MANAGEMENT SCIENCE
Vol. 52, No. 8, August 2006, pp. 1258–1272
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5208 | 1258
DOI 10.1287/mnsc.1060.0517

Modeling the "Pseudodeductible" in Insurance Claims Decisions

Michael Braun
Sloan School of Management, MIT, Cambridge, Massachusetts 02139, braunm@mit.edu

Peter S. Fader, Eric T. Bradlow, Howard Kunreuther
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
{faderp@wharton.upenn.edu, ebradlow@wharton.upenn.edu, kunreuther@wharton.upenn.edu}

In many different managerial contexts, consumers "leave money on the table" by, for example, their failure to claim rebates, use available coupons, and so on. This project focuses on a related problem faced by homeowners who may be reluctant to file insurance claims despite the fact their losses are covered. We model this consumer decision by introducing the concept of the "pseudodeductible," a latent threshold above the policy deductible that governs the homeowner's claim behavior. In addition, we show how the observed number of claims can be modeled as the output of three stochastic processes that are separately, and in conjunction, managerially relevant: the rate at which losses occur, the size of each loss, and the choice of the individual to file or not file a claim. By allowing for the possibility of pseudodeductibles, one can sort out (and make accurate inferences about) these three processes.

We test this model using a proprietary data set provided by State Farm, the largest underwriter of personal lines insurance in the United States. Using mixtures of Dirichlet processes to capture heterogeneity and the interplay among the three processes, we uncover several relevant "stories" that underlie the frequency and severity of claims. For instance, some customers have a small number of losses, but all are filed as claims, whereas others may experience many more losses, but are more selective about which claims they file. These stories explain several observed phenomena regarding the claims decisions that insurance customers make, and have broad implications for customer lifetime value and market segmentation.

Key words: duration models; Dirichlet process priors; insurance claims; semiparametric Bayesian statistics; underreporting
History: Accepted by David E. Bell, decision analysis; received June 9, 2005. This paper was with the authors 1 month for 1 revision.

1. Introduction

In many different managerial contexts, customers may "leave money on the table" by, for example, their failure to claim rebates, use available coupons, and so forth. When customers act in this way, models that do not incorporate this aspect of the decision, and thus fail to consider the censored nature of the observed data, can lead to biased inferences about the "true" underlying processes. In this research, we show that by incorporating a customer's choice to "act" into a model, one can improve both the prediction and understanding of many interesting and managerially important features of an observed data set. Our approach is general and can be used directly in *any* setting in which the transactions are observed if and only if the magnitude of the transaction is sufficiently large, such as the estimation of reservation prices, or the analysis of customers who call a technical support line only for those problems that are sufficiently complex. The focus of this research paper is one such instance: the decision of households to file claims on their homeowner's insurance policies.

This research is motivated by access to a unique and rich data set provided by State Farm Fire and Casualty Company, the largest seller of personal homeowners' insurance policies in the United States, in an attempt to better understand the choices its customers make when they decide to file claims. A specific goal is to improve the prediction of the rate at which customers file small claims (claims amounting to less than $1,000) and to understand why the average size of claims increases (within a household) from claim to claim, even after adjusting for inflation. We present an explanation for both of these phenomena that not only accounts for the rate of small claims and the nonstationarity of all claims, but also allows for the segmenting of customers according to their selectivity in filing claims.

To do this, we construct a probability model that allows some insurable losses (those greater than the specified policy deductible) to remain unclaimed. As an illustration, consider a customer who experiences a loss that is covered by his homeowner's insurance. Once the amount of that loss is determined, the homeowner can claim an indemnity (a reimbursement) from his insurer for the amount that the severity of the loss exceeds the deductible in the policy. For large losses, such as the destruction of a home by fire, one would expect most homeowners to file a claim without hesitation. But for a loss that exceeds the deductible by a more modest amount, the customer may decide to forgo the indemnity and absorb the amount of the damage himself. For example, suppose this homeowner has an insurance policy with a $1,000 deductible, meaning that the first $1,000 of any loss is his responsibility, with the remainder eligible for reimbursement. If the fire damage is $50,000, then the homeowner can receive an indemnity of $49,000, a claim that he will likely file. But if the homeowner suffers only $1,200 in damage from a ball thrown through his window, it is not certain that the homeowner would "act" by filing a claim for a $200 indemnity.

This paper introduces to the literature the idea of the *pseudodeductible*, a latent, unobserved threshold that determines whether or not an insured loss is large enough to trigger the policyholder to file a claim on that loss. Although the amount of the policy deductible is known to both the policyholder and the insurer (indeed, it is specified in the insurance contract), and serves as a hard floor on the size of a loss that may be claimed, we believe that the true lower bound for the severity of claimed losses is somewhat higher. In the example above, the homeowner would file a claim on the window damage if and only if the pseudodeductible is less than $1,200. Otherwise, the loss remains unclaimed and unobserved. From the perspective of the insurer, the loss never happened.

Our analysis centers on determining jointly the size of the pseudodeductible, the frequency of all insurable losses (both claimed and unclaimed) and the severities of these losses. This is a formidable task, because neither the unclaimed losses nor the pseudodeductible threshold are observed directly. By exploiting a rich proprietary source of household-level data, we can (and do) infer a relationship between a customer's rate of losses (both claimed and unclaimed) and the pseudodeductible. We do this using methods of Bayesian inference (Gelman et al. 2004, Congdon 2001), where the likelihood function is constructed from three distinct latent stochastic processes:

(1) a severity (magnitude) process for the size of each loss (either claimed or unclaimed);

(2) a timing process for the occurrence, or "arrival" of the losses; and

(3) a choice process that determines whether the customer will claim the loss.

These processes work together such that the severity and choice processes determine which of the transactions in the timing process are observed. By decomposing the generating process of the observed data into these three subprocesses, we can identify relationships between the rate at which all losses occur and the size of the difference between the policy deductible and the pseudodeductible.

Our empirical findings are consistent with a story that, when deciding which losses to claim, those customers who experience frequent losses may be more selective than those with fewer losses. There are many reasons why this might be, such as the presence of transactions costs associated with filing a claim (either explicit or implicit) or anticipated future premium increases associated with filing a claim. Our focus, however, is to use the pseudodeductible as a tool to explain directly observed phenomena, such as

(1) the percentage of households filing small claims (in our sample, 18% of households filed at least one claim of less than $1,000 during the six-year observation period);

(2) the average increase in claim severity from claim to claim ($114 in our sample); and

(3) the percentage of households whose claim severity increases from claim to claim (53.1% in our sample).

In other words, in developing a model that allows for the possibility that customers leave money on the table in the short term, we can make better predictions about many different features of the observed customer activity. Furthermore, our model allows us to segment the existing customer base according to their estimated latent loss rates and claim thresholds, rather than depending solely on the directly observed transactions history.

To our knowledge there have been no previous attempts to either estimate the size of the pseudodeductible, or to use the pseudodeductible as an empirical modeling tool. However, there are several streams of research that touch on many of the concepts that we use in this paper. The idea of an optimal claims decision is analogous to other settings in which individuals may choose to leave money on the table, such as the failure to redeem rebates, or to participate in welfare programs (Moffitt 1983) or retirement plans (Choi et al. 2005). In what Lemaire (1995) calls "a rare example of research duplication in actuarial science," various solutions for optimal decision rules for claims on automobile insurance policies (assuming various forms of rewards and penalties for favorable or detrimental claims history) have appeared in journals of operations research (Haehling von Lanzenauer 1974, Hastings 1976, Norman and Shearn 1980), economics (Venezia and Levy 1980, Venezia 1984, Dellaert et al. 1993), and actuarial science (De Leve and Weeda 1968, de Pril 1979, Lemaire 1995), with the common thread among all results being that a claim would be filed if its value is greater than the resulting expected discounted utility. However, for all of the interest in this topic, there has been no attempt (as far as we know) to estimate the claims "rule" that individuals actually apply. Newhouse et al. (1980) comes closest, recognizing that the effective deductible (what we call a pseudodeductible) for medical insurance may be somewhat larger than the policy deductible. More recently, Israel (2004) showed that drivers with past automobile insurance claims tend to drive more safely as the number of claims increases, because incremental claims become more and more expensive in terms of both premium charges and the looming possibility that a policy might be cancelled. In the marketing literature, the beta-binomial/negative-binomial distribution (BB-NBD) model is used frequently in situations in which the transaction rates and reporting probabilities are randomly distributed across the population (Schmittlein et al. 1985, Fader and Hardie 2000). Instead of using a beta distribution to model the probability of reporting a count event, as in the BB-NBD model, van Praag and Vermeulen (1993) assume that an event is reported if and only if another variable exceeds some known threshold.

Section 2 formally presents the pseudodeductible model in which the pseudodeductible is assumed to be stationary. Sections 3 and 4 respectively describe model estimation and parameter inferences, including a posterior predictive check of the model assessed for many of the interesting features of the data (Rubin 1984, Gelman et al. 1996). In §5, as a demonstration of the power of the model, we segment the customer base according to the posterior probabilities of having particular combinations of loss rates and pseudodeductibles. Section 6 refines the model by using nonstationary pseudodeductibles to explain the phenomenon of increasing claim severities. Finally, in §7, we discuss our results, extend our insights outside of the insurance field and propose topics for future related research.

