CLAIMS AND DEDUCTIBLES FOR HOMEOWNER'S INSURANCE

Michael Braun

A DISSERTATION

in

Operations and Information Management

For the Graduate Group in
Managerial Science and Applied Economics

Presented to the Faculties of the
University of Pennsylvania
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2006

Supervisor of Dissertation

Supervisor of Dissertation

Graduate Group Chair

UMI Number: 3225432

This dissertation is dedicated to my family and friends who supported me along the way.

You know who you are.

ABSTRACT

CLAIMS AND DEDUCTIBLES FOR HOMEOWNER'S INSURANCE

Michael Braun

Supervisors: Peter S. Fader and Howard Kunreuther

This dissertation examines the interplay between claims and deductibles in homeowner's insurance. In the first essay, I examine the effect of marketing cues on asymmetric information. Economic theories of asymmetric information predict that customers who are more likely to file claims during the coverage period will choose lower deductibles on their policies. One of the challenges in testing this theory is that customers may exhibit inertia and not consider their insurance options for an extended period of time. Furthermore, the ultimate "trigger" for a deductible change could be a factor unrelated to that customer's expected claim propensity, such as a pricing change or marketing activity. Hence, the rate at which a customer files a claim after a deductible change may be related to why the deductible change was made at all.

In the second essay, I introduce the concept of the "pseudodeductible," a latent threshold above the policy deductible that governs the propensity (or reluctance) of a homeowner to file a claim when the underlying loss is covered by insurance. I show how the observed number of claims can be modeled as the output of three stochastic processes: the rate at which losses occur, the size of each loss, and the choice of the individual to file or not file a claim. Using mixtures of Dirichlet processes to capture heterogeneity, I uncover several relevant "stories" that underlie the frequency and severity of claims. For instance, some customers have few losses, but all are filed as claims, while 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.

Acknowledgements

I could never have completed this dissertation alone. My advisors, Pete Fader and Howard Kunreuther, contributed more of their time, effort and patience than anyone could have expected. Eric Bradlow found the perfect way to balance support and encouragement with honest feedback and constructive criticism. And Stephen Shore and Alex Muermann taught me everything I know about insurance markets and economic modeling. I am a better researcher and richer person for having worked with this first-class committee.

Other individuals who contributed either directly or indirectly to this research include David Croson, Rachel Croson, Fred Feinberg, Jean Lemaire, Bob Meyer and Uri Simonsohn. Paul Kleindorfer got me started in the field of decision processes early in my doctoral career at Wharton. Jim Smith and Kevin McCardle introduced me to the field of decision analysis even earlier, during my time as an MBA student at Duke.

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.

Support from the Wharton Risk Management and Decision Processes Center is gratefully acknowledged. I also thank the Wharton School of the University of Pennsylvania for providing me with four years of fellowship support.

And, of course, I thank my parents, Linda and Barry Braun, for all of their love and support during an academic career that began long ago and will never, ever end.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this dissertation I investigate the interplay between claims and deductibles in homeowner's insurance. The approach is interdisciplinary, drawing from concepts developed in actuarial science, statistics, marketing, psychology, operations research and economics. While researchers in each of these disciplines have addressed specific elements of decision-making for insurance customers, there has been limited observational empirical research on the choices that individual policyholders make regarding how much insurance they purchase and how expensive these customers are to service. My goal is to address individual behavioral and economic issues that influence both the cost and revenue elements of the insurance value chain. In particular, I am interested in the role of asymmetric information in deductible choice, and the propensity of customers to file incurred insured losses as claims.

Consider the following example. Suppose a homeowner insures his home, valued at $300,000, against losses due to typical covered perils (such as weather, accident, theft or fire, but not flood, earthquake or personal liability). This policy has a $1000 deductible, so the homeowner is responsible for the first $1000 of a loss, with the remainder eligible to be paid by the insurer. The duration of an insurance policy is one year, and can be renewed annually if both the customer and insurer choose to do so. In exchange for this coverage of risk, the homeowner pays an annual premium of $750 to the insurer.

There are several decisions that this customer has made–or will make. First, at some point in the initial policy year, the customer decides to purchase an insurance policy with a defined deductible amount. In each subsequent year that the customer retains his policy, he could retain his current deductible or switch to a new one. Furthermore, in the years since the initial purchase of the insurance policy, the customer may experience one or more insurable losses, and may decide to file a claim for reimbursement for none, one, some, or all of them. In exchange for this insurance coverage, the customer agrees to pay an annual premium that is determined by the insurer, and is based on the amount of claims that the insurer expects the customer to file during that year. The insurer's expectations for the cost to cover this customer may be influenced by a number of factors, such as geographical location of the home, the value and type of construction of the home, the customer's claims history, and the customer's chosen deductible level. Consequently, since the claims history influences the premium that the customer will pay in the future, the customer may not request reimbursements to which he may otherwise be entitled. Thus, the decisions about how much insurance to purchase, and how to utilize the insurance coverage once it is in force, are intertwined.

