$

One can express the probability that i is aware at time t as a function of the amount of information he is exposed to, say I_{it} . This information exposure can be a function of how much commercial information and word-of-mouth is circulating in general, of i 's media exposure, media habits, perceived source credibility, etc., and of random shocks. Actor i becomes aware of the innovation and its (purported) benefits when his information exposure crosses his perceptual threshold, say π_{it} . Similarly, we express the probability that i chooses to adopt (being aware) as a

function of the innovation's perceived utility to him, say U_{it} . Actor i adopts when that evaluation meets his minimum threshold, say ϕ_{it} . Finally, we assume that I_{it} and U_{it} are random variables additively composed of deterministic elements (V_{it1} and V_{it2}) and stochastic elements ($-\varepsilon_{it1}$ and $-\varepsilon_{it2}$), independently distributed over i and t with cdf F_1 and F_2 , respectively. Collecting all the explanatory variables of the first and second stage in the vectors \mathbf{x}_{it1} and \mathbf{x}_{it2} respectively, such that $V_{it1} - \pi_{it} = \boldsymbol{\alpha}_1 \mathbf{x}_{it1}$ and $V_{it2} - \phi_{it} = \boldsymbol{\alpha}_2 \mathbf{x}_{it2}$, we can write:

$$\Pr[a_{it}=1 \mid y_{it-1}=0] = \Pr[I_{it} \geq \pi_{it}] = \Pr[V_{it1} - \pi_{it} \geq \varepsilon_{it1}] = F_1(\mu_1 \boldsymbol{\alpha}_1 \mathbf{x}_{it1}) \quad [6a]$$

$$\Pr[e_{it}=1 \mid a_{it}=1, y_{it-1}=0] = \Pr[U_{it} \geq \phi_{it}] = \Pr[V_{it2} - \phi_{it} \geq \varepsilon_{it2}] = F_2(\mu_2 \boldsymbol{\alpha}_2 \mathbf{x}_{it2}) \quad [6b]$$

where μ_1 and μ_2 are positive scale parameters related to the variance of ε_{it1} and ε_{it2} , respectively.

For convenience, as before we assume that $\mu_1 = \mu_2 = 1$.

The two-stage models we present below further assume that ε_{it1} and ε_{it2} are independently distributed across stages. This assumption follows from the interpretation of the evaluation process as being conditional on being aware, and does not require the events of the joint process {being aware, evaluating positively} to be independent (cf. Abowd and Farber 1982). This is an important assumption, as one can expect that an actor with very high interest in a particular area is more likely to actively seek out information about innovations in that area, and once knowledgeable about one such innovation and its purported benefits, to evaluate it highly.

We now proceed by presenting three different two-stage models. The first specification is appropriate in the ideal case where one has data on both time of awareness and time of adoption (full observability). The other two specifications can be used in the more common situation where one observes only the time of adoption (partial observability).

Specification 1: Full observability

When both awareness and adoption are observed, and one does not consider post-adoption observations, the data comprise three possible states:

State 0: $a_{it} = 0$, e_{it} not relevant (no awareness and hence no adoption)

State 1: $a_{it} = 1$, $e_{it} = 0$ (awareness, but no positive evaluation; hence no adoption).

State 2: $a_{it} = 1$, $e_{it} = 1$ (both awareness and positive evaluation, hence adoption).

We use P_{0it} , P_{1it} and P_{2it} to denote the probability of i being in each state at time t , conditional upon $y_{it-1} = 0$ ($P_{0it} + P_{1it} + P_{2it} = 1$). The log likelihood will then have the following form (cf. Maddala 1983):

$$LL = \sum_i \sum_t [1 - d_{it}] [(a_{it} e_{it}) \ln P_{2it} + a_{it} (1 - e_{it}) \ln P_{1it} + (1 - a_{it}) \ln P_{0it}] \quad [7]$$

We also have (assuming $\mu_1 = \mu_2 = 1$):

$$\begin{aligned} P_{2it} &= \Pr[e_{it}=1, a_{it}=1 | y_{it-1}=0] = \Pr[a_{it}=1 | y_{it-1}=0] \Pr[e_{it}=1 | a_{it}=1, y_{it-1}=0] \\ &= F_1(\alpha_1 \mathbf{x}_{it1}) F_2(\alpha_2 \mathbf{x}_{it2}) \end{aligned} \quad [8a]$$

$$\begin{aligned} P_{1it} &= \Pr[e_{it}=0, a_{it}=1 | y_{it-1}=0] = \Pr[a_{it}=1 | y_{it-1}=0] \Pr[e_{it}=0 | a_{it}=1, y_{it-1}=0], \\ &= F_1(\alpha_1 \mathbf{x}_{it1}) [1 - F_2(\alpha_2 \mathbf{x}_{it2})], \end{aligned} \quad [8b]$$

$$\begin{aligned} P_{0it} &= \Pr[a_{it}=0 | y_{it-1}=0] \\ &= 1 - F_1(\alpha_1 \mathbf{x}_{it1}) \end{aligned} \quad [8c]$$

Substituting equations 8a-8c into equation 7, we can write the total log likelihood as:

$$\begin{aligned} LL = \sum_i \sum_t [1 - d_{it}] [&(a_{it} e_{it}) \ln \{ F_1(\alpha_1 \mathbf{x}_{it1}) F_2(\alpha_2 \mathbf{x}_{it2}) \} \\ &+ a_{it} (1 - e_{it}) \ln \{ F_1(\alpha_1 \mathbf{x}_{it1}) [1 - F_2(\alpha_2 \mathbf{x}_{it2})] \} \\ &+ (1 - a_{it}) \ln \{ 1 - F_1(\alpha_1 \mathbf{x}_{it1}) \}] \end{aligned} \quad [9]$$

One can estimate parameter vectors α_1 and α_2 for each stage separately (Maddala 1983, p. 50). Vector α_1 can be estimated from a standard hazard model for awareness using all data points, and vector α_2 from a standard hazard model using only those actor-time periods in which actors are aware. Variants of this full observability model have appeared in prior research. Mare's (1981) sequential logit model is, in essence, a full observability multi-stage discrete-time hazard model (Mare 1993, p. 353), and Hogan and Kitagawa (1985), Longini et al. (1989), and Dekimpe, Parker and Sarvary (2000) have estimated continuous-time variants.

Specification 2: Partial observability without memory

In situations of partial observability, one cannot distinguish between states 0 (no adoption because no awareness) and 1 (no adoption because no positive evaluation). As a result, one does not observe a_{it} and e_{it} separately, but only their product $a_{it} \times e_{it} = y_{it}$. We present two models for such data structures: one model assumes that being aware is a zero-order process and that actors, once aware, can immediately become unaware again (i.e., they have no memory); the other model assumes that actors never forget once aware (perfect memory).