2. The Model

We take a Bayesian approach to modeling this problem, which involves constructing a posterior distribution of the frequency of all losses (both claimed and unclaimed), the severity of these losses and the latent threshold that defines the pseudodeductible. This posterior distribution is composed of two parts: the likelihood of the observed data, and a prior (or mixing) distribution that allows for heterogeneity of the likelihood parameters across the population. In this section, we begin by constructing the likelihood of the observed data by integrating together the distributions of all of the data. We then introduce a semiparametric prior distribution on the parameters of the likelihood.

2.1. Notation

Let $f^o(t^o, y^o \mid \theta)$ be the probability density function for the timing and severity of observed claims for all H households, such that

$$f^o(t^o, t_s^o, y^o \mid \theta) = \prod_{h=1}^{H} f_h^o(t_h^o, t_{hs}^o, y_h^o \mid \theta_h), \tag{1}$$

where $f_h^o(\cdot)$ is the observed data likelihood for household h, $t_h^o = (t_{h1}^o, \ldots, t_{hK_h}^o)$ is the vector of claim interarrival times for the K_h claims that are filed by household h during the observation period, t_{hs}^o is the "survival time" between the K_hth filed claim and the end of the observation period, and $y_h^o = (y_{h1}^o, \ldots, y_{hK_h}^o)$ is the vector of the severities of the *losses* that were filed as claims.[1] θ_h is the parameter vector associated with household h (the contents of θ_h will be defined in more detail as we develop the model). By definition, $T_h^o = (T_{h1}^o, \ldots, T_{hK_k}^o)$ is the vector of claim arrival times. Additionally, T_{h0}^o is the beginning of the observation period for household h, T_x^o is the end of the observation period, $t_{hk}^o = T_{hk}^o - T_{hk-1}^o$ for all $k = 1, \ldots, K_h$ and $t_{hs}^o = T_x^o - T_{hK_h}^o$.

In contrast to T_h^o, the vector T_h contains the arrival times for *all* I_h losses experienced by household h, including both those that are claimed and those that are unclaimed and therefore unobserved. Thus, if T_h^u is the vector of arrival times of *unobserved* (unclaimed) losses, then $T_h = (T_h^o, T_h^u)$. The loss interarrival times for all losses are denoted as $t_{hi} = T_{hi} - T_{hi-1}$ for $i = 1, \ldots, I_h$. However, it is often more convenient to describe the arrival time of the ith loss, T_{hi}, as $T_{hk,j}$, the time of the jth loss that occurs between the arrival times of the $(k-1)$th and kth claims. When we use this notation, the interarrival time between two losses is $t_{hk,j} = T_{hk,j} - T_{hk,j-1}$. If there are no observed claims, then $K_h = 0$, the t_h^o vector is empty, and $t_{hs}^o = T_x - T_{h0}$. A graphical representation of this arrival process for a hypothetical household with two observed claims is presented in Figure 1.

In addition, define D_{hk} as the amount of the policy deductible in force at the time of claim k. Thus, the amount of the indemnity received by the policyholder is $\breve{y}_{hk} = y_{hk}^o - D_{hk}$, and y_{hk}^o, the full amount of a claimed loss, is the sum of the policy deductible and the indemnity received by the policyholder. If $K_h = 0$, then y_h^o is an empty vector.[2] The vector y_h is

[1] A ^{o} indicates an observed data vector or a distribution of observed values. Notations without a ^{o} will be used for constructs in which some items are unobserved, or for distributions of all transactions where some may be unobserved.

[2] We assume that the value of the loss is known at the time the homeowner decides to file the claim, that the homeowner files claims for exactly the correct and reimbursable amount, and that the insurer pays all eligible claims for the full amount. Issues of insurance fraud and insurer solvency go beyond the scope of this research. Furthermore, insurers of personal lines typically do not negotiate with claimants on indemnity amounts.

Figure 1 Loss and Claim Arrival Process for a Hypothetical Household with Two Claims That Occur at Times T^o_{h1} and T^o_{h2}

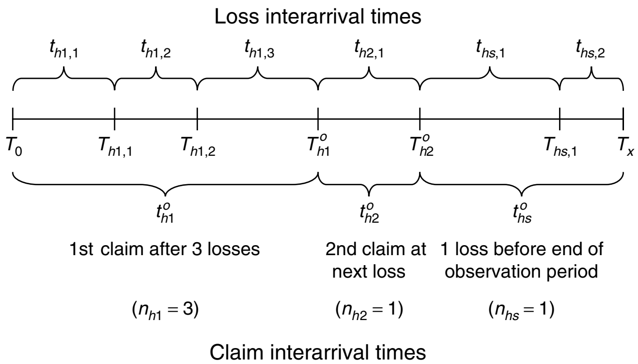
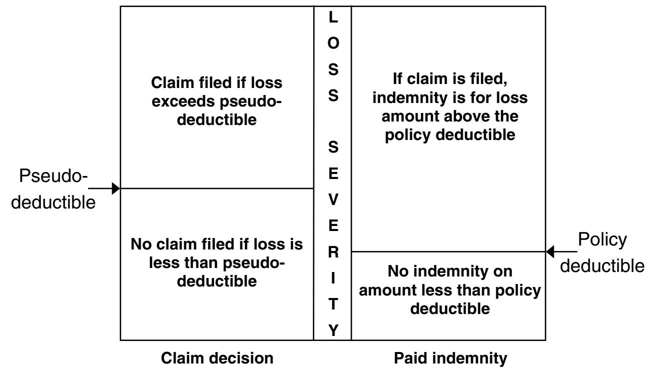


Figure 2 Relationship Between Policy Deductible and Pseudodeductible



the set of severities for all I_h losses, both observed and unobserved. We do not know when the unobserved losses occur (if there are any), so we have to make an assumption about the deductible levels that are in force at the times of these losses. We define $\bar{D}_{hk}$, the time-weighted average of D_{hk} and D_{hk-1}, as the average deductible in force during the period between the kth and $(k-1)$th claims, and $\bar{D}_{hs}$ as the average deductible in force during the period between the kth claim and the end of the observation period. If $D_{hk} = D_{hk-1}$, then $\bar{D}_{hk} = D_{hk} = D_{hk-1}$.

The pseudodeductible for household h at the time of claim k is designated as $\Psi(D_{hk}, \psi_h)$, where $\Psi(\cdot)$ is a function and ψ_h is a latent, household-specific parameter that determines how much larger the pseudodeductible is over the policy deductible. $\Psi(\bar{D}_{hk}, \psi_h)$ is the pseudodeductible in force during the period between the arrival times of claim $k-1$ and claim k. The function $\Psi(\cdot)$ potentially could take an infinite number of functional forms, but in this paper we will consider three:

(1) Identity model: $\Psi(D_{hk}, \psi_h) = D_{hk}$;
(2) Additive model: $\Psi(D_{hk}, \psi_h) = D_{hk} + \psi_h$; or
(3) Multiplicative model: $\Psi(D_{hk}, \psi_h) = D_{hk}(1+\psi_h)$.

Stated explicitly, the pseudodeductible, $\Psi(D_{hk}, \psi_h)$, is the threshold that determines whether a loss is large enough to be claimed, whereas the pseudodeductible *factor*, $\psi_h \geq 0$, controls the relationship between the policy deductible and the pseudodeductible. Under the Identity model (our baseline), the pseudodeductible is the same as the policy deductible, so customers claim all losses above D_{hk}. The Additive model assumes that the maximum amount of money that a customer would forgo does not depend on the size of that policy deductible, while the Multiplicative model assumes that the pseudodeductible increases proportionally with the policy deductible. Figure 2 illustrates the relationship between the policy deductible and the pseudodeductible for those models in which $\Psi(D_{hk}, \psi_h) > D_{hk}$, such as the Additive and Multiplicative models.

The existence of the pseudodeductible in insurance decisions suggests that policyholders are willing to forgo reimbursements to which they are otherwise entitled. In the short term, this might be explained by the costs associated with filing a claim, such as the opportunity cost of time for filing the claim, or more direct expenses such as the cost of gathering damage repair estimates. Longer-term costs may matter as well. The pseudodeductible could offset the expected future discounted cash flows associated with a claim, such as premium increases or policy cancellation. By allowing for unobserved heterogeneity in the pseudodeductible estimates, we can account for the variance in thresholds and risk perceptions that are likely to exist in the population.

One explanation of the behavioral distinction between additive and multiplicative pseudodeductibles is related to the diminishing marginal value of money. The policy deductible represents the amount of a loss the customer must absorb before making any claims decisions. Hence, D is the amount the customer is "in the hole" before he decides what additional amount of the loss he is willing to pay by himself. The customer will then absorb incremental loss dollars until this cost of not filing a claim offsets the benefit (e.g., saving transaction costs or preventing increases in future premia). Under an Additive pseudodeductible, the amount of money the customer is willing to absorb is independent of D, but under the multiplicative model, this amount increases with D. Because the benefits of not filing do not depend on D in either case, the Additive model implies that the value function of money, in the domain of losses, is linear, whereas the value function for the Multiplicative model is convex. Consequently, the Multiplicative model implies a value function that is consistent with Prospect Theory (Kahneman and Tversky 1979), and suggests that proportions, rather than absolute amounts, impact decisions (Thaler 1980, Kahneman and Tversky 1984). For example, suppose a customer has a policy deductible of \$100 and faces a \$300 loss.