1.1 Deductible choice and asymmetric information

Each chapter in this dissertation focuses on a different way in which deductible and claims decisions are related. In the first section of this dissertation, I examine some behavioral and economic determinants of deductible choice for homeowner's insurance. Rather than model the decision as the choice among a set of deductible options, I concentrate on the decision of a customer to increase his deductible level from one year to the next. There are many different reasons why a customer might increase his deductible. One is related to budget constraints, either actual or perceived, that might induce customers to reduce their insurance premium expenditures through a deductible increase. Another is that the customer exploits private information about his own propensity to file claims during the upcoming policy year, and increases his deductible accordingly.

This research addresses mental budget constraints by looking at the impact of pricing cues on customers' propensities to increase their deductibles. I find that customers who are faced with premium increases from the previous year are more likely to increase their deductibles, even after controlling for the amount of reduction in premium that the deductible increase offers. The underlying idea is that if customers segregate insurance expenditures into a separate mental account, then customers may need to take some action to ensure that the account does not become "overdrawn".

My test for the presence of private information is based on one proposed by Chiappori and Salanie (2000), who posit that a negative correlation between the amount of insurance purchased and the ex-post incidence of claims is an indicator of information asymmetries in the market. In many cases, however, studies that use this test have been limited by any of three factors:

1. the confounding effect of customer inertia when choosing deductibles in sequential years;
2. the use of only a single observation per household (which precludes examination of within-household correlations); or
3. endogeneity between the deductible chosen by a customer and the manner in which deductible options are presented to the customer.

My results are more convincing results because I exploit the *dynamics* of the deductible choice decision from year to year. Instead of looking at the correlation between the claims and the size of the deductible, I focus on the relationship between claims and the incidence of deductible increases. Because I observe claim-switch pairs for each household in multiple years , I can draw inferences using within-household correlations, which avoids the endogeneity problem. Furthermore, the model conditions the incidence of deductible increases on factors that might knock a customer out of an inertial state. Indeed, in this sample of homeowner's insurance policies, I find that there is a negative correlation between the incidence of filing claims and the incidence of increasing ones deductible in that year. This is exactly what one would expect; riskier customers who anticipate filing claims are less likely to reduce the amount of reimbursement they would receive after filing those claims.

However, my interest in asymmetric information in insurance markets is not as much on its presence or absence, but more on whether the factors that trigger customers to switch their deductibles might also affect the risk types of the customers who do switch. For example, a customer who increases his deductible could be someone who is either low risk or high risk, conditional on the information used by the insurer to price the policy. If a high deductible policy is priced by the insurer to be profitable for low-risk customers, then it may be less profitable if it is selected by too many high-risk customers (since the claims paid out on that policy may exceed the premium revenues that the insurer collects). If a premium increase, or some other marketing cue, triggers too many high-risk customers to switch, relative to the number of low-risk customers, then these triggers may have unintended negative consequences for the profitability of the insurer.

1.2 The "Pseudodeductible"

In the second section of this dissertation, I shift attention to the underlying processes that generate claims on insurance policies. Once a customer chooses a deductible, he may experience some losses that are covered by insurance. If the amount of the loss (the severity) is less than the deductible that is specified in the policy, then the customer certainly receives no reimbursement. But if the loss is greater than the policy deductible, then the customer has a choice: file a claim on that loss, or forgo the reimbursement to which he is entitled. It is not immediately obvious that a customer should file claims on all eligible losses, since filing a claim in one period could affect premiums in subsequent periods.

I 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. If the severity of a loss exceeds the pseudodeductible, then the loss is observed as a claim; otherwise, it remains unobserved to the insurer (and the researcher). Thus, 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 either file or not file a claim. Estimates of an individual's loss rate can be considered to be an inference about that customer's riskiness, while the estimate of the pseudodeductible estimates that individual's propensity to convert losses into claims. Such inference would not be possible in models that do not allow for the possibility of pseudodeductibles.

Using mixtures of Dirichlet processes to capture heterogeneity and the interplay among the three processes, one can 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, while 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. In addition, I find that for a majority of customers, the posterior expected pseudodeductible is larger than the next highest deductible level that was available to the customer. This result suggests that customers may be selecting deductibles that are too low, since customers could have saved money on premiums by taking a higher deductible.