When actors have no memory, the probability that actor i adopts at time t is still as specified in equation 8a. The corresponding log likelihood of adoption equals:

$$LL = \sum_i \sum_t [1 - d_{it}] [y_{it} \ln\{ F_1(\alpha_1 x_{it1}) F_2(\alpha_2 x_{it2})\} + (1 - y_{it}) \ln\{1 - F_1(\alpha_1 x_{it1}) F_2(\alpha_2 x_{it2})\}] \quad [10]$$

This model has not been used before to study innovation adoption, but is essentially identical to the one developed for cross-sectional binary data by Abowd and Farber (1982). Their context was a two-sided labor market, with the first event being a prospective employee applying for a position and the other event being the employer deciding to hire the applicant or not. Though the

process is sequential, Abowd and Farber observed only whether or not an employer hired a job seeker. In spite of these differences in research context, Abowd and Farber’s model and ours are identical, apart from their use of cross-sectional data versus our use of panel data on time of adoption.

This partial observability model of innovation adoption without memory has two disadvantages. First, estimation problems may arise if both stages share even a few variables (Mohanty 1998; Schneider 1993), even though, as Abowd and Farber note, the model is theoretically identified when there is at least one non-overlapping variable in \mathbf{x}_{it1} or \mathbf{x}_{it2} . The source of the estimation problem is the perfect symmetry in the way F_1 and F_2 enter the log likelihood function (eq. 10). As a result, the only thing that distinguishes the two stages is their set of covariates. Hence, the discrimination between stages and the empirical identification of the model hinges on the explanatory variables, and variable selection requires particular care. The second disadvantage is the assumption that people become unaware at the end of every time period. The opposite assumption—that actors do not forget once aware—seems more reasonable.

Specification 3: Partial observability with perfect memory

If one assumes that actors never forget once they have become aware of an innovation’s existence, one obtains a very different model structure. The key idea is to keep track of the many ways in which someone can be both aware and positively inclined (“appreciative”) at some time t .⁴ For ease of exposition, we simplify the notation and write $F_m(\boldsymbol{\alpha}_m \mathbf{x}_{itm})$ as $F_m(t)$. For actor i to adopt at $t = 1$, he must have become both aware and appreciative at $t = 1$. Hence,

⁴ The two-stage continuous-time epidemiological model developed by Klein, Klotz and Grever (1984) uses the same approach and, as a result, is similar in structure to our discrete-time model (compare our equation 13 with their equation 2.3). An alternative procedure is the convolution approach used in continuous-time modeling (e.g., Coale and McNeil 1972; Sawhney and Eliashberg 1996).

$$\Pr(T = 1) = F_1(1) F_2(1). \quad [11]$$

An actor i who adopts at $t = 2$, may have got there through two routes: either he became aware at $t = 1$ but became appreciative only at $t = 2$, or he became aware only at $t = 2$ and became appreciative immediately. Hence,

$$\begin{aligned} \Pr(T = 2) &= F_1(1) [1 - F_2(1)] F_2(2) + [1 - F_1(1)] F_1(2) F_2(2) \\ &= F_2(2) \{ F_1(1) [1 - F_2(1)] + [1 - F_1(1)] F_1(2) \}. \end{aligned} \quad [12]$$

In general, we know that someone who adopted at time $T = t$ became appreciative for the first time at time t , but do not know when he became aware. He may have become aware at any time between 1 and t . We can write the general expression:

$$\begin{aligned} \Pr(T = t) &= F_2(t) \{ F_1(1) [1 - F_2(1)] [1 - F_2(2)] [1 - F_2(3)] \dots [1 - F_2(t-1)] + \\ &\quad [1 - F_1(1)] F_1(2) [1 - F_2(2)] [1 - F_2(3)] \dots [1 - F_2(t-1)] + \\ &\quad [1 - F_1(1)] [1 - F_1(2)] F_1(3) [1 - F_2(3)] \dots [1 - F_2(t-1)] + \\ &\quad \dots + \\ &\quad [1 - F_1(1)] [1 - F_1(2)] [1 - F_1(3)] \dots [1 - F_1(t-1)] F_1(t) \} \\ &= F_2(t) \sum_{s \leq t} \{ \prod_{k < s} [1 - F_1(k)] \} F_1(s) \{ \prod_{s \leq q < t} [1 - F_2(q)] \} \end{aligned} \quad [13]$$

By the same rationale, the probability that an actor has not adopted by time t is the sum of the probabilities that he is not aware and the probability that he is aware but not appreciative:

$$\begin{aligned} \Pr(T > t) &= \prod_{p \leq t} [1 - F_1(p)] + \\ &\quad [1 - F_2(t)] \left[\sum_{s \leq t} \{ \prod_{k < s} [1 - F_1(k)] \} F_1(s) \{ \prod_{s \leq q < t} [1 - F_2(q)] \} \right] \end{aligned} \quad [14]$$

The log-likelihood for the partial observability model with perfect memory results from substituting equations 13 and 14 for $\Pr(T = t)$ and $\Pr(T > t)$ in the general expression for the log likelihood for discrete-time hazard models (eq. 1).

An important distinction between the partial observability model without memory and the one with perfect memory is that F_1 and F_2 do not enter symmetrically into the latter's likelihood function. As a result, the perfect memory model may be better able to tease out effect of a variable on awareness and evaluation. Also, assuming perfect memory (i.e., being aware as an absorbing state) is probably a more realistic than assuming the complete absence of memory, and is likely to yield models with better descriptive fit.

ILLUSTRATION: APPLICATION TO THE *MEDICAL INNOVATION* DATA

We have introduced a range of models here, varying in their assumptions and data requirements. The two-stage partial observability models are clearly more complex than the widely used single-stage model. In this section we investigate whether the cost of that complexity can be justified in incremental explanatory power. We use the *Medical Innovation* data for this purpose.

Medical Innovation is a study of the adoption of tetracycline, a broad-spectrum antibiotic, by 125 physicians in four small cities in Illinois between the drug's launch in November 1953 and February 1955. The study is often credited for documenting that innovation diffusion is a social process in which adoption is driven by social contagion (Rogers 1995). The study has more than just historical interest, though. Its data on the diffusion of tetracycline have become "a strategic research site for testing new propositions of how social structure drives contagion" (Burt 1987, p. 1301) and for assessing the performance of new modeling techniques (Marsden and Podolny

1990; Strang and Tuma 1993; Valente 1996). More recently, Van den Bulte and Lilien (2001) have argued that contagion was probably rather weak and that marketing effort by drug manufacturers, especially Lederle which was the first to launch a tetracycline-based product, was likely to be a more important factor driving physicians' adoption. Table 1, based on original reports on the *Medical Innovation* study, shows that physicians indeed considered both marketing efforts (rows 1-3) and colleagues (row 4) to be sources of information and influence, but considered the former much more so than the latter. The table also suggests that, relatively speaking, the effect of marketing efforts was concentrated in the first stage of the adoption process, while the effect of colleagues was concentrated in the later stage. This pattern in physicians' self-reports, combined with Van den Bulte and Lilien's findings suggesting that previous analyses may have confounded contagion with the effect of marketing effort, make the data set especially interesting for our purposes. Specifically, *Medical Innovation* may be a case where the strong effect of marketing swamps the more moderate effect of social contagion in a single-stage model, and where only a two-stage model can detect contagion. In that respect, *Medical Innovation* is a challenging test case for the value of two-stage modeling.