He has to decide whether or not to pay an additional \$200 out of his own pocket. If he had a \$1,000 deductible instead, and faced a \$1,200 loss, he would also have to decide whether or not to pay \$200. But in the first scenario, he is only committed for \$100, and in the second, he is committed for \$1,000. The Multiplicative model suggests that the marginal value of the \$200 is greater when the customer has paid only \$100 than when he has paid \$1,000.

2.2. Deriving the Likelihood

The likelihood for the observed claims data, conditional on the parameters, can be factored into distributions for the severity vector, the interclaim arrival times and the survival time. Thus, from (1),

$$f_h^o(t_h^o, t_{hs}^o, y_h^o \mid \theta_h) = f_y^o(y_h^o \mid t_h^o, t_{hs}^o, \theta_h) f_s^o(t_{hs}^o \mid t_h^o, \theta_h) f_t^o(t_h^o \mid \theta_h), \quad (2)$$

where $f_y^o(\cdot)$, $f_s^o(\cdot)$, and $f_t^o(\cdot)$ are the likelihoods of the severities, survival time, and interclaim times, respectively. In the following sections we derive each of these distributions.

2.2.1. The Choice and Severity Processes. Let $F_y(y_{hi} \mid \theta_h)$ be the cumulative distribution function for a single loss severity, y_{hi}, on the domain $(0, \infty)$, and let $f_y(y_{hi} \mid \theta_h)$ be its density. By definition, the pseudodeductible is the latent threshold that determines whether a loss y is observed as a claim y^o. This means that the likelihood of observing a specific claim y_{hk}^o is zero if $y_{hk}^o < \Psi(D_{hk}, \psi_h)$. Because we know that any observed y_{hk}^o must be greater than $\Psi(D_{hk}, \psi_h)$, the distribution of y_{hk}^o is conditional on this restriction. Therefore, the severity density of an observed claim is

$$f_y^o(y_{hk}^o \mid \theta_h) = \frac{f_y(y_{hk}^o \mid \theta_h)}{1 - F_y(\Psi(D_{hk}, \psi_h) \mid \theta_h)} \mathbf{1}\{y_{hk}^o \geq \Psi(D_{hk}, \psi_h)\}, \quad (3)$$

where $\mathbf{1}\{\cdot\}$ is an indicator function that takes a value of one if the argument inside the braces is true, and zero otherwise.

The likelihood depends, of course, on the parametric family that we choose for $F_y(y_{hi} \mid \theta_h)$. We have chosen a flexible family by assuming that a loss, y_{hi}, is drawn from a Weibull distribution with shape parameter c and scale parameter μ_{hi}. In addition, we assume that μ_{hi} is distributed across all losses according to a gamma distribution with shape parameter r and scale parameter a. Hence,

$$F_y(y_{hi} \mid \mu_{hi}, c) = 1 - \exp[-\mu_{hi} y_{hi}^c]$$

and

$$g_\mu(\mu_{hi} \mid r, a) = \frac{a^r \mu_{hi}^{r-1} \exp[-a\mu_{hi}]}{\Gamma(r)}$$

(Hardie and Fader 2005). By integrating across the distribution of μ_{hi}, the marginal cumulative distribution function is

$$F(y_{hi} \mid r, c, a) = \int_0^\infty F_y(y_{hi} \mid \mu_{hi}) g_\mu(\mu_{hi})\, d\mu_{hi} = 1 - \left(\frac{a}{a + y_{hi}^c}\right)^r, \quad (4)$$

and the density function is

$$f(y_{hi} \mid r, c, a) = \frac{rcy_{hi}^{c-1}}{a + y_{hi}^c}. \quad (5)$$

This distribution is known as the Weibull-gamma distribution, or alternatively as the three-parameter Burr XII distribution (Hardie and Fader 2005, Johnson et al. 1994, Klugman et al. 1998). We use this mixture distribution for severities for several reasons:

(1) It allows for heterogeneity in losses across all households *and* across losses within each household. Loss-level heterogeneity allows for the possibility that a single household may experience many different kinds of losses, some more damaging than others (such as the distinction between a broken window from a ball and a shredded roof from a tornado). Assuming loss-level heterogeneity is necessary because, otherwise, we would be using the same distribution to model severities of different kinds of losses that might be experienced by the same household.

(2) With three parameters, the Weibull-gamma is a flexible distribution that can take a number of different shapes.

(3) The Weibull-gamma has a cumulative distribution function in closed form, making its use mathematically tractable.

The severity portion of the likelihood, $f_y^o(y_h)$, is formed by substituting (4) and (5) into (3) and multiplying across claims such that

$$f_y^o(y_h \mid r, c, a) = \begin{cases} \prod_{k=1}^{K_h} f_y^o(y_{hk}^o \mid r, c, a) & \text{if } K_h > 0 \\ 1 & \text{if } K_h = 0. \end{cases} \quad (6)$$

2.2.2. Timing and Survival Likelihoods. The distributions for $f_t^o(\cdot)$ and $f_s^o(\cdot)$ are based on $f_t(t_{hi} \mid \theta_h)$, the probability density function of a single loss interarrival time for household h. By definition, those losses whose severities are less than the pseudodeductible are unclaimed and unobserved. Let n_{hk} be the number of losses since the claim at time T_{hk-1}^o up to and including the claim at time T_{hk}^o; n_{hk} is the number of unobserved losses it takes to get to the next observed claim. The time between claim $k-1$ and claim k, denoted by t_{hk}^o, is equal to $T_{hk}^o - T_{hk-1}^o$, which in turn is equal to the sum of the n_{hk} loss interarrival times between t_{hk}^o and t_{hk-1}^o. Hence, the distribution of $f_t^o(t_{hk}^o \mid \theta_h)$, the claim interarrival time for a single claim, is equivalent to the n_{hk}-fold convolution of $f_t(t \mid \theta_h)$.

Let $f_t(t_{hi} \mid \theta_h)$ be an exponential distribution with rate λ_h, such that

$$f_t(t_{hi} \mid \lambda_h) = \lambda_h \exp[-\lambda_h t_{hi}].$$

This distributional choice underlies an assumption that losses arrive according to a Poisson process at the household level, with each household having its own loss arrival rate. The n_{hk}-fold convolution of an exponential distribution is an Erlang distribution with shape parameter n_{hk} and scale parameter λ_h (Kulkarni 1995). Therefore, if n_{hk} were known, the density of the claim interarrival times would be

$$f_t^o(t_{hk}^o \mid n_{hk}, \lambda_h) = \frac{\lambda_h^{n_{hk}} t_{hk}^{o^{n_{hk}-1}} \exp[-\lambda_h t_{hk}^o]}{(n_{hk}-1)!}. \tag{7}$$

But n_{hk} is not known. We have already characterized the claim/no-claim decision as a Bernoulli process, so n_{hk} is a random variable with a geometric distribution with "success" parameter p_{hk}, where

$$p_{hk} = 1 - F_y(\Psi(\bar{D}_{hk}, \psi_h) \mid \theta_h) \tag{8}$$

(we use the average policy deductible $\bar{D}_{hk}$ because we are interested in p_{hk} for the time interval between two observed claims). Because (7) is the distribution of the claim interarrival times conditional on n_{hi}, we uncondition (7) by summing over the distribution of n_{hk}. This gives us the distribution of claim interarrival times of

$$\begin{aligned} f_t^o(t_{hk}^o \mid \theta_h) &= \sum_{m=1}^{\infty} \frac{\lambda_h^m t_{hk}^{o^{m-1}} \exp[-\lambda_h t_{hk}^o]}{\Gamma(m)} \cdot (1-p_{hk})^m p_{hk} \\ &= \lambda_h p_{hk} \exp[-\lambda_h p_{hk} t_{hk}^o] \end{aligned} \tag{9}$$

(Kulkarni 1995).

Notice that (9) is an exponential distribution with rate parameter $\lambda_h p_{hi}$. Therefore, the density of t_{hs}^o, the survival time, is equal to the probability that the arrival time of the next claim is greater than t_{hs}^o. Using the cumulative distribution function of the exponential distribution, we get

$$f_s^o(t_{hs}^o \mid \theta_h) = 1 - \exp(-\lambda_h p_{hs} t_{hs}), \tag{10}$$

where $p_{hs} = 1 - F_y(\Psi(\bar{D}_{hs}, \psi_h))$.

As in (6), the likelihood vector for the interclaim times is

$$f_t^o(t_h^o \mid \theta_h) = \prod_{i=1}^{K_h} f_{hk}^o(t_{hk}^o \mid \theta_h). \tag{11}$$

By substituting (6), (10), and (11) into (2), and in turn substituting (2) into (1), we get the likelihood of the complete observed data set in (1).