1.3 Data

The research for both dissertation sections utilizes a data set provided by State Farm, the largest underwriter of homeowner's insurance in the United States. This data set includes information on over one million policies that were in effect for single-family homes in six midwestern states from 1998 to 2004. Each record includes geographical information (e.g., county and zip code); policy information (e.g., coverage limits, deductibles and premiums); and claims history (including dates, amounts and reasons for claims). Records are taken as snapshots on December 31 of each of the years in the observation period. The data are proprietary to State Farm and were provided to me solely for use in this research. State Farm uses these data for critical operations (such as the disbursement of claims to policyholders), so I am extremely confident in their quality and accuracy.

1.4 Implications

Both sections of this dissertation, taken individually and together, offer insights into customer behavior that can assist insurers in making both strategic and tactical decisions. For example, in a model of asymmetric information in insurance markets, customers who increase their deductibles might be either conditionally low-risk types or conditionally high-risk types. There may be several ways to induce customers to increase their deductibles (increasing premiums is one such method). But what type of customers switch in response to each of the many kinds of marketing tactics? If too many high-risk customers take high-deductible policies, the profitability of those policies could be adversely affected. These implications are discussed in greater detail in Chapter 2.

Similarly, once a customer chooses his deductible, what will his claiming behavior be? Does deductible choice effect the rate at which customers experience losses, or the rate at which they file claims on those losses. This is the fundamental implication of Chapter 3. One way insurers assess the riskiness of customers (and their propensity to file future claims) is to look at the customer's history of claims (or experience rating). Indeed, this is exactly how I assess riskiness in Chapter 2. But riskiness can take different forms: exposure to losses and selectivity in filing claims. Given a customer's claims history, one can discern which definition of risk is most appropriate for that customer. For example, suppose a customer did not file any claims during the previous year. This result may have occurred because the customer did not experience any insurable losses, or because he did experience some losses, but decided not to seek reimbursement. These outcomes reflect different underlying behavioral processes that the insurer could influence separately. They also reveal some surprising observations about customer choices, especially related to the size of a customer's pseudodeductible relative to his policy deductible. The goal of this dissertation is to establish a framework that allows both customers and insurers to understand these issues better.

Chapter 2

The Effect of Cues on Asymmetric Information in Homeowner's Insurance Markets

2.1 Introduction

Markets for personal-lines insurance, such as homeowner's insurance, offer a fertile field on which to study economic hypotheses about asymmetric information. Insurers sell policies to customers without observing the customer's true probability of filing a claim; instead, they must infer that probability from observed data. The insurer's estimate of that probability is a function of insurance policy features (e.g., deductible) and observable information about the customer (e.g., claim history, geographical location). While insurers typically have the most comprehensive sources of data to estimate these probabilities, customers often have additional information (e.g., the customer's attention to safety) that the insurer can only partially observe, or not observe at all.

One would expect customers to exploit this informational asymmetry when choosing the amount of insurance to purchase. This informational asymmetry in insurance markets typically takes one of two forms. Adverse selection suggests that customers with private information that they are inherently high-risk will choose to purchase more insurance than those who have private information that they are low-risk (Rothschild and Stiglitz, 1976). Moral hazard suggests that customers who have purchased more insurance will choose to invest less in safety when that safety investment is private information. Both moral hazard and adverse selection imply that after conditioning on variables used in setting premiums, the chosen deductible and the probability of filing a claim will be negatively correlated (Chiappori, 2000; Chiappori and Salanie, 2000, 2003).

One challenge in testing this prediction using data on homeowners' insurance is that customers are inertial. In the State Farm data set (which is described in section 1.3), about 10% of customers who have a choice of remaining with their current deductibles still switch their deductibles in a typical year, and nearly all of these switches are increases. One possible explanation of this fact is that customers always rationally choose the optimal deductible level, implying that they rarely receive new private information about their risk type that would trigger a deductible increase, or that such new private information is nearly always "good" news. Alternatively, customers may face high switching costs or may only occasionally consider the deductible choice problem at all. In this case, the decision to switch deductibles may have as much to do with external cues (e.g., pricing) that encourage switching (or at least thinking about the deductible choice problem at all) as with the arrival of new private information. For example, for this data set, switch rates increased from 8% to 22% in 2003, when customers received a bill message saying that they could save money on premiums by increasing their deductibles.[1]

[1]All State Farm customers in our subject market who could save money in the short term on premiums by increasing their deductibles received a message on thier 2003 renewal notices. The text of a typical message was: "Changing your policy deductible to 1% of your dwelling coverage amount will impact your annual premium by $601 for a premium of $807. Please note, higher deductibles mean lower premiums but may reduce our payment to you for a covered property claim. Please refer to the enclosed insert for details."