[Table 1 about here]

Substantive Assumptions

We assume that to adopt, physicians must both be aware of the innovation and evaluate it positively. Using insights from social network threshold modeling and random utility modeling, we assume that both awareness and utility thresholds may vary as a function of physician characteristics (Erikson 1998; Granovetter 1978; Hedström 1994; McFadden 1986; Valente 1996). Further, to better separate marketing effects from social contagion, we assume that marketing effort and social influence each affects a separate stage in the adoption process. We

assume that marketing efforts, such as sales calls and advertising, can affect awareness but not evaluation, and that exposure to socially relevant peers who have adopted previously can affect evaluation but not awareness. (More complex models where a stage can be affected by both marketing effort and social influence did not lead to better fit and are not reported here).

We assume that social contagion stems from exposure to others who have already adopted and that it operates over personal relationships. This social influence may consist of information transfer improving a physician's evaluation of the drug's therapeutic merits or may also consist of social normative pressure among cohesive peers. The data do not allow one to distinguish between these two "social cohesion" processes, but we will distinguish them from competitive mimicry of structurally equivalent colleagues who a physician feels compelled to imitate for fear of losing status in the community (cf. Burt 1987). Our inability to uniquely operationalize each social contagion mechanism is not important, as our purpose is not to adjudicate between alternative contagion processes, but rather to assess the performance of two-stage models, regardless of the nature of social contagion.

Data and variables

Since Coleman et al. (1966) provide a detailed description of the population, the sample, and data collection procedures, and the portion of the original data set that we use is publicly available (Burt 1986), we limit our discussion to the variables we used for our own analysis.

We define the first stage as awareness and merge evaluation and adoption in the second stage. Most of the covariates that we enter in the second stage are posited to explain whether and when a physician arrives at a positive evaluation of the drug. However, it is unlikely that the time lag between positive evaluation and actual prescription is zero. Not accounting for this second source of variation in the time lag between awareness and adoption may mask the role of social

network exposure on evaluation. For clarity of exposition, we organize and describe the explanatory variables by awareness, evaluation, and adoption stage, even though we merge the latter two in the analysis.

Awareness.--We include three physician characteristics to account for heterogeneity in physicians' tendency to become aware early. The first is the physician's scientific orientation, an attitudinal measure coded as 1 if the physician agreed with the statement that it is more important for a physician to "keep himself informed of new scientific developments [rather than to] devote more time to his patients," and as 0 otherwise. The second physician characteristic is the number of journals he receives or subscribes to, a measure of media exposure varying across physician but constant over time. We use the logarithm to reflect decreasing returns to scale. The third characteristic is advisor status. This is a sociometric indegree variable, and is computed as the percentage of colleagues in a physician's town who cited him as an advisor on medical practice. High status physicians may feel compelled to systematically scan for innovations and may also be the target of special marketing efforts, both resulting in them becoming aware faster.

For our measure of marketing effort, we use Van den Bulte and Lilien's advertising data. Based on their a priori arguments and results, we include only the marketing efforts of the first entrant and market leader, Lederle. To allow for carry-over effects over time, we use a measure of depreciation-adjusted stock of marketing effort. Let m_t be the amount of advertising in month t (in hundreds of pages), and let δ be the monthly decay rate ($0 \leq \delta \leq 1$). The stock of marketing effort M_t is then defined as:

$$M_t = m_t + (1 - \delta) M_{t-1} = \sum_{\tau=0}^t (1 - \delta)^{t-\tau} m_{\tau} \quad [15]$$

Following Van den Bulte and Lilien, we do not include an interaction effect between marketing effort and the number of journals received. However, we do include an interaction term between

advisor status and marketing effort, to allow for the possibility that Lederle salespeople disproportionately called on leaders in the medical community, in the hope of converting them quickly and then capitalizing on their status to convince less prestigious physicians to start prescribing tetracycline as well.

Evaluation.--We include three physician characteristics to account for heterogeneity in physicians' tendency to evaluate the drug highly. The first two are, again, the physician's scientific orientation and his advisor status. Science-oriented and highly reputed physicians are likely to be both more interested in trying a new drug and be more confident in doing so. We also include professional age, a measure (on a 1-6 scale) of how long ago the physician graduated from medical school. We include both a linear and quadratic term to account for a possible inverse U-shaped relationship between professional age and evaluation: compared to mid-career physicians, older physicians may be more conservative and very inexperienced physicians more risk-averse. To reduce collinearity, we mean-center age before squaring it.

We define the social influence that physician i is subject to at time t as a function of whether other physicians $\{j\}$ have adopted previously (indicated by $\{y_{jt-1}\}$) and how important each physician j is to i (indicated by the social weight w_{ij}). Many authors express the extent of social network exposure (SNE) i is experiencing as a lagged network autocorrelation variable, such that $SNE_{it} = \sum_j w_{ij} y_{jt-1}$ (e.g., Marsden and Podolny 1990; Strang 1991). The actual social contagion, i.e. influence on adoption behavior, is then $\alpha_{SNE} \sum_j w_{ij} y_{jt-1}$, where α_{SNE} is a parameter to be estimated.

Note that this standard operationalization of social network exposure assumes that influence increases linearly with the number or proportion of socially relevant physicians who have already adopted. This may be overly restrictive (cf. Hedström, Sandell and Stern 2000; Myers

1997). To the extent that physicians look for information from their peers, it may not matter how many peers have adopted, as long as at least some of them have. The marginal contagion effect may be quite high for low values of SNE , but very quickly taper off. In contrast, to the extent that physicians look for consensus, the marginal effect may be almost nil for low values of SNE , and increase only after a large proportion of peers has adopted. Both such behaviors can be represented flexibly by specifying (cf. Easingwood, Mahajan and Muller 1982, 1983):

$$Social\ contagion = \alpha_{SNE} [SNE_{it}]^\gamma = \alpha_{SNE} [\sum_j w_{ij} y_{jt-1}]^\gamma \quad [16]$$

where $\gamma > 0$ is a parameter to be estimated. If $\gamma < 1$, the effect of SNE is concave: it first increases more quickly than in the linear case, but then tapers off. If $\gamma > 1$, the effect is convex: it increases slowly, but approaches its maximum value rapidly towards the end (Figure 1).