2.3. Heterogeneity Across Households

With the data likelihood established, we now turn to the question of modeling heterogeneity across households by specifying a class of mixing distributions for household-specific parameters. From (9), (10), and (11), we see that the entire parameter space for the model is $(\lambda_1, \ldots, \lambda_H, \psi_1, \ldots, \psi_H, r, c, a)$. In §2.2.1, we explained that r is assumed to be homogeneous across households, and c and a are the parameters of the mixing distribution on μ_{hi}. The issue remaining is how to allow $\theta_h = (\lambda_h, \psi_h)$ to vary across households. One option, of course, is to include no heterogeneity at all. Another would be to apply a known parametric mixing distribution to λ_h and ψ_h.

Instead, we apply heterogeneity semiparametrically by using a mixture of Dirichlet processes (MDP) as a prior on (λ_h, ψ_h) (see Walker et al. 1999). Dirichlet processes were first presented by Antoniak (1974) and Ferguson (1983). The statistical idea behind an MDP is that the mixing distribution for $(\log \lambda_h, \log \psi_h)$, converted to a $(-\infty, \infty)$ scale, is itself a mixture of some unknown number of "kernel" distributions. In our case, we use the common kernel choice of the bivariate normal distribution. The advantage of using an MDP is that any distribution can be approximated by incorporating a sufficient number of these kernels without resorting to a specific parametric form (Kim et al. 2004). Furthermore, an MDP can be used as a proxy for latent class models without a need to specify the number of classes up front (Escobar and West 1995), and can relax restrictions that are often imposed by parametric priors, such as unimodality (Draper 1999). Additionally, the MDP model allows for within-class heterogeneity, unlike standard latent class approaches. Examples of the use of MDP in a variety of settings are available in Congdon (2001) and an application of the MDP to discrete choice models is developed by Kim et al. (2004).

The distribution of λ_h and ψ_h can be defined as a mixture of L component distributions of (λ_l, ψ_l), each weighted by a corresponding element of the probability vector $\pi = (\pi_1, \ldots, \pi_L)$, where $\sum_{l=1}^{L} \pi_l = 1$. If $\pi_l \approx 0$ for any l, we call that component "empty," and thus L is an upper bound on the number of nonempty components in the mixture (which may be as high as the number of households). We then place a "constructive" (or "stick-breaking") Dirichlet prior, with control variable α, on π (Sethuraman 1994, Walker et al. 1999, Congdon 2001), a diffuse bivariate normal hyperprior on $(\log \lambda_l, \log \psi_l)$ for $l = 1, \ldots, L$ and diffuse hyperpriors on all variance-covariance matrices.[3] All together, we denote this prior on $(\log \lambda_h, \log \psi_h)$ as $MDP(G_0, \alpha)$, where G_0 (the "kernel" distribution) is

[3] α has no clear interpretation other than a determinant of the level of smoothing in the model (Walker et al. 1999).

the bivariate normal hyperprior on $(\log\lambda_h, \log\psi_h)$. By constructing the prior on (λ_h, ψ_h) in this way, our estimation process (discussed below) allows us to simulate from the marginal posterior distributions of λ_l, ψ_l, and π_l for all l. Because the MDP is a mixture of component distributions, groups of customers with (λ_h, ψ_h) pairs that are similar to one another are likely to draw their parameters from the same component. We say that these customers are of the same "type." The posterior distribution of a household's λ_h and ψ_h can then be defined by both the posterior probability of being in each type and the distributions of λ_l and ψ_l within that type.

3. Estimation

Our goal is to provide inference regarding the posterior distribution

$$g(\theta \mid t^o, y^o) \propto f^o(t^o, y^o \mid \theta) g(\theta), \tag{12}$$

where $f^o(t^o, y^o \mid \theta)$ is distributed according to (1) and $g(\theta)$ is the full hierarchical prior for $\theta = (\lambda_1, \ldots, \lambda_L, \psi_1, \ldots, \psi_L, r, c, a)$:

$$(\log\lambda_l, \log\psi_l) \sim MDP(\alpha, G_0)$$
$$G_0 \sim MVN(\beta, \tau)$$
$$\beta \sim MVN(\beta_0, \Omega).$$

We placed proper but weakly informative hyperpriors on β_0, $\log r$, $\log c$, $\log a$, τ, and Ω. The control parameter α was set at 0.5. Details of the computation are provided in §3.2. We next describe an application of this pseudodeductible model to a data set of insurance claims.

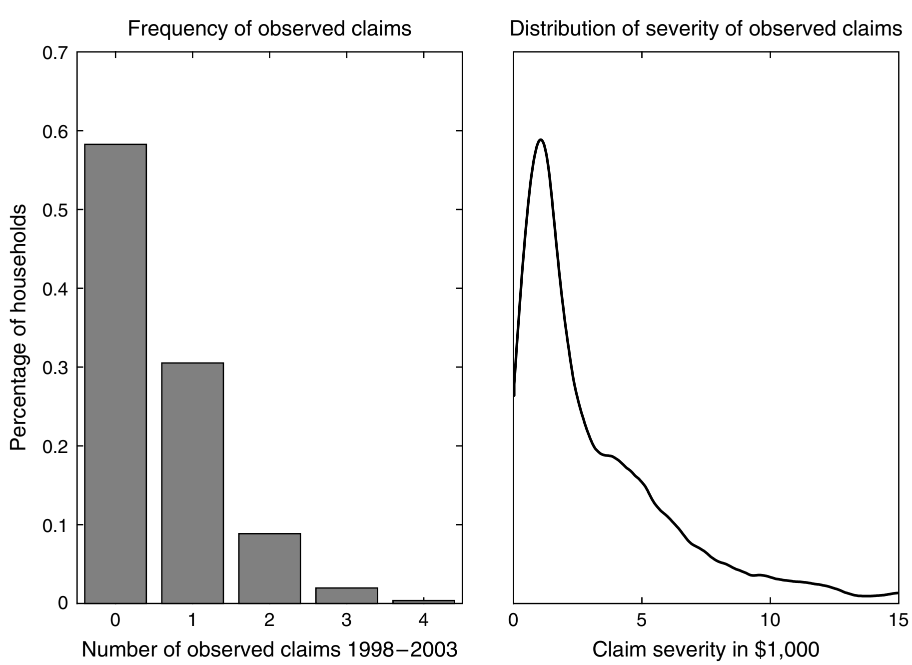
3.1. The Data

State Farm provided deductible and paid-claims history information for all homeowners policies on single-family homes (more than one million records) that were in effect in six midwestern states for any of the years from 1998 to 2003. The records are taken as snapshots on December 31 of each of the six years in the observation period. The data are proprietary to the company and were provided to us solely for use in the present study. State Farm uses these data for critical operations (such as the disbursement of claims to policyholders), so we are extremely confident in their quality and accuracy.

To make the analysis computationally feasible, we selected a subset of households from the comprehensive data set to create a smaller working data set. First, we restricted the analysis to those policies that were in force on December 31 for the entire six-year duration of the panel. This step reduced the number of eligible households to about 435,000. Second, we generated two systematic random samples of 0.75% of these homes, leaving calibration and holdout data sets each containing 3,267 homes. We present a graphical summary of this data in Figure 3. Note that nearly 60% of the households have filed no claims at all during this period. Even though the data is sparse, we will employ a Bayesian inferential approach that shares information across subjects so that we can draw within-subject inferences about household-level propensities, even when households do not file any claims at all.

The timing of a claim, t^o_{hk} is expressed in terms of the number of years (or fraction of years, by week)

Figure 3 Distributions of Claim Frequencies (Left) and Severities (Right) for Observed Data



between the times of the previous and current claims. For the first claim, t^o_{h1} is the number of years since the 1998 effective date of the policy for household h. The survival time is the difference between the time of the last claim (or, if there are no claims, the 1998 policy effective date) and December 31, 2003. All severity amounts were adjusted to 2004 U.S. dollars using the U.S. Consumer Price Index for the Midwest (Bureau of Labor Statistics 2004).

3.2. Estimation Method

Our inferential approach is to simulate draws from the marginal posterior distribution defined by (12) using Markov chain Monte Carlo (MCMC) methods. In particular we utilized the freely available WinBUGS Bayesian modeling software package (Spiegelhalter et al. 1996). The WinBUGS code for implementing the MDP prior was adapted from code described in Congdon (2001). To improve identification and MCMC convergence properties, we set a baseline class where $\psi_1 = 0.001$ (a pseudodeductible of \$1 per \$1,000 of deductible in the Multiplicative model, and \$1 for all deductible levels in the Additive model), so $\Psi(D_{hk}, \psi_1) \approx D_{hk}$ for at least one type. Potential label switching was addressed by postprocessing MCMC draws according to the algorithm proposed by Stephens (2000). After several trial runs with L (the maximum number of components) set very high, we observed that many of the posterior estimates for π_l were indistinguishable from zero. We interpreted those components as being "empty," and eventually limited the number of classes under consideration to $L = 8$. This choice of L dramatically reduces computation time, and we confirmed that a higher maximum value for L would have no impact on our results. The three models we tested correspond to the three alternative definitions of the pseudodeductible listed in §2.1: Identity, Additive, and Multiplicative. We ran two independent Markov chains for each model and, after a burn-in period, we selected the final 6,000 draws from each chain, for a total MCMC sample of 12,000 draws per pseudodeductible model.[4]

4. Results

In this section we begin by comparing the log-marginal likelihoods and severity parameter estimates for the three models under consideration. After concluding that the Multiplicative model has the best fit relative to the other models, we will examine the parameter estimates for that model's loss frequency and pseudodeductible estimates. In addition, we will show how the Multiplicative model explains the observed data well in an absolute sense.