This inertia, and the cues that knock people out of it, could have an important impact on relationship between risk type and deductible choice by encouraging people to act or not act on their private information. Consider a standard Rothschild-Stiglitz (RS) competitive equilibrium without cues or inertia in which customers must decide whether or not to switch to a higher deductible. The customers who receive new private information that they are lower risk (L-types) will switch to a higher deductible; the customers who receive private information that they are higher risk (H-types) will not switch. In the presence of inertia, however, these L-types may not switch at all. Cues, such as premium increases or bill messages, may then increase the role of informational asymmetry in the market by inducing L-types (who would switch in the absence of inertia) to switch. Alternatively, cues may reduce the role of informational asymmetry in the market by pushing H-types (who would not switch in the absence of cues) to switch along with the L-types. This paper asks whether cues (and which ones) increase or decrease the role of informational asymmetry in markets. In the data, this is analogous to looking for factors that increase or decrease the conditional correlation between switching deductibles and filing a claim.

This problem is important for two reasons. First, from an economic perspective, empirical observations can highlight the role of individual behavior in either amplifying or attenuating theoretical predictions. For example, one cue that I will investigate in this chapter is the increase in premiums from year to year. The effect of this cue on switching behavior can be explained by theories of mental accounting (Thaler, 1980, 1999) or mental budgeting (Heath and Soll, 1996), since customers may segregate the effect of paid premiums from the effect of potential indemnity in case of a covered loss. Second, from a marketing perspective, the types of customers that switch in response to a cue may or may not be the ones whose switching is most profitable. For example, if a high-deductible policy were priced with a premium that breaks even for low-risk customers (as would be the case in the competitive equilibrium story of RS), then a cue that induces too many high-risk customers to switch (e.g., a premium increase that the customer tries to mitigate with a deductible switch) could be a money-loser for the insurer. In this chapter I do not make any judgments about which cues are necessarily profitable or unprofitable, since one does not know which policies are designed to be profitable for which customers. Rather, I develop additional insight into how these cues could induce behavior that might not be desirable for the insurer.

The State Farm data set includes information on filed claims, coverage limits, deductible amounts and premiums, from 1999 to 2004. Because I observe policy characteristics and customer claims behavior for multiple periods, I can draw inferences about the within-household correlation between switching and claiming. This is a particularly important feature of the data, since by examining behavior of individual households over time, I avoid certain endogeneity problems that might be caused by the self-selection of customers into policies according to their risk types (Chiappori and Salanie, 2003). By looking at within-household correlations between switching and claims behavior, I can observe how these correlations (and, of course, the switching and claims behavior independently) respond to exogenous cues from year to year. If I did not have a dynamic data set with multiperiod observations, then I could only relate claims to existing deductible levels (rather than switching behavior), which, as discussed above, may be confounded by customer inertia or other biases in the initial selection of deductible (Abbring et al., 2003a,b).

The standard test for asymmetric information is whether or not "conditionally on all information that is available to the insurance company, is the choice of a particular contract correlated to risk, as proxied ex post by the occurrence of an accident?" (Chiappori and Salanie, 2003). I operationalize this test by examining the within-household correlation between the propensity for a household to increase its deductible (which I generally call "switching") and the propensity of that household to file a claim in the subsequent policy year. There are four main results. First, I find that pricing cues influence deductible switching behavior in a way that is consistent with theories of mental accounting (Thaler, 1980, 1999) and mental budgeting (Heath and Soll, 1996). Second, I reveal a negative correlation between switching and claims, suggesting that the conditionally low-risk customers are the ones who, on average, increase their deductibles. Third, I find that pricing cues, such as the amount of money a customer who increases his deductible can save on premiums, increase the within-household correlation and induce more conditionally high-risk customers to switch. Finally, I find that informational asymmetries are more pronounced for conditionally high-risk customers.