[Figure 1 about here]

We use the two exposure variables computed and published by Burt (1986). Each variable assumes a different influence mechanism represented by the w_{ij} weights. The direct ties weights reflect whether i nominated j as an interaction partner for advice or discussions, such that i might have gained information or experience social normative pressure from j . The structural equivalence weights indicate whether i might mimic j out of fear of losing out in the competition for status. We impute missing variables as described by Van den Bulte and Lilien (2001). The weights are normalized such that $SNE_{it} = \sum_j w_{ij} y_{jt-1}$ is bounded between 0 and 1. Note that for any value of γ , $[SNE_{it}]^\gamma$ will be similarly bounded.⁵

We also include an interaction between advisor status and social network exposure. Coleman et al. (1966, p. 102) report that the empirical hazard rate of physicians receiving nominations as

⁵ Further generalizing the social network exposure variable by allowing the “contagiousness” of adopters to decay over time, as in Kalish and Lilien (1986), Urban et al. (1990), and Myers (2000), did not lead to improvements in fit.

discussion partner or advisor increased over time while that of non-nominated physicians did not rise, which they interpret as evidence of contagion operating through social networks. However, elsewhere (p. 109) they report another set of empirical hazard rates, broken down by both sociometric status and prescription volume, which show that each group's hazard rate increased but at different times. As they state (p. 126), such a difference between the nominated and non-nominated physicians suggests a somewhat different result than their earlier, and more famous, interpretation: "not that the isolated [i.e., non-nominated] were unaffected by social networks, but rather that the effect upon them came considerably later. ... A contagion process has occurred among the relatively isolated doctors as well as among the integrated [i.e., nominated] ones, although it was of much smaller magnitude and started later." Both interpretations by Coleman et al. suggest a positive interaction between sociometric status and social network exposure. However, it is possible that the increase in empirical hazards that Coleman et al. observed among nominated physicians in the first eight months was not due to contagion but to marketing efforts, and that only the increase in hazards late in the diffusion process, when marketing efforts had flattened out, actually reflected contagion. The behavioral explanation is that physicians with high status are likely to be more self-confident and hence less sensitive to their colleagues' behavior. High status physicians would then either adopt rapidly, suggesting a main effect posited earlier, or not at all, even in later periods when most of their peers have adopted. Such behavior would suggest a negative interaction between advisor status and social network exposure.

Trial stage. -- An obvious and important factor contributing to a lag between evaluation and prescription is the mere need to actually prescribe the drug. The more patients a physician sees who may benefit from the drug, the more likely the physician will adopt rapidly. This effect is

borne out by the original analysis by Coleman, Katz, and Menzel (1966), who reported that a physician's volume of prescriptions of drugs similar to tetracycline was the single most powerful predictor of the timing of his first use of tetracycline. Unfortunately, prescription volume is not included in the publicly available data set (Burt 1986), and we are not able to directly control for variation in prescription volume among physicians. However, a dummy variable indicating whether a physician had a chief or honorary position in his hospital, can be viewed as an indirect control, since such physicians were likely to be less involved in actual medical practice than active or regular staff. To capture seasonal variations in prescription volume, we include a dummy variable for the summer months of July and August when the weather is warmer and schools are closed, limiting the spread of contagious diseases.

After constructing the variables, we deleted four physicians, due to missing covariates. The data set for estimation contains 17 monthly observations for 121 individuals, 105 of whom had adopted by the last observation period. Table 2 presents descriptive statistics for the data, after excluding post-adoption observations irrelevant to explain adoption.

[Table 2 about here]

Estimation

We do not estimate the full observability model because the *Medical Innovation* data set does not contain awareness data of sufficient quality. Of the 109 physicians for whom the data of adoption has been established, only 44 ventured an estimate of the two-month period in which they first heard of tetracycline. In addition, for 12 of these physicians (27%), the recalled date of awareness fell after the observed date of adoption, indicating a lack of reliability of the awareness data.

We estimate the models by maximizing the likelihood, using a quasi-Newton algorithm. For the partial observability models, we try multiple sets of starting values to minimize the risk of converging to a local maximum. We test for statistical significance using likelihood ratio tests, which are more powerful and better behaved than Wald tests based on the asymptotic standard errors (Agresti 1990, p. 89). Our two-stage models do not allow for unobserved heterogeneity. This limitation is unlikely to affect our findings, as three different tests for unobserved heterogeneity in single-stage models performed by Van den Bulte and Lilien (2001) provided negative results.

Because the single-stage and the two partial observability models are not nested in each other, we cannot use likelihood ratio tests to assess fit. Instead, we use the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both are based on the log likelihood, but penalize models with a larger number of parameters. Specifically, using n to denote the number of observations and p to denote the number of parameters in the model, the AIC and BIC are defined as (e.g., Agresti 1990):

$$\begin{aligned} \text{AIC} &= -2 \text{ LL} + 2p \\ \text{BIC} &= -2 \text{ LL} + (\log n)p \end{aligned}$$

As is clear from these formulas, the BIC penalizes complex models more severely than the AIC does. As recommended by Raftery (1995) and following Xie (1994) and Yamaguchi (1998), we use the number of events (105 adoptions) as n when calculating the BIC. We estimated the three models using the logit, probit and complementary log-log specifications. The last specification fit slightly better, but otherwise the results were consistent across specifications. We present results for the complementary log-log specification only.

Results

Table 3 presents the results for the social contagion operating via direct ties. The first key result is that, based on the log likelihoods and the AIC, the two-stage model with perfect memory fits better than the two-stage model with zero memory, which in turn fits better than the single-stage model. Using the BIC as criterion, only the two-stage model with perfect memory outperforms the simpler single-stage model. The difference in BIC is about 3, which Raftery (1995) suggests is large enough to be evidence of superior fit. To further assess descriptive fit, we omit the effects that are non-significant in all three models: advisor status and its interaction with marketing effort in the first stage, and science orientation in the second stage. The two-stage models then have the same number of covariates as the single-stage model. The $-2LL$ values for single stage, the two-stage model without memory and two-stage model with perfect memory are now 602.87, 593.89 and 587.54, respectively. The BIC values are 663.37, 659.05 and 652.70. The BIC indicates that both two-stage models now fit better than the single-stage model, and the 10-point difference between the first and third model provides strong to very strong evidence that the two-stage model with perfect memory is superior to the standard single-stage model (Raftery 1995). In short, the first key finding is that the theoretically most appealing partial observability model fits better than the standard model does.

[Table 3 about here]

The second key finding concerns social contagion. All three models find evidence of marketing effects, as expected based on the situational analysis reported by Van den Bulte and Lilien (2001), but vary in their conclusions about social contagion. The standard model finds no evidence of positive contagion. Interestingly, the results suggest that high-status physicians were less susceptible to social contagion than were non-prestigious physicians. This finding runs

counter to the results reported by Coleman et al. (1966), but could be anticipated if higher-status physicians have more confidence in their own judgments. The two-stage model without memory finds some evidence of contagion, but the effect is only significant at the 10% level. The theoretically more appealing model with perfect memory finds evidence of contagion significant at 5%. In short, as we use model specifications that better reflect the theory of the adoption process with mass media and interpersonal influence operating at different stages, we are able to document contagion effects that a single-stage model is unable to detect.

Apart from these important differences in model fit and conclusions about contagion, the three model specifications are consistent in their findings about the effect of the other covariates. A few differences are of interest, though. The first is in the decay rate of marketing efforts: 0.40 in the two-stage model with perfect memory, but only about 0.25 in the other two models. This is not surprising: a specification that assumes that people do not forget—i.e. that they carry over their awareness into future periods—can allow for a stronger decay rate of marketing efforts before people become aware. Another important difference is the effect size of being a chief or administrator. The effect is negative, and significant, in all three models, but very large (in an absolute sense) only in the two-stage models. The latter finding suggests that, keeping other factors constant, physicians in leading, honorary or administrative positions were much less likely to prescribe the drug. This large effect is consistent with the finding by Coleman et al. (1966) that a physician's volume of prescriptions of drugs similar to tetracycline was the single best predictor of time of adoption.