Table 1 Log-Marginal Likelihoods

	Log-marginal likelihood	
Model	In-sample	Holdout
Identity	−16,665	−16,708
Additive	−11,255	−11,299
Multiplicative	−7,999	−8,046

4.1. Log-Likelihood Comparison

As a comparative global measure of model fit, we use the log of the marginal likelihood, computed using the importance sampling method proposed by Newton and Raftery (1994). The difference between any two log-marginal likelihoods is the log of the Bayes factor, reflecting the relative strength of support for those models. The log-marginal likelihoods for both the calibration and cross-sectional holdout samples are presented in Table 1. Holdout log-marginal likelihoods were determined by computing the log-marginal likelihood of the holdout data using the draws from the predictive posterior distribution based on the calibration data set. It is clear that the Multiplicative model fits best, because the logs of the Bayes factors comparing it to the Additive and Identity models are over 3,000 and 8,000, respectively. The Identity model is nested within both the Additive and Multiplicative models, so the difference in log-marginal likelihoods can be interpreted as the improvement in model fit contributed by the introduction of the pseudodeductible.[5]

4.2. Estimated Severity Distributions

One way to present the variation in the inferences derived from these three models is to look at the severities of a "central" loss that is implied by each model. We focus on the median of the loss severity distribution, and present the quantiles for the posterior distributions of the median loss severities

[4] The results presented here utilized 100,000 draws for burn-in. Future runs indicated much faster convergence, as indicated by a z-test due to Geweke (1992). Hence, in practice, much shorter runs than those utilized here are reasonable.

[5] Although there is evidence to support the Multiplicative model on both theoretical and empirical grounds (through the use of the log-marginal likelihoods in this section and posterior predictive checks below), one should use caution when interpreting these values. This is because we are placing noninformative priors on ψ, which has different interpretations in the Additive and Multiplicative models. For a further discussion on this issue, see Bernardo and Smith (2000, Chapter 6). We thank an anonymous reviewer for pointing out this concern regarding the use of log-marginal likelihoods (and, similarly, Bayes factors) with weakly informative priors.

Table 2 Estimated Median Loss Severities

	Estimated median loss (1,000 US$)		
Model	10%	50%	90%
Identity	1.108	1.248	1.374
Additive	0.317	0.433	0.564
Multiplicative	0.355	0.456	0.609

in Table 2. These quantiles are computed from the MCMC samples described in §3.[6]

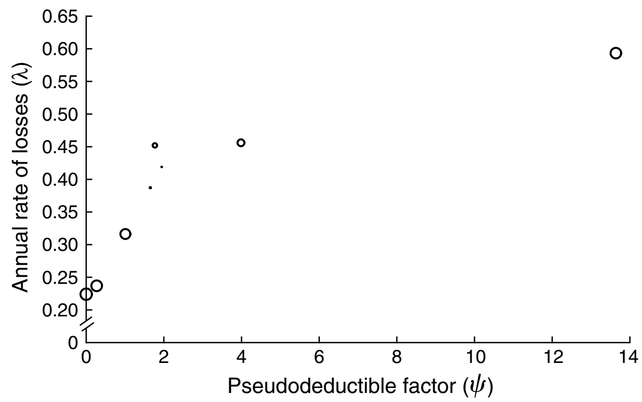
The median of the median loss distribution for the Identity model is $1,248, whereas the estimated medians for the Additive and Multiplicative models are significantly lower: $433 and $456, respectively. These estimates characterize the distribution of *all* losses, not just those that are claimed. So, given the observed claims, the results from the models that include a pseudodeductible imply that there are more losses that remain unclaimed than is implied by the nonpseudodeductible Identity model. Thus, the Identity model overestimates the median loss severity. We have confirmed this pattern through extensive simulation studies, and found that underestimating the "true" pseudodeductible can lead to overestimation of the median of the underlying severity distributions.

4.3. Timing and Pseudodeductible Estimates

The Multiplicative model is our best-fitting model, as indicated in Table 1, and therefore all subsequent reported results are based on this model. Like the MDP prior distribution on (λ_h, ψ_h), the posterior distribution is also semiparametric. The density at any point in the (λ_h, ψ_h) space is a mixture of the L component densities, each with a median at (λ_l, ψ_l). One way to simplify the presentation of the posterior distribution is to focus on those L medians, which are analogous to the support points of the mixing distribution in frequentist latent class modeling (Escobar and West 1995).

We characterize the posterior distribution for λ_l, ψ_l, and π_l by plotting the medians of the component distributions in Figure 4, and by summarizing the quantiles in the online supplement.[7] We interpret λ_l as the rate (in annual units) at which losses arrive for type l households. Thus, λ_l is the expected number of losses in a year, and $1/\lambda_l$ is the expected time between losses. Furthermore, ψ_l is the percentage *above the policy deductible* that characterizes the pseudodeductible for type l, and π_l is the proportion of homes that are estimated to be of type l. The rates of losses range from 0.22 to 0.59, which during the six-year observation period translates to 1.3 to 3.5 losses during that period. The range for ψ is much greater; not counting type 1 (for which $\psi_1 = 0.001$ by definition), the median pseudodeductibles range from 127% to 1,464% of the policy deductible, with a weighted average (including type 1) of about 451%. The relationship between λ and ψ, as illustrated in Figure 4, means that those households with higher rates of losses also tend to have higher-percentage differences between their policy deductibles and pseudodeductibles and that those households with low rates of loss occurrence also have lower pseudodeductible factors. Put another way, the relative loss-claim threshold for a "frequent-loss" homeowner is higher than the threshold for an "infrequent-loss" homeowner.

Figure 4 Loss Rate and Pseudodeductible Factors for Each Type



Note. Circle diameters are proportional to the weighting probabilities (π_l) for each class.

If we believe that individuals are expecting additional future costs from filing a claim, this result makes sense. A policyholder that has already filed a claim might expect that an additional claim would lead to a substantial increase in premiums, or that the policy might be canceled altogether. As a result, the policyholder who has a higher rate of losses would be more selective about filing claims, and would then have a higher pseudodeductible. Although this model suggests that this relationship is static, we will show in §6 that the pseudodeductible actually evolves in magnitude from claim to claim.

4.4. Posterior Prediction

Now that we have simulated draws from the posterior distribution, we can use these draws to show how the pseudodeductible model can explain those features of the data, of greatest interest, that we presented in the introduction. Our method to accomplish this is the posterior predictive check (Rubin 1984, Gelman et al. 1996). The idea behind using a

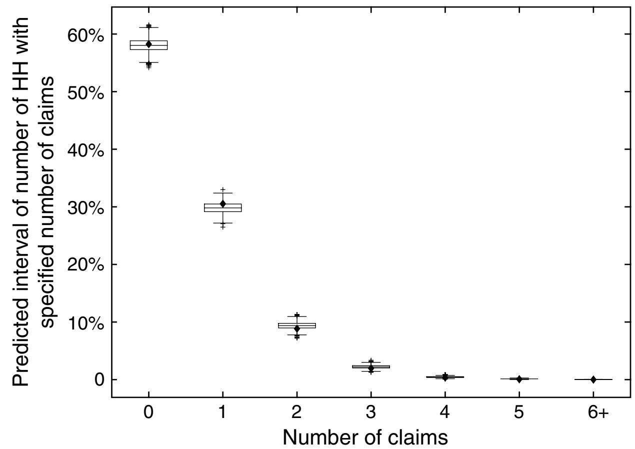
[6] Using the posterior predictive draws of the loss severity parameters r, c, and a, the median m of the Weibull-gamma distribution is computed by solving the equation $F_y(m) = 1/2$. Hence, $m = [a(2^{1/r} - 1)]^{1/c}$. We get a posterior distribution for m by computing m for each posterior predictive draw.

[7] The online supplement is available on the *Management Science* website at http://mansci.pubs.informs.org/ecompanion.html.

posterior predictive check (PPC) is that a model fits well if simulated data sets that are generated from the model "look like" the original data. The similarity between the observed and simulated data sets is assessed on the basis of carefully chosen test statistics that reflect those aspects of the observed data that we care about most. If the simulated test statistics appear to be generated from a model that is consistent with the observed ones, then we can accept the model in an absolute sense (unlike a comparison of log-marginal likelihoods, which assesses only the relative fit). If the observed test statistic falls to the left of the distribution of the simulated test statistics, the model is overestimating that particular characteristic of the model (and vice versa for underestimation). A relevant question, then, is the degree of fit, which is commonly summarized by a posterior predictive p-value (tail area) computed as the proportion of simulated test statistics that are greater than the observed one.