Note that I am testing the effects of cues on asymmetric information in general, and not distinguishing between adverse selection and moral hazard. For the purpose of clarity, I use the terms "low risk" and "high risk" (or L-type and H-type) to differentiate between two types of customers. These labels are consistent with the adverse selection literature in general, and with RS in particular. Under moral hazard, customer types are not risk types, but rather the responsiveness of customers' willingness to invest in safety (or other precautionary measures) as a result of the amount of insurance they have purchased. Empirically, the manifestation of the actions of a customer who has a low deductible and then files a claim (e.g., a customer who bought a lot of insurance and behaved "unsafely" as a result) might look the same as a customer who, knowing he is a conditionally high-risk type, purchases the policy with the low deductible. I am essentially agnostic to which of these two stories to tell, so I pick one nomenclature–that used in the adverse selection literature–to describe customer types. Also, note that this paper is not specifically testing the RS story per se, since the market under study is probably not perfectly competitive and there may be more than two types of customers in the market. Nevertheless, the RS framework provides some theoretical background that highlights the impact and importance of these results.

With this clarification in hand, I proceed to show how pricing cues influence asymmetric information in insurance markets. First, in section 2.2, I discuss key contributions already made to the multiple disciplines that this research builds upon. In section 2.3, I present the formal empirical model for testing these hypotheses. In section 3.4 I present and discuss the results, and in section 2.4, I present some policy implications and prescriptive recommendations.

2.2 Literature Review

There are two types of questions that I address in this research: the external factors that trigger deductible switching, and whether or not these factors influence informational asymmetries in the market for homeowner's insurance. In this section, I discuss some of the previous research on these two key areas.

2.2.1 Deductible Choice

Economists have studied the deductible choice question in the context of individuals who are expected utility maximizers with von Neumann-Morgenstern (vN-M) preferences. The simplest models assume, for example, that customers face losses that have known probabilities of occurring and that the premium is calculated as a fixed percentage above the expected value of the distribution of that loss. Using this framework, Mossin (1968) and Schlesinger (1981) computed optimal deductibles for customers facing this proportionally-determined premium. In particular, they demonstrate that for policies that are "fairly priced," (whose premiums equal the expected value of the loss), it is optimal for customers to take a policy with no deductible (if such a policy were even available). Furthermore, they show that there exists some premium, higher than the fairly priced one, for which a customer will not purchase any insurance at all. Models that incorporate additional complexity into the deductible choice model will yield different results. For example, Doherty and Schlesinger (1990) show that customers with constant absolute risk aversion will take higher deductibles as their belief that the insurer will default on payments increases.

However, as more and more researchers recognize that customers often deviate from the restrictive assumptions of EU maximization, there has been greater interest in formulating optimal deductible choice strategies for customers who do not adhere to the vN-M axioms. Machina (2000) proved that many of the predictions of deductible choice that would ordinarily require adherence to the vN-M axioms will still hold when some (such as linearity in probabilities) are relaxed, as long as the individual is risk averse. Braun and Muermann (2004), examining the impact of regret on the demand for insurance, apply Regret Theory (Loomes and Sugden, 1982; Bell, 1982) to show that, depending on the premium loading and degree of regret aversion, customers who are regret averse may demand larger or smaller deductibles than those who are not. However, the predictions implied by Dual Theory (Yaari, 1987) suggest that customers will take either full insurance or no insurance at all, with no preferences for intermediate deductible levels (Doherty and Eeckhoudt, 1995). Hence, even in an analytical exercise, optimal deductible choices are quite sensitive to the assumptions that one makes about individual choice behavior.

Attempts to model the choice of deductible using observational data have been sparse. Pashigian et al. (1966) used aggregated data on auto insurance policies to show that while the vast majority of customers chose low deductibles for their auto insurance, these choices are inconsistent with expected utility theory. More recently, Grace et al. (2003) attempt to estimate overall demand for catastrophic insurance in Florida and New York, but they assume that both the insurers and customers are expected utility maximizers, and that the available deductible is a continuous, rather than discrete, measure. There have been more attempts to examine insurance decisions experimentally. In fact, the original articles on Prospect Theory (Kahneman and Tversky, 1979) and Regret Theory (Bell, 1982) use insurance-related experiments to lend support to those theories. Slovic et al. (1977) show that individuals are more likely to purchase insurance for a high-probability/low-severity event than for a low-probability/high-severity event with the same expected value. Hsee and Kunreuther (2000), while investigating the role of affect on insurance decisions, show that subjects who are primed to feel some affection to a good are more likely to purchase insurance on that good. Framing of deductibles matters as well; Johnson et al. (1993) demonstrated that subjects are less likely to purchase an insurance policy with a deductible than an actuarially identical policy that instead pays a rebate.