Three other results are worth highlighting. The two-stage model estimates suggest that the same variable can have different effects in the awareness and evaluation stages (as with media coverage in the *Monica* anecdote). Scientific orientation is positively associated with awareness,

but not with evaluation. The reverse is true for advisor status: it is associated with arriving at a positive evaluation quickly, but is unrelated to when a physician became aware of the drug. Finally, the γ estimates are all well above 1, indicating that the response to social network exposure is strongly convex. Physicians tended to respond to their peers' adoptions, but not before more than 80% of their contacts had done so (compare Figure 1).⁶

The results for models with social contagion from structurally equivalent peers are similar (Table 4). The two-stage model with perfect memory fits best, the single-stage model does not find evidence of contagion while the two-stage models do, and the other similarities and disparities in Table 3 discussed earlier are also present. Two differences between the tables are worth highlighting. First, the γ estimates are lower in Table 4, and not significantly different from 1. If contagion between structurally equivalent actors is driven by competitive concerns, as commonly accepted, then this result is to be expected. When contagion is a reaction to competitive threats, then actors should respond sooner, i.e. at lower levels of adoption by peers,

⁶ In the empirical analyses reported here, we allowed advertising to affect only awareness and social network exposure to affect only evaluation. Because tetracycline was very aggressively marketed and received wide media coverage, as noted by Coleman et al. (1966) and described by Van den Bulte and Lilien (2001), we believe that contagion is unlikely to have been important in creating awareness. Furthermore, given tetracycline's product advantages, simplicity of use, and low risk, we do not believe that advertising and other marketing efforts were necessary to convince physicians to give the new drug a try. Yet, we recognize that assuming advertising to operate exclusively at the awareness stage and contagion exclusively at the evaluation stage is a stronger assumption than what theory and prior research suggests, namely that advertising operates mostly at the awareness stage and contagion mostly at the evaluation stage. We therefore also estimated models with social network exposure entering the awareness stage and marketing effort entering the evaluation stage. These model extensions did not improve model fit (the value of -2LL decreased by less than 1) and did not affect the other parameter estimates much. Hence, for this particular data set, our assumptions are not only reasonable but also statistically validated.

than if contagion is a response to maintain attitudinal balance. Second, models with contagion operating over direct ties fit marginally better than do models with contagion operating between structural equivalents, and this small difference holds across all three process specifications.

[Table 4 about here]

CONCLUSION

In this paper, we address the gap between theory positing distinct stages in the adoption process and data recording only the final outcome of the process, i.e. the time of adoption. We bridge this gap in richness between theory and data by developing new models, which we call partial observability models of innovation adoption. These event history models incorporate the theoretical assumption of adoption being a two-stage process, but can be estimated without using data on awareness and only using traditional data about the time of adoption. These models enable one to more sharply operationalize theoretical propositions about the effects of mass media and social networks, and to empirically differentiate their effects on awareness and evaluation.

An application to the classic *Medical Innovation* data illustrates that these finer-grained models can not only provide better descriptive fit but, more importantly, can also substantially affect research conclusions. Specifically, while a single-stage model incorporating marketing effort does not produce statistically significant estimates of social contagion, a theoretically more refined two-stage model with perfect memory in which marketing effort affects only awareness and social contagion operates at the evaluation stage finds significant effects for both marketing effort and social network exposure. This result is consistent with theory as well as with answers by physicians to direct inquiries about sources of influence presented in original reports of the *Medical Innovation* study. Yet, this result had not been obtained in previous empirical analyses

using single-stage event history models of adoption. As Besharov and Greif (2001, p. 263) note, “models are exercises in selective blindness” and even though part of normal science consists in making finer distinctions, “at some point, the fineness of the distinction is not worth the additional complexity.” The application to *Medical Innovation*, though only illustrative, clearly suggests that the insights one can gain from two-stage partial observability models may indeed be worth the additional complexity.

Partial observability models may be methodologically sound and theoretically valuable tools in several research areas other than innovation adoption. As Yamaguchi (1991, p. 162) notes, two-step analysis is generally recommended “if (a) there is a logically preceding event before which the event of interest is zero, and (b) the timing of the logically preceding event is known for each subject.” The second condition, however, is not necessary with partial observability models. This, we believe, offers some interesting research opportunities. For instance, in their discussions of neo-institutional theory and organizational decision-making processes, Tolbert and Zucker (1996) and Roberts and Greenwood (1997) distinguish between a search/awareness stage and an evaluation stage. Neo-institutional theory suggests that institutionalization is a cognitive process at the former stage, and a normative or sociopolitical process at the latter stage. Hence, models that differentiate between these two stages may allow empirical studies to more precisely pinpoint the nature of mimesis and institutionalization in organizational fields. Life course research is another area where partial observability models may be useful tools to study multi-stage processes. The analysis of premarital pregnancy is a good example (Hogan and Kitagawa 1985; Yamaguchi 1991). The risk of becoming pregnant before marriage is zero before the occurrence of puberty or the first premarital intercourse. Moreover, certain covariates, such as juvenile delinquency, may affect the time of premarital sexual intercourse (stage 1) but not the

time of premarital pregnancy (stage 2). The reverse may hold for other covariates, such as knowledge about and availability of contraceptives. With such differences in the effect of covariates, it is best for one's analysis to reflect the sequenced nature of the process leading to premarital intercourse (Yamaguchi 1991).

Partial observability models do have an important limitation: to make up for the lack of data, they require stronger a priori assumptions than do full observability models. Partial observability models allow researchers to incorporate the theoretical concept of a sequenced process into their analysis even when their dependent variables data are only about the final outcome. But this benefit comes at a cost. The onus of identifying the two stages is shifted from (a) collecting data recording when actors become aware to (b) having a rich set of covariates and sufficient confidence—based on a priori theory, cumulative research evidence, and one's own understanding of the actors' situation—to use these covariates to reflect the sequential nature of the adoption process. As a result, partial observability models may be appropriate only for well-developed areas of inquiry. The upside, of course, is that two-stage models force empirical researchers to think harder about the causal mechanisms driving the process they study, and to better represent them in models amenable to empirical validation. In that respect, two-stage modeling helps narrow the gap between “substantive or theoretical models” and “methodological or statistical models,” as called for by Goldthorpe (2000), Skovertz (1998) and Sørensen (1998, 1999).

Even though the partial observability models allow one to relax important behavioral assumptions about the adoption process, the models we have presented are rather simple from a statistical point of view. Allowing for unobserved heterogeneity across actors' thresholds would be an important advance, especially for use with large data sets with relatively few covariates.

Combining both partial observability models into a discrete mixture model allowing some adopters to have no memory and others to have perfect memory would be another possible enrichment.

As Merton (1968, p. 73) noted, research rests on a “triple alliance between theory, method and data.” Like most streams of active research, the study of innovation adoption and social contagion has had to confront not only gaps between theory and data, which we focused on here, but also gaps between theory and models and even gaps between models and data. Two-stage models with partial observability contribute to the continuing quest to more tightly link theory, data, and models in the study of innovation adoption and social contagion.