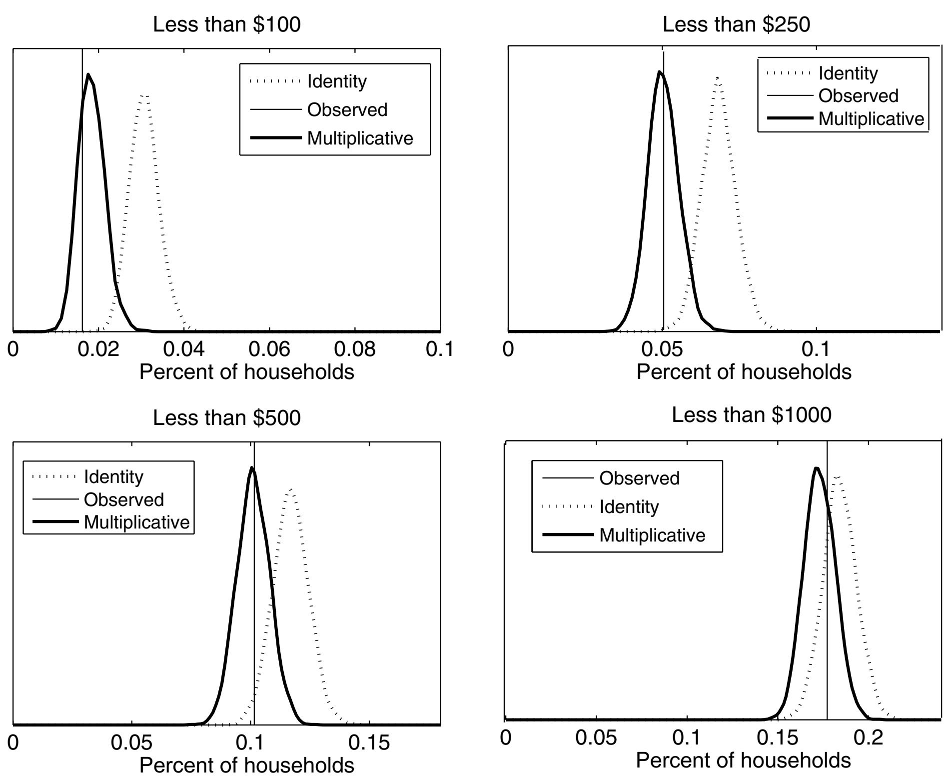
The distributions of the number of claims per household for the replicated data are presented in Figure 5. However, one of the additional features of the data in which we are interested, as described in the introduction, is the proportion of households with at least one small claim. We consider the test statistics of the percentage of households with at least one claim smaller than one of four thresholds—$100, $250, $500, and $1,000—that might determine what constitutes a small claim. Here, the fits of the models diverge. Figure 6 plots these PPCs. From these plots we see that the Identity model overestimates

Figure 5 Posterior Predictive Intervals for Number of Claims



Notes. Each solid diamond represents the observed proportion of households with the given number of claims. The box plot indicates the distribution of proportions across the replicated data sets. The probabilities that the proportion of households for a replicated data set exceeds the observed values are 0.43, 0.24, 0.81, 0.79, 0.66, 0.46, and 0.38.

Figure 6 Posterior Predictive Intervals for Percentage of Households With at Least One Small Claim



Note. Posterior predictive checks for the Identity (dotted line) and Multiplicative (solid line) models for the percentage of households with at least one claim smaller than $100, $250, or $500. The probabilities that a value from a replicated data set exceeds the observed value are 1, 1, 0.98, and 0.78 for the Identity model, and 0.73, 0.39, 0.46, and 0.31 for the Multiplicative model.

the percentage of households that file small claims, whereas the Multiplicative model closely predicts this feature of the observed data. These test statistics are of particular interest because they go to the heart of our story—that the size of the pseudodeductible and the propensity to file small claims are intertwined. Including a pseudodeductible in the model allows us to replicate this important feature of the data almost exactly.

5. Inference

5.1. Posterior Classification Probabilities

Another set of inferences under the model relates to the posterior classification of households to one of the L types, as given by π_l^h, the posterior probability that (λ_h, ψ_h) for household h is drawn from component $l = 1, \ldots, L$ of the posterior mixing distribution. Applying Bayes's Theorem

$$\pi_l^h = \frac{f^o(y_h^o, t_h^o \mid \lambda_l, \psi_l)\pi_l}{\sum_{l'=1}^{L} f^o(y_h^o, t_h^o \mid \lambda_{l'}, \psi_{l'})\pi_{l'}},$$

we see that the household-level posterior distributions for losses depends not just on the number and timing of observed claims, t_h^o, but also on the severity of those claims, y_h^o. As an example, a policyholder with low-severity claims may be less likely to be selective about the claims he files, an inference that one would not be able to draw by looking at claim frequencies alone. The finding here is that when we take into account the possibility of unfiled claims, looking at claims alone can be a misleading measure of a household's proneness to "risk."

The claim information for some selected households (arbitrarily labeled A though H and chosen for their illustrative value and representativeness) are summarized in Table 3 and the means of the posterior distributions of π_l^h, the posterior classification probabilities, are presented in Table 4 (a table that includes the 10% and 90% quantiles of the posterior classification probabilities is available in the online supplement). Household A filed no claims during the observation period. There are two stories that can explain this observation: either no losses occurred (so there was nothing to claim), or at least one loss occurred and all losses went unclaimed. The first story suggests that λ_A and ψ_A are low, and the second suggests that they may be high. If we were to assume that because household A has no claims its loss rate must also be low, we would be ignoring the 29% total posterior probability that A is in one of the four "high (λ, ψ)" types, as described in Table 4. But even if A were from one of these types, it is still possible that either no losses occurred or that there were some unclaimed losses. These posterior probabilities provide more information about households like household A than looking at the observed claims data alone ever could.

Now let's consider two households with exactly one claim, B and C. If we were to consider the number of claims alone, we would predict identical loss rates for both of these households. But because the size of the claim for household B is only \$506, B cannot have a large pseudodeductible. That is why B is more likely to have a "low (λ, ψ)" type. C submitted a larger claim, so ψ_C *could* have a "high (λ, ψ)" type, but not necessarily.

The separation in "type" prediction becomes even more interesting when we look at the two-claim households. One might first think that if a household files more than one claim in a six-year period, it would automatically have a high probability of being a "high (λ, ψ)" type. But household E filed at least one relatively small claim, so it cannot be in the highest pseudodeductible class. This restriction would hold even if the second claim from E were extremely high. D and F, on the other hand, did not file small claims, so there is a nonzero probability that these households are of the "high (λ, ψ)" type. Even though households D, E, and F had the same number of claims, the story of "many losses, selectively claimed" is more plausible for D and F, whereas the story of "few losses in the first place" is more plausible for E. It is the size of the *smallest* claim that informs us of which story applies to which household. These patterns apply to households with three or four claims (households G and H) as well.

Table 3 **Severity of Claim**

HH	Claim 1	Claim 2	Claim 3	Claim 4
A	—	—	—	—
B	506	—	—	—
C	34,840	—	—	—
D	8,397	3,879	—	—
E	203	682	—	—
F	1,688	128,928	—	—
G	1,019	1,391	563	—
H	2,590	3,717	2,177	2,010

5.2. Deductible "Upgrades"

If a policyholder's pseudodeductible is higher than the next highest available deductible, one could argue that he could have saved money by taking a policy with the higher deductible (and lower premium). Using the posterior classification probabilities that we described in §5.1, we can compute the posterior expected pseudodeductible for each individual. Not surprisingly, the expected pseudodeductibles for the households with no claims are so large that they always exceed the next highest deductible levels. But pseudodeductibles of households that have had at least one

Table 4 **Posterior Means of Classification Probabilities for Selected Households**

Class	1	2	3	4	5	6	7	8
λ	1.32	1.44	1.92	2.34	2.52	2.7	2.76	3.54
ψ	0	0.27	1.01	1.65	1.77	1.94	3.99	13.64
Prior (%)	24	21	18	<0.5	<0.5	4	9	21
A (%)	26	22	20	2	1	6	9	14
B (%)	31	28	31	1	0	9	0	0
C (%)	17	15	17	1	0	6	11	32
D (%)	8	8	11	1	0	4	11	57
E (%)	50	49	<0.5	<0.5	<0.5	0	0	0
F (%)	21	22	33	2	1	14	8	0
G (%)	17	18	36	3	2	23	0	0
H (%)	9	11	30	4	2	26	17	0

claim tend to have lower expected pseudodeductibles than those with no claims (because a relatively small claim could make higher pseudodeductibles impossible). For example, for household C in Table 4, the posterior expected ψ is 5.97. That household has a $500 policy deductible, so its pseudodeductible is about $3,000. Because household C would not file a claim on a loss below $3,000 anyway, it could have saved money on premiums by taking a policy deductible of $1,000 or $2,000. In fact, we find that 52% of all households with at least one claim (and 80% of all households in our sample) could have saved money in this way by taking a higher policy deductible, with no change in their actual claiming decision.