In the research in this chapter, one of the more intriguing observations is the relatively low number of customers change their deductible in a given year. This behavior is consistent with status quo bias (Samuelson and Zeckhauser, 1988), which refers to the tendency of customers to remain at some position, state or choice, rather than act and move to another state. Such customer inertia has been observed previously in insurance markets. Johnson et al. (1993) studied insurance customers in Pennsylvania and New Jersey who were faced with a choice of the "full tort" or "limited tort" option on their auto insurance. This study found that customers tended to take the option that was the default ("do nothing") option in their particular state. Status quo bias can be explained by loss aversion (Kahneman et al., 1991) and is typically thought to be closely related to the endowment effect (Kahneman et al., 1990).

There are many reasons why customers might exhibit inertia in their deductible choices, and why these customers may retain low deductibles for a number of years. It is possible that customers are either not reevaluating their insurance choices annually, and if they are, there may be no immediate reason for them to change their choices even if they were behaving rationally. Furthermore, customers might be "shaken out" of inertial behavior when their experience certain events, such as a premium increase or the filing of an insurance claim. In addition, customers may experience myopia regarding the probability that certain events might occur in the future, and not adjust their deductibles until after the fact (Kunreuther and Pauly, 2005).

So, if insurance customers, in general, evaluate their deductible options regularly, are their external factors or cues that might stimulate them to do so? One such cue might be the cost of the insurance itself, and how that cost changes from year to year. Consider the story of a customer who purchases an insurance policy that offers \$300,000 in coverage with a \$500 deductible for a premium of \$800 a year. Suppose this customer lives in an area that is experiencing an (exogenous) increase in crime, and therefore his premium rises to \$1000 in the next year. In response to this \$200 increase, the customer calls his insurance company, who suggests an increase in deductible from \$500 to \$1000 that would result in a premium of \$900. The customer takes the insurer's suggestion and raises his deductible, saving himself money in the short term, but exposing himself to greater exposure in the event of a loss in the future. In this case, the deductible increase is triggered by both the premium change he would face by retaining his current deductible, and the amount of money that he could save by increasing his deductible. Of course, the potential for saving money on premiums alone, in the absence of a premium increase on the current deductible, might be sufficient to trigger a deductible switch.

Regardless, the reaction of the customer to the premium change can be explained if customers consider insurance premiums to be an expense, and process these expenses separately from other gains on losses that they might accrue or incur. Mental budgeting (Heath and Soll, 1996) implies that additional expenses in a budget for a particular item are resisted once total expenditures for that item get too large. So, if customers think of insurance premiums as a separate budget category, then if insurance expenses grow much larger than some reference point or budget threshold, the customer will look for ways to reduce them. Reductions in insurance premiums could come from either an increase in the deductible, or possibly defection to an alternative insurer who offers lower premiums. Another cue that might trigger a deductible increase is the opportunity for the customer to reduce his premium, even if the premium at the current deductible level does not go up. This story is more consistent with mental accounting (Thaler, 1980, 1999) if the customer emphasizes the gains from premium savings more than the additional incremental loss that he might incur in case of a covered event.

2.2.2 Asymmetric information

The classic theory of asymmetric information in insurance markets is the adverse selection story proposed by Rothschild and Stiglitz (1976). The RS model assumes that there are high-risk (H-type) and low-risk (L-type) customers that can choose an insurance contract that offers either full insurance (no deductible) or partial insurance (positive deductible). The market is assumed to be perfectly competitive, so the insurer prices these policies at the break-even point. If there were complete information in the market, then insurers would observe the risk types of the customers, and offer a distinct contract to each type of customer. However, if asymmetric information exists, then only the customer knows his own risk type. The insurer could, potentially, offer a single insurance contract to all customers; this is what would happen under a pooling equilibrium. However, under the assumptions of the RS model, only a separating equilibrium exists, and the insurer will offer policies with different deductibles. Under this equilibrium, H-type customers will select the no-deductible policy and L-type customers will select the deductible policy.

Of course, these assumptions are quite restrictive. There many be many different types of customers in the market, and insurers may offer more than two deductibles (and, for the customers in this data set, they do). Wilson (1977) presents a model that does admit a pooling equilibrium, in which firms anticipate the behavior of competitors and customers. Grossman (1979) describes a dissembling equilibrium in which H-type customers have an incentive to act like L-types, since insurers might deny coverage to customers who signal that they are high-risk. Furthermore, the market may not be perfectly competitive, so the profitability of a policy with any particular deductible is unknown.