Table 1. Doctors' self-reported sources of information and influence when adopting tetracycline ^a

	Percentage of physicians				
	Crediting Source with		Mentioning Source of Information as		
	Original influence	Most influence	First	Intermediate	Final
Detail men (salespeople)	57	38	52	27	5
Direct mail	18	8	22	16	14
Drug house periodicals	4	5	3	11	21
Colleagues	7	20	10	15	28
Journal articles	7	23	6	21	21
Professional meetings	3	...	3	4	8
All other media	4	6	3	7	3

^a Source: Van den Bulte and Lilien (2001, p. 1417), based on Katz (1961, p. 77) and Coleman et al. (1966, p. 59).

Table 2. Descriptive statistics for *Medical Innovation data*^a

	Mean	SD	Min	Max	Correlations															
					1	2	3	4	5	6	7	8	9	10						
1. y (adoption indicator)	.111	.314	0	1	...															
2. Number of journals (log)	1.484	.392	.69	2.20	.10															
3. Science orientation	.268	.443	0	1	.12	-.02														
4. Advisor status (indegree)	2.780	6.052	0	45.45	.04	.45	.07													
5. Summer	.084	.278	0	1	-.05	-.07	-.08	-.06												
6. Age	.000	1.706	-2.54	2.46	-.09	-.16	-.01	.02	.06											
7. Age ²	2.907	2.561	.21	6.48	-.08	-.03	.01	-.03	.05	.17										
8. Chief	.098	.298	0	1	-.05	.06	.29	.08	.00	.31	.26									
9. Marketing effort ($\delta = .25$)	.155	.087	0	0.24	.05	-.17	-.19	-.09	.30	.14	.13	.05								
10. SNE (Direct ties) ^b	.332	.322	0	1	.07	-.04	-.13	.00	.18	.15	.09	.09	.68							
11. SNE (Structural equivalence)	.451	.419	0	1	.02	-.15	-.15	-.10	.27	.18	.10	.15	.75	.72						

^a $N = 947$ (I.e., the unit of observation is the physician-month).

^b SNE stands for social network exposure.

Table 3. Results for models with social contagion from direct ties

	Single-Stage Model	Two-Stage Models	
		Zero Memory	Perfect Memory
Intercept	-4.14****	-4.12***	-3.47****
Number of journals (log)	0.86***	0.73**	0.68**
Science orientation	1.05****	1.15****	0.89***
Marketing effort	3.76***	4.77****	3.22***
Decay rate (δ)	0.26	0.22**	0.40***
Advisor status (indegree)	-0.06	-0.04	-0.10
Advisor status x Marketing effort	0.42	0.07	0.58
Intercept	...	2.52	-0.57
Summer	-0.77*	-2.95**	-1.62**
Age	-0.13*	-0.64**	-0.47**
Age ²	-0.11**	-0.58*	-0.38
Chief	-0.90**	-10.90**	-8.95***
Science orientation	...	-1.02	0.60
<i>SNE</i> (Direct ties) ^a	1.19	4.95*	2.98**
γ	6.26**	12.32 ^b	9.05**
Advisor status (indegree)	...	3.71***	3.25****
Advisor status x <i>SNE</i>	-0.13**	-0.05	0.72
-2LL	600.34	591.09	583.33
df	14	17	17
AIC	628.34	625.09	617.33
BIC	665.50	670.21	662.45

Note.—Results are from complementary log-log models. The significance levels reported are for likelihood ratio tests that the parameter of interest is zero, except for tests of γ , where the test is $\gamma = 1$.

^a *SNE* stands for social network exposure

^b Nested model with $\gamma = 1$ does not converge.

* $P < .10$

** $P < .05$

*** $P < .01$

**** $P < .001$

Table 4. Results for models with social contagion from structurally equivalents

	Single-Stage Model	Two-Stage Models	
		Zero Memory	Perfect Memory
Intercept	-4.13****	-3.49****	-3.49****
Number of journals (log)	0.87***	0.59**	0.68**
Science orientation	1.08****	0.88***	0.97****
Marketing effort	3.30***	4.56****	2.86***
Decay rate (δ)	0.28***	0.31**	0.40**
Advisor status (indegree)	-0.07	-0.07**	-0.10
Advisor status x Marketing effort	0.62*	0.26	0.61
Intercept	...	0.07	-0.37
Summer	-0.83*	-2.67**	-2.00***
Age	-0.11	-0.37**	-0.41**
Age ²	-0.10**	-0.35***	-0.49***
Chief	-0.90**	-5.44***	-6.56***
Science orientation	...	-0.54	0.40
<i>SNE</i> (Structural equivalence) ^a	0.51	1.83**	1.58**
γ	1.79	3.33 ^a	3.05
Advisor status (indegree)	...	2.34***	2.77****
Advisor status x <i>SNE</i>	-0.10**	0.07	0.30
-2LL	603.56	591.82	585.01
df	14	17	17
AIC	631.56	625.82	619.01
BIC	668.72	670.94	664.13

Note.—Results are from complementary log-log models. The significance levels reported are for likelihood ratio tests that the parameter of interest is zero, except for tests of γ , where the test is $\gamma = 1$.

^a *SNE* stands for social network exposure

^b Nested model with $\gamma = 1$ does not converge.