This poses an interesting policy issue for State Farm, which has commenced a marketing campaign that asks customers to switch from low-deductible to high-deductible policies. The customers with large differences between their policy deductibles and pseudodeductibles might benefit, and State Farm could potentially save money by processing fewer small claims. But the insurer would receive less revenue as deductibles increase. Understanding these trade-offs and how pseudodeductibles affect deductible choices are topics for future research.

6. Incorporating Nonstationarity

While the static multiplicative pseudodeductible version of our model explains the observed distributions of claim counts and severities, as well as the percentage of households filing small claims, it does not explain the remaining two test statistics of interest well: the amount of the increase in claim severity from claim to claim and the percentage of claims that are larger than the previous claim. Naturally, a static model should predict that half of claims are greater than the previous claim with, on average, no increase from claim to claim. However, we observe an average increase of $114 from claim 1 to claim 2. Also, 53% of claims are larger than the previous claim (standard error = 0.076%).

Allowing for a pseudodeductible that evolves after each claim helps explain this phenomenon. Instead of assuming that ψ_h is the same for all claims, we let each ψ_h adjust to a new value after each claim. We can now define a pseudodeductible factor, ψ_{hk} as the pseudodeductible in force for household h immediately *before* the kth claim. So, from the start of the observation period until the time of the first claim, the pseudodeductible is $\Psi(\bar{D}_{h1}, \psi_{h1})$; between claims 1 and 2, the pseudodeductible is $\Psi(\bar{D}_{h2}, \psi_{h2})$; and so forth.

We determine ψ_{hk+1} by multiplying ψ_{hk} by ν_{hk}, a random variable defined on $(0, \infty)$. The "starting" parameter ψ_{h1} is the pseudodeductible factor that is in effect from the beginning of the observation period until the time of the first observed claim. Once household h files its next claim, its pseudodeductible factor is adjusted by a factor of ν_{hk}, so $\psi_{h2} = \nu_{h1}\psi_{h1}$, $\psi_{h3} = \nu_{h2}\psi_{h2}$, and so forth. Because any single draw of ν_{hk} may be either greater than or less than one, this specification of ψ_{hk} allows for pseudodeductibles that can either increase or decrease after each claim. The evolutionary factors ν_{hk} are all drawn from household-specific distributions, similar to the formulation that Moe and Fader (2004a) used to model evolutionary behavior in website visits.

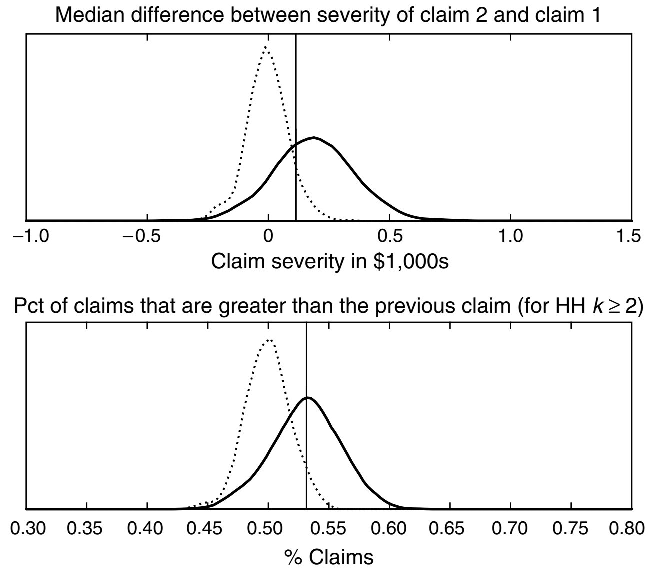
In our case, the ν_{hk} are modeled as being drawn from lognormal distributions with parameters b_h and σ^2, where b_h can be interpreted as the median of the distribution of $\log \nu_h$ for household h, and σ^2 is the variance of $\log \nu_h$ for all households. We incorporate the heterogeneity on b_h by adding an additional dimension to the MDP that was used in the static pseudodeductible model. Thus, just like in the static model, the mixing distribution on the vector (λ_h, ψ_h, b_h) can be described by the L component distributions of (λ_l, ψ_l, b_l) and the L weighting proportions. Each ν_{lk} is then drawn from a lognormal distribution with parameters b_l and σ^2, so each household of type l draws a random vector $(\lambda_l, \psi_l, b_l, \nu_{l1}, \ldots, \nu_{lK_h})$.

We estimated this nonstationary model by allowing for $L = 11$ distinct components on the MDP prior (as in the stationary model, L is simply a chosen upper bound on the number of nonempty components and was chosen in the same manner as described in §3.2). The log-marginal likelihood of the nonstationary multiplicative model is $-3,696$, a dramatic improvement over that of the static model (see Table 1). The quantiles of these components and proportions are described in the online supplement. The quantiles of b_l are transformed from the logarithmic scale, so the values are equivalent to the medians of ν_{li}. Because the medians of the MCMC samples for the medians of ν_{l1} are greater than one, we can say that pseudodeductibles generally increase from claim to claim (this is clearly not true for every household or every claim, because those draws of ν_{li} that are less than one will trigger a decrease in the pseudodeductible). We expect pseudodeductibles to increase for the same reasons that we discussed in §4.3—that customers who have filed claims previously might expect that another claim would lead to premium increases or cancellation of the policy. Decreasing pseudodeductibles may be due to customers who may have been more selective on earlier losses, but on their subsequent losses are more likely to extract a payment from their insurance policies. Another possible explanation is that after a claim, customers adjust their expectations about the costs of filing claims downwards. Regardless, there are dynamics pushing in both directions, but the increasing pseudodeductible appears to be the dominating progression.

Figure 7 illustrates the PPC for the median increase in claim severity from the first to the second claim and the percentage of claims that are greater than the previous claim (for households with $K_h \geq 2$). The vertical lines represent the observed values. The dotted line in each plot is the density of simulated values for the static multiplicative model and the solid line is the density for the nonstationary model. We see that the stationary models are miscalibrated for these inherently nonstationary test statistics, and that the miscalibration is corrected when we add nonstationarity to the model.

This is a particularly important result, because it suggests that observed increases in claim severities may not be caused by an increase in the extent to which customers become more "risk prone" as they file more claims (or become more brazen in filing large claims). Instead, these increases may be caused by customers who become more selective after each claim that they file. For example, consider a household with a starting pseudodeductible of \$800 that files two claims of severities \$1,200 and \$2,500. If one looks only at the claims, one might think that the household is becoming more risk prone, because the severity of the claims is going up. But now suppose that after the first claim, the pseudodeductible increases from \$800 to \$1,500, and that the household experiences a loss of \$1,000 sometime between the two claims. We do not observe the \$1,000 loss, because it is less than the new \$1,500 pseudodeductible. So even though the second claim (third loss) is greater than the first claim, the second loss is *smaller* than the first loss, casting doubt on a hypothesis that increasing claim severity necessarily indicates increasing loss severity. An expectation that the next claim might result in a larger premium (or perhaps cancellation of the policy) could be leading customers to absorb moderately severe losses themselves.

Figure 7 **Posterior Predictive Interval for Nonstationarity of Claims**



Notes. For the median difference between the severities of claim 1 and claim 2, the probabilities that a replicated value exceeds the observed value are 0.076 for the stationary model (dotted line) and 0.501 for the nonstationary model (solid line). For the percentage of claims larger than the previous claim, the probabilities are 0.042 for the stationary model and 0.486 for the nonstationary model.

7. Discussion and Future Research

In this paper, we have shown that by decomposing an observed process into its latent, unobserved subprocesses, one can gain a great deal of insight into consumer behavior, as well as make more accurate predictions regarding the observed data. Some may argue that understanding "how" a transaction moves from unobserved to observed is somehow managerially irrelevant—that just as a purveyor of a packaged consumer goods might care only about modeling sales, and not missed sales opportunities, an insurer cares only about the number and size of claims it is ultimately asked to pay out. But this view ignores the benefits of probability modeling. We model observed consumer behavior as the output of multiple error-laden stochastic processes, allowing us to understand individual behavior despite its sometimes random and unpredictable nature. We argue that a model that is sufficiently complex (but no more complex than that) will offer the practitioner better information about how his customers are making their transaction decisions. For an example of how including transaction opportunities in a model outperforms models where only executed transactions are considered, see Moe and Fader (2004b), who incorporated website *visits* in a model that predicts website *purchases*.

Given that we have presented empirical evidence that there are multiple underlying processes that affect the number and size of claims, an insurer can engage in activities to influence each process separately. For example, suppose an insurer wants to reduce the amount of money paid out in claims, and to do so, he needs to select which customers will not have their policies renewed in the coming year. Traditional models would suggest that the customer with the largest aggregate claim severity would be the least profitable. In fact, the most important theoretical (and practical) implication of this research is that the observed number of claims is the decisive indicator of neither the "riskiness" of a customer nor his propensity to file claims in the future. The customer with a smaller claim may be less risk prone but more likely to file a small claim. The customer with the larger claims may be less likely to file, but might have more unclaimed losses. For each case, this model assigns household-specific posterior probabilities that can be incorporated in other methods of decision analysis.