The data that I am using in this dissertation comes from a world that is *not* consistent with the assumptions of RS. The homeowner's insurance market is not perfectly competitive due to state regulation of premiums, and customers may engage in types of decision-making (such as multiperiod learning) that are not expressly considered by RS. Nevertheless, the RS model provides a framework to discuss what happens when customers switch their deductibles. Suppose that the market is at equilibrium, and that H-types and L-types select low deductible and high deductible policies respectively. Then, when a customer increases his deductible, one possible story is that an H-type customer with a low deductible has received new information about his own type (and, as such, an equilibrium condition is maintained), or that some cue (pricing, behavioral or otherwise), may have triggered the H-type customer to make the deductible switch, jarring the system out of equilibrium. In fact, even if the customer's true type does not change, but the customer misperceives his own risk type, the customer may switch to the higher deductible that is inconsistent with the true risk type (Kleindorfer and Kunreuther, 1983). The alternative story is that the system is already out of equilibrium, with some L-type customers in the policy that was designed for the H-type customers, and that the switch moves the L-type customer to the equilibrium policy with the higher deductible. In short, when an insurer observes a customer increase his deductible, that customer could be either a conditional H-type or a conditional L-type. Of course, this story could apply to a wide range of scenarios, and not just the ones that adhere to the assumptions of the RS model.

Although risk types are not observed directly, one can use the ex-post incidence of claims as an approximation to the risk level in a given year. If a customer files a claim, he is more likely to be a high-risk customer than a customer who does not file a claim. Thus, one would expect that customers who increase their deductibles would be less likely to file claims, since the economic theory predicts that these switchers are more likely to be low-risk types. The examination of the correlation between deductible choice and claims incidence forms the basis of the standard tests for asymmetric information in insurance markets (Chiappori and Salanie (2003) offer a detailed survey). Chiappori and Salanie (2000) propose three methods to measure this correlation: (1) the correlation between the residuals from independent probit estimations of deductible choice and claims incidence; (2) the correlation coefficient from a bivariate probit estimation (which is the method I use); and (3) a χ^2 test for independence of the number of customers choosing each deductible with the number of claims. They then test for asymmetric information in this way using data from a French underwriter of automobile insurance, and find, as expected, that there is a negative correlation between deductible and claims incidence. Puelz and Snow (1994) and Dionne et al. (2001) also find evidence of asymmetric information in automobile markets, and Finkelstein and McGarry (2003) and Finkelstein and Poterba (2004) find evidence of asymmetric information in annuity markets. However, Cawley and Philipson (1999) do not find such evidence in their study of a life insurance market. To my knowledge, there is no research on the presence of informational asymmetries in the market for homeowner's insurance.

A common thread across all of these papers is their emphasis is that, when using the correlation between the amount of insurance purchased and the ex-post claims incidence, it is essential to include all information that is observable to the insurer that is used when setting premiums (Finkelstein and Poterba, 2004). This information could be at the market level (such as whether or not the home is in a high-crime area), or at the individual level (such as a customer's experience rating or claims history). Otherwise, the correlation will reflect not only asymmetric information, but also other factors that are related to both deductible choice and risk type (Finkelstein and McGarry, 2003). For example, a customer might increase his deductible because he has some private information about his risk type, or because the higher deductible is priced such that it is simply a better buy. By conditioning on the factors that go into the pricing process, all that is left are decisions that are made in response to these information asymmetries. Therefore, when I talk about high-risk or low-risk customers, I am always conditioning of these other factors.

2.3 The Model

2.3.1 Model Description

The model jointly estimates the probability that a customer will switch his deductible and the probability that he will file a claim. These events are observed in each year as a pair of binary outcomes. I define the first element of this outcome pair as "claim," and assign that element a value of 1 if the customer files a claim on the policy in that year, and 0 if he does not file a claim. Similarly, I define the second element as "switch," which takes a value of 1 if the customer increases his deductible and 0 if he does not. Hence, in a given year the observed data for a particular customer is either $(0, 0)$, $(0, 1)$, $(1, 0,)$, or $(1, 1)$. If the outcomes for a specific household, across multiple time periods, are negatively correlated, then that customer is less likely to file a claim after increasing his deductible.

One way to think about an observed binary outcome is to define it in terms of the value a continuous latent variable, defined on the $(-\infty, \infty)$ interval, that is not directly observed. In the simple univariate case, if the binary observed data is 1, then the corresponding latent variable must be less than some threshold, and if the observed value is 0, the latent variable must be greater than that threshold. If the latent variable has a standard normal distribution, then this model is the well-known (univariate) probit model for binary choice. One can then extend this model to one in which there are two correlated binary outcomes that are determined by the values of a latent bivariate normal (BVN) distribution. This is a bivariate probit model, where the probabilities of switching or claiming are controlled by the means of the underlying BVN, and the correlation between the two observed binary outcomes is controlled by the correlation of that BVN. If the correlation of the latent variables is positive, the correlation between the binary outcomes is positive as well. In a typical bivariate probit model, the variance parameters for the BVN cannot be identified (probabilities do not change when the variances are multiplied by a scalar), so I constrain them to be 1 without any loss of generality (Chib and Greenberg, 1998). Thus, while the BVN has five parameters, only three need to be estimated for a bivariate probit: the two means and the correlation coefficient. Chiappori and Salanie (2000) used the bivariate probit to test for asymmetric information in automobile insurance, and my approach follows theirs closely. The multivariate probit model (of which the bivariate probit is the two-dimensional special case) is, in fact, a well-established statistical technique that has been used in many diverse contexts in which the researcher observes correlated binary data, such as voter behavior (Chib and Greenberg, 1998), survey analysis (Edwards and Allenby, 2003) and multiple product purchasing (Manchanda et al., 1999).

I operationalize the bivariate probit model in a mathematically equivalent way that is more easily interpretable than the standard approach. Rather than estimating the means of the underlying BVN and examining the probability that the latent variable is either positive or negative, I fix the means at zero and estimate two thresholds that define how much of the mass of the BVN distribution is above or below that threshold. The two thresholds (one for the switch dimension and one for the claim dimension) define four quadrants in which a random draw from this distribution may fall. If the customer switches, then the draw for the switch must have been less than the switch threshold, and if he doesn't switch, the latent variable must have exceeded the threshold. The same goes for claims. To change the probability of a switch or claim, just change the corresponding threshold. Therefore, as the threshold moves away from its parallel axes in the positive direction, the likelihood of that event goes up as well.

To define this model more formally, let s_{ht} be a binary variable that equals 1 when household h increases its deductible in year t, and 0 otherwise. Similarly, c_{ht} equals 1 when a customer files a claim in year t, and 0 otherwise. Hence, for each household in each year, the binary pair (c_{ht}, s_{ht}) is observed. Furthermore, let w_{ht} be the bivariate draw from a standard bivariate normal distribution with mean vector $(0, 0)$, and covariance matrix $\rho_{ht} I_2$, where I_2 is a 2x2 identity matrix and ρ_{ht} is a correlation coefficient that is household- and time-specific. This latent variable w_{ht} is partitioned into its two components w_{sht} and w_{cht}. Two threshold parameters, μ_{sht} and μ_{cht}, correspond to the switching and claiming behavior for household h at time t. Therefore,

$$s_{ht} = \begin{cases} 1 \text{ if } w_{sht} \le \mu_{sht} \\ 0 \text{ if } w_{sht} > \mu_{sht} \end{cases} \text{ and } c_{ht} = \begin{cases} 1 \text{ if } w_{cht} \le \mu_{cht} \\ 0 \text{ if } w_{cht} > \mu_{cht} \end{cases} \tag{2.1}$$

At time t, when μ_{sht} is large, is is more likely that h increases its deductible, and if μ_{cht} is large, it is more likely that h files a claim.

The values of μ_{sht}, μ_{cht} and ρ_{ht} are determined by functions of observed covariates and unobserved parameters. Let x_{ht} be a vector of covariates for household h in year t, and let β be a vector of coefficients (the same length as x_{ht}) that correspond to each of the covariates in x_{ht}. Both x_{ht} and β can be partitioned into subvectors that represent covariates and coefficients that drive μ_{sht}, μ_{cht} and ρ_{ht} respectively, such that $x_{ht} = (x_{sht},\ x_{cht},\ x_{\rho ht})$ and $\beta = \left(\beta_s,\ \beta_c,\ \beta_\rho\right)$. Hence,

$$\begin{aligned} \mu_{sht} &= \beta_s x_{sht}; \\ \mu_{cht} &= \beta_c x_{cht}; \text{ and} \\ \rho_{ht} &= 2 \cdot \text{logit}^{-1}\left(\beta_\rho x_{\rho ht}\right) - 1 \end{aligned} \tag{2.2}$$

The rescaling for the correlation term ensures that ρ_{ht} is between -1 and 1. Thus, x_{ht} and β, through μ^s_{ht}, μ^c_{ht} and ρ_{ht}, determine the joint probabilities of increasing a deductible and filing a claim.

The estimates for β_s and β_c are the marginal effects of changes in x_{sht} and x_{cht} on μ_{cht} and μ_{sht}, respectively, while β_ρ is the marginal effect of changes in $x_{\rho ht}$ on $0.5\cdot$ logit(ρ_{ht}). While these values are helpful