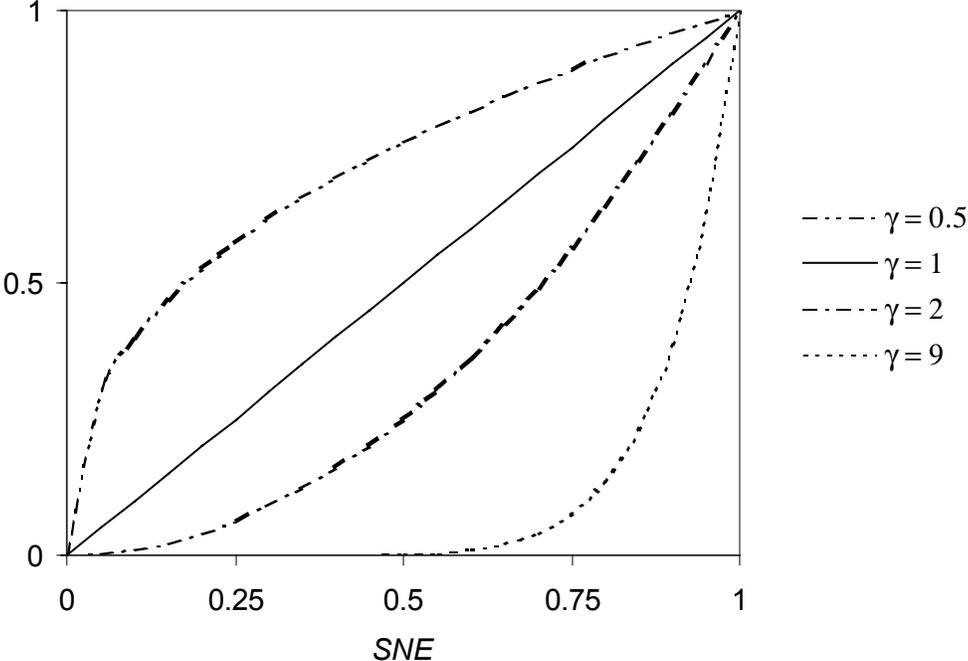
* $P < .10$

** $P < .05$

*** $P < .01$

**** $P < .001$

Figure 1. How SNE^γ varies as a function of γ



References

- Abowd, John M., and Henry S. Farber. 1982. "Job Queues and the Union Status of Workers." *Industrial and Labor Relations Review* 35: 354-67.
- Agresti, Alan (1990), *Categorical Data Analysis*. New York: John Wiley.
- Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories," Pp. 61-98 in *Sociological Methodology 1982*, edited by Samuel Leinhardt. San Francisco, CA: Jossey-Bass.
- Bandura, Albert. 1977. *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain. 1994. "Why the Bass Model Fits without Decision Variables." *Marketing Science* 13: 203-23.
- Besharov, Gregory and Avner Greif. 2001. "Comments and Further Thoughts on 'Organizational Genesis, Identity and Control: The Transformation of Banking in Renaissance Florence,'" Pp. 258-69 in *Networks and Markets*, edited by James E. Rauch and Alessandra Casella. New York: Russell Sage Foundation.
- Bonus, Holger. 1973. "Quasi-Engel Curves, Diffusion, and the Ownership of Major Consumer Durables." *Journal of Political Economy* 81: 655-77.
- Burt, Ronald S. 1986. "The *Medical Innovation* Network Data." Technical Report No. 3, Center for the Social Sciences, Columbia University.
- 1987. "Social Contagion and Innovation: Cohesion versus Structural Equivalence." *American Journal of Sociology* 92: 1287-335.
- Coale, A.J. and D.R. McNeil. 1972. "The Distribution by Age of the Frequency of First Marriage in a Female Cohort." *Journal of American Statistical Association* 67: 743-9.
- Coleman, James S., Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis, IN: The Bobbs-Merrill Company.
- Dekimpe, Marnik G., Philip M. Parker, and Miklos Sarvary. 2000. "Global Diffusion of Technological Innovations: A Coupled-Hazard Approach." *Journal of Marketing Research* 37: 47-59.
- Dodson, Joe A., Jr., and Eitan Muller. 1978. "Models of New Product Diffusion through Advertising and Word-of-Mouth." *Management Science* 24: 1568-78.

- Easingwood, Christopher J., Vijay Mahajan and Eitan Muller. 1981. "A Nonsymmetric Responding Logistic Model for technological Substitution." *Technological Forecasting and Social Change* 20: 199-213.
- , 1983. "A Nonuniform Influence Innovation Diffusion Model of New Product Acceptance." *Marketing Science* 2: 273-96.
- Erikson, Robert. 1998. "Thresholds and Mechanisms: A Comment on Hedström and Swedberg's Chapter," pp. 88-93 in *Rational Choice Theory and Large-Scale Data Analysis*, edited by Hans-Peter Blossfeld and Gerald Prein, Boulder, CO: Westview Press.
- Freedman, Ronald, and John Y. Takeshita. 1969. *Family Planning in Taiwan: An Experiment in Social Change*. Princeton, NJ: Princeton University Press.
- Gatignon, Hubert, and Thomas S. Robertson. 1986. "Integration of Consumer Diffusion Theory and Diffusion Models: New Research Directions," Pp. 37-59 in *Innovation Diffusion Models of New Product Acceptance*, edited by Vijay Mahajan and Yoram Wind. Cambridge, MA: Ballinger Publishing Company.
- Golder, Peter N. 2000. "Historical Method in Marketing Research with New Evidence on Long-Term Market Share Stability." *Journal of Marketing Research* 37: 156-72.
- Goldthorpe, John H. 2000. *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.
- Granovetter, Mark S. 1978. "Threshold Models of Collective Action." *American Journal of Sociology* 83: 1420-43.
- Granovetter, Mark S., and Roland Soong. 1983. "Threshold Models of Diffusion and Collective Behavior." *Journal of Mathematical Sociology* 9: 165-79.
- Greene, William H. 2000. *Econometric Analysis, 4th ed.* Upper Saddle River, NJ: Prentice-Hall.
- Hauser, John R., and Glen L. Urban. 1977. "A Normative Methodology for Modeling Consumer Response to Innovation." *Operations Research* 25: 579-619.
- Hedström, Peter. 1994. "Contagious Collectivities: On the Spatial Diffusion of Swedish Trade Unions, 1890-1940." *American Journal of Sociology* 99: 1157-79.
- Hedström, Peter, Rickard Sandell, and Charlotta Stern. 2000. "Mesolevel Networks and the Diffusion of Social Movements: The Case of the Swedish Social Democratic Party." *American Journal of Sociology* 106: 145-72.
- Hogan, Dennis P. and Evelyn M. Kitagawa. 1985. "The Impact of Social Status, Family Structure, and Neighborhood on the Fertility of Black Adolescents." *American Journal of Sociology* 90: 825-55.

- Horsky, Dan, and Leonard S. Simon. 1983. "Advertising and the Diffusion of New Products." *Marketing Science* 2: 1-17.
- Jedidi, Kamel, Jehoshua Eliashberg, and Wayne S. DeSarbo. 1989. "Optimal Advertising and Pricing for a Three-Stage Time-Lagged Monopolistic Diffusion Model Incorporating Income." *Optimal Control Applications & Methods* 10: 313-31.
- Kalish, Shlomo. 1985. "A New Product Adoption Model with Price, Advertising, and Uncertainty." *Management Science* 31: 1569-85.
- Kalish, Shlomo, and Gary L. Lilien. 1986. "Applications of Innovation Diffusion Models in Marketing," Pp. 235-79 in *Innovation Diffusion Models of New Product Acceptance*, edited by Vijay Mahajan and Yoram Wind. Cambridge, MA: Ballinger.
- Katz, Elihu. 1961. "The Social Itinerary of Technical Change: Two Studies on the Diffusion of Innovation." *Human Organization* 20 (Summer): 70-82.
- Klein, John P., Jerome H. Klotz and Michael R. Grever. 1984. "A Biological Marker Model for Predicting Disease Transitions." *Biometrics* 40: 927-36.
- Lieberson, Stanley. 2000. *A Matter of Taste: How Names, Fashions, and Culture Change*. New Haven, CT: Yale University Press.
- Lin, Nan. 1971. "Information Flow, Influence Flow and the Decision-Making Process." *Journalism Quarterly* 48: 33-40, 61.
- Lin, Nan, and Ronald S. Burt. 1975. "Differential Effects of Information Channels in the Process of Innovation Diffusion." *Social Forces* 54: 256-74.
- Longini, Ira M. Jr., W. Scott Clark, Michael Haber and Robert Hosburgh, Jr. 1989. "The Stages of HIV Infection: Waiting Times and Infection Transmission Probabilities," Pp. 111-37 in *Mathematical and Statistical Approaches to AIDS Epidemiology (Lecture Notes in Biomathematics, Vol. 83)*, edited by C. Castillo-Chavez. Berlin: Springer-Verlag.
- Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Mahajan, Vijay, Eitan Muller, and Roger A. Kerin. 1984. "Introduction Strategy for New Products with Positive and Negative Word-of-Mouth." *Management Science* 30: 1389-404.
- Mare, Robert D. 1981. "Change and Stability in Educational Stratification." *American Sociological Review* 46: 72-87.
- 1993. "Educational Stratification on Observed and Unobserved Components of Family Background," Pp. 351-76 in *Persistent Inequality*, edited by Yossi Shavit and Hans-Peter Blossfeld. Boulder, CO: Westview.

- Marsden, Peter V., and Joel Podolny. 1990. "Dynamic Analysis of Network Diffusion Processes," Pp. 197-214 in *Social Networks through Time*, edited by Jeroen Weesie and Henk Flap. Utrecht: ISOR/Rijksuniversiteit Utrecht.
- McFadden, Daniel. 1986. "The Choice Theory Approach to Marketing Research." *Marketing Science* 5: 275-97.
- Mesak, Hani I. 1996. "Incorporating Price, Advertising and Distribution in Diffusion Models of Innovation: Some Theoretical and Empirical Results." *Computers and Operations Research* 23: 1007-23.
- Mohanty, Madhu S. 1998. "The Role of the Desire for Union Status in the Decision to Enter Local Government Job Queues: The US Evidence." *Applied Economics* 30: 1369-78.
- Muller, Eitan. 1983. "Trial/Awareness Advertising Decisions: A Control Problem with Phase Diagrams with Non-Stationary Boundaries." *Journal of Economic Dynamics & Control* 6: 333-50.
- Myers, Daniel J. 1997. "Racial Rioting in the 1960s: An Event History Analysis of Local Conditions." *American Sociological Review* 62: 94-112.
- . 2000. "The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks." *American Journal of Sociology* 106: 173-208.
- Pareek, Udai and Y.P. Singh. 1969. "Communication Nets in the Sequential Adoption Process." *Indian Journal of Psychology* 46: 33-55.
- Peay, Marilyn Y., and Edmund R. Peay. 1984. "Differences among Practitioners in Patterns of Preference for Information Sources in the Adoption of New Drugs." *Social Science in Medicine* 18: 1019-25.
- Pelz, Donald C. 1983. "Use of Information Channels in Urban Innovations." *Knowledge: Creation, Diffusion, Utilization*. 5: 3-25.
- Prentice, R. L., and L. A. Gloeckler. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics* 34: 57-67.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." Pp. 111-63 in *Sociological Methodology 1995*, edited by Peter V. Marsden. Oxford: Blackwell Publishers.
- Roberts, Peter W., and Royston Greenwood. 1997. "Integrating Transaction Cost and Institutional Theories: Toward a Constrained-Efficiency Framework for Understanding Organizational Design Adoption." *Academy of Management Review* 22: 346-73.

- Robertson, Thomas S. 1971. *Innovative Behavior and Communication*. New York: Holt, Rinehart, and Winston.
- Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: Free Press.
- 1995. *Diffusion of Innovations, 4th ed.* New York: Free Press.
- Saloner, Garth, and Andrea Shepard. 1995. "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines." *Rand Journal of Economics* 26: 479-501.
- Sawhney, Mohanbir S., and Jehoshua Eliashberg. 1996. "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures." *Marketing Science* 15: 113-31.
- Schneider, H. 1993. "Binary Discrete Choice under Asymmetric Conditions." *Statistical Papers* 34 (March), 43-57.
- Silver, Steven D. 1984. "A Simple Mathematical Theory of Innovative Behavior: Comment." *Journal of Consumer Research* 10: 441-4.
- Skvoretz, John. 1998. "Theoretical Models: Sociology's Missing Link," Pp. 238-52 in *What is Social Theory?* edited by Alan Sica. Oxford: Blackwell.
- Simon, Hermann, and Karl-Heinz Sebastian. 1987. "Diffusion and Advertising: The German Telephone Campaign." *Management Science* 33: 451-66.
- Snyder, Leslie B. 1991. "Modeling Dynamic Communication Processes With Event History Analysis." *Communication Research* 18:464-86.
- Sørensen, Aage B. 1998. "Theoretical Mechanisms and the Empirical Study of Social Processes." Pp. 172-203 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by Peter Hedström and Richard Swedberg. Cambridge: Cambridge University Press.
- 1999. "Statistical Models and Mechanisms of Social Processes." Unpublished draft, Department of Sociology, Harvard University.
- Strang, David. 1991. "Adding Social Structure to Diffusion Models: An Event History Framework." *Sociological Methods & Research* 19: 324-53.
- Strang, David, and Sarah A. Soule. 1998. "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills," Pp. 265-90 in *Annual Review of Sociology*, edited by John Hagan and Karen S. Cook. Palo Alto, CA: Annual Reviews.
- Strang, David, and Nancy Brandon Tuma. 1993. "Spatial and Temporal Heterogeneity in Diffusion." *American Journal of Sociology* 99: 614-39.

- Tapiero, Charles S. 1983. "Stochastic Diffusion Models with Advertising and Word-of-Mouth Effects." *European Journal of Operational Research* 12: 348-56.
- Tolbert, Pamela S., and Lynne G. Zucker. 1996. "The Institutionalization of Institutional Theory," Pp. 175-90 in *Handbook of Organization Studies*, edited by Stewart R. Clegg, Cynthia Hardy, and Walter R. Nord. London: Sage.
- Urban, Glen L., John R. Hauser, and John H. Roberts. 1990. "Prelaunch Forecasting of New Automobiles." *Management Science* 36: 401-21.
- Valente, Thomas W. 1995. *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press.
- 1996. "Social Network Thresholds in the Diffusion of Innovations." *Social Networks* 18: 69-89.
- Valente, Thomas W. and Walter P. Saba. 1998. "Mass Media and Interpersonal Influence in a Reproductive Health Communication Campaign in Bolivia." *Communications Research*, 25: 96-124.
- Van den Bulte, Christophe and Gary L. Lilien. 1997. "Bias and Systematic Change in the Parameter Estimates of Macro-level Diffusion Models." *Marketing Science* 16: 338-53.
- 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology* 106: 1409-35.
- Weenig, Mienieke W.H., and Cees J.H. Midden. 1991. "Communication Network Influences on Information Diffusion and Persuasion." *Journal of Personality and Social Psychology* 61: 734-42.
- Xie, Yu. 1994. "The Log-Multiplicative Models for Discrete-Time, Discrete-Covariate Event-History Data," Pp. 301-40 in *Sociological Methodology 1994*, edited by Peter V. Marsden. Oxford: Blackwell Publishers.
- Yamaguchi, Kazuo. 1991. *Event History Analysis*. Newbury Park, CA: Sage.
- 1998. "Mover-Stayer Models for Analyzing Event Nonoccurrence and Event Timing with Time-Dependent Covariates: An Application to an Analysis of Remarriage," Pp. 327-61 in *Sociological Methodology 1998*, edited by Adrian E. Raftery. Oxford: Blackwell Publishers.
- Zaltman, Gerald and Ronald Stiff. 1973. "Theories of Diffusion," Pp. 416-68 in *Consumer Behavior: Theoretical Sources*, edited by Scott Ward and Thomas S. Robertson. Englewood Cliffs: Prentice-Hall.