Thus, customer decisions about whether or not to file a claim will directly impact the aggregate indemnity that the insurer must pay out, making pseudodeductible models a critical component in estimating the customer lifetime value (CLV) of policyholders. Most CLV models incorporate the cost of goods in some form, but for most products (e.g., packaged goods), those costs are known at the time of the transaction (Berger and Nasr 1998, Gupta et al. 2004). In contractual settings, however, the full cost is not known until the customer uses the service (Fader and Hardie 2006). The pseudodeductible may also play a part in the revenue side of CLV calculations. In §5.2, we mentioned that although customers with high pseudodeductibles could benefit by increasing their policy deductibles, it is unclear whether such a move is in an insurer's interest, because there is a trade-off between savings from processing small claims and the reduction in premium revenue that comes with selling high-deductible policies. Also, as our nonstationary pseudodeductible model suggests, the filing of a claim makes the filing of future claims less likely. Therefore, an insurer might be better off keeping a recent claim-filer as a customer, as opposed to canceling the policy altogether, because that customer pays a higher premium *and* will be more selective about filing claims going forward. Of course, the removal of the cancellation threat might alter the nature of the pseudodeductible nonstationarity, hence the need for more research to understand these trade-offs as well. Additionally, we are interested in whether the filing of claims influences the likelihood that a customer would change his deductible from one year to the next, or cancel his policy altogether.

Notwithstanding the managerial usefulness of this model in and out of the domain of financial services, this framework could be used to improve our understanding about why insurance customers behave the way they do. We did not set out to predict *ex ante* which customers have the larger pseudodeductibles, but one could adapt our model to make the pseudodeductible a function of explanatory variables. For example, it might be interesting to see whether the cause of a loss (e.g., natural factors or customer carelessness) alters the propensity of the customer to file a claim. We envision this paper as the start of a potentially fruitful stream of research into better understanding about how and why customers "leave money on the table."

An online supplement to this paper is available on the *Management Science* website (http://mansci.pubs.informs.org/ecompanion.html).

Acknowledgments
The authors thank Rachel Croson, Fred Feinberg, Jean Lemaire, Bob Meyer, Alexander Muermann, Stephen Shore, Uri Simonsohn, and participants in research seminars at Massachusetts Institute of Technology and the Universities of Maryland, North Carolina, Virginia, and Pennsylvania for their helpful comments and suggestions. Support from the Wharton Risk Management and Decision Processes Center is gratefully acknowledged. Michael Braun also acknowledges support of a Wharton doctoral fellowship at the University of Pennsylvania. In addition, this research would not have been possible without the involvement of State Farm Fire and Casualty Company, and in particular Jim Ament, David Cummings, Gary Grant, Robin Hawkins, and Dennis McNeese.

References

Antoniak, C. E. 1974. Mixtures of Dirichlet processes with applications to nonparametric problems. *Ann. Statist.* **2**(4) 1152–1174.

Berger, P. D., N. I. Nasr. 1998. Customer lifetime value: Marketing models and applications. *J. Interactive Marketing* **12**(1) 17–30.

Bernardo, J. M., A. F. M. Smith. 2000. *Bayesian Theory*, 2nd ed. John Wiley and Sons, Chichester, UK.

Bureau of Labor Statistics. 2004. Consumer price index. Retrieved September 23, 2004, http://www.bls.gov/cpi, series ID cuur0200sa0, cuus0200sa0.

Choi, J. J., D. Laibson, B. C. Madrian. 2005. $100 bills on the sidewalk: Violations of no-arbitrage in 401-k accounts. Working Paper 11554, NBER, Cambridge, MA.

Congdon, P. 2001. *Bayesian Statistical Modelling*. Wiley Series in Probability and Statistics, John Wiley and Sons, Chichester, UK.

De Leve, G., P. J. Weeda. 1968. Driving with Markov-programming. *ASTIN Bull.* 5(1) 62–86.

de Pril, N. 1979. Optimal claim decisions for a bonus-malus system: A continuous approach. *ASTIN Bull.* **10** 215–222.

Dellaert, N. P., J. B. G. Frenk, L. P. van Rijsoort. 1993. Optimal claim behavior for vehicle damage insurances. *Insurance: Math. Econom.* **12**(3) 225–244.

Draper, D. 1999. Discussion on "Bayesian Nonparametric Inference for Random Distributions and Related Functions." *J. Roy. Statist. Soc., Ser. B* **61**(3) 510–513.

Escobar, M. D., M. West. 1995. Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**(430) 577–588.

Fader, P. S., B. G. S. Hardie. 2000. A note on modelling underreported Poisson counts. *J. Appl. Statist.* **27**(8) 953–964.

Fader, P. S., B. G. S. Hardie. 2006. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. Working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Ferguson, T. S. 1983. Bayesian density estimation by mixtures of normal distributions. H. Rizvi, J. Rustagi, eds. *Recent Advances in Statistics*. Academic Press, New York, 287–302.

Gelman, A., X.-L. Meng, H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6**(4) 733–807.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Rubin. 2004. *Bayesian Data Analysis*. Chapman and Hall, Boca Raton, FL.

Geweke, J. F. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds. *Bayesian Statistics*, Vol. 4. Oxford University Press, Oxford, 169–194.

Gupta, S., D. R. Lehmann, J. A. Stuart. 2004. Valuing customers. *J. Marketing Res.* **41**(1) 7–18.

Haehling von Lanzenauer, C. 1974. Optimal claim decisions by policyholders in automobile insurance with merit-rating structures. *Oper. Res.* **22**(5) 979–990.

Hardie, B. G. S., P. S. Fader. 2005. Applied probability models with applications in marketing. Unpublished manuscript.

Hastings, N. A. J. 1976. Optimal claiming on vehicle insurance. *Oper. Res. Quart.* **27**(4) 805–813.

Israel, M. 2004. Do we drive more safely when accidents are more expensive? Identifying moral hazard from experience rating schemes. Working paper, Northwestern University, Evanston, IL.

Johnson, N. L., S. Kotz, N. Balakrishnan. 1994. *Continuous Univariate Distributions*, Vol. 1. John Wiley and Sons, New York.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–291.

Kahneman, D., A. Tversky. 1984. Choices, values and frames. *Amer. Psychologist* **39**(4) 341–350.

Kim, J. G., U. Menzefricke, F. M. Feinberg. 2004. Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Rev. Marketing Sci.* **2**(1) 1–39.

Klugman, S. A., H. H. Panjer, G. E. Willmot. 1998. *Loss Models: From Data to Decisions*. John Wiley and Sons, New York.

Kulkarni, V. G. 1995. *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, London, UK.

Lemaire, J. 1995. *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Moe, W. W., P. S. Fader. 2004a. Capturing evolving visit behavior in clickstream data. *J. Interactive Marketing* **18**(1) 5–19.

Moe, W. W., P. S. Fader. 2004b. Dynamic conversion at e-commerce sites. *Management Sci.* **50**(3) 326–335.

Moffitt, Robert. 1983. An economic model of welfare stigma. *Amer. Econom. Rev.* **73**(5) 1023–1035.

Newhouse, J. P., J. E. Rolph, B. Mori, M. Murphy. 1980. The effect of deductibles on the demand for medical care services. *J. Amer. Statist. Assoc.* **75**(371) 525–533.

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

Norman, J. M., D. C. S. Shearn. 1980. Optimal claiming on vehicle insurance revisited. *J. Oper. Res. Soc.* **31**(2) 181–186.

Rubin, D. B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12**(4) 1151–1172.

Schmittlein, D. C., A. C. Bemmaor, D. G. Morrison. 1985. Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Sci.* **4**(3) 255–266.

Sethuraman, J. 1994. A constructive definition of Dirichlet priors. *Statist. Sinica* **4**(2) 639–650.

Spiegelhalter, D. J., A. Thomas, N. G. Best, W. R. Gilks. 1996. BUGS: Bayesian inference using Gibbs sampling. Computer software, version 1.4. Retrieved 2004, http://www.mrc-bsu.cam.ac.uk/bugs/.

Stephens, M. 2000. Dealing with label switching in mixture models. *J. Roy. Statist. Soc., Ser. B* **62**(4) 795–809.

Thaler, R. H. 1980. Toward a positive theory of consumer choice. *J. Econom. Behav. Organ.* **1**(1) 39–60.

van Praag, B. M. S., E. M. Vermeulen. 1993. A count-amount model with endogenous recording of observations. *J. Appl. Econometrics* **8**(4) 383–395.

Venezia, I. 1984. Aspects of optimal automobile insurance. *J. Risk Insurance* **51**(1) 63–79.

Venezia, I., H. Levy. 1980. Optimal claims in automobile insurance. *Rev. Econom. Stud.* **47**(3) 539–549.

Walker, S. G., P. Damien, P. W. Laud, A. F. M. Smith. 1999. Bayesian nonparametric inference for random distributions and related functions. *J. Roy. Statist. Soc., Ser. B* **61**(3) 485–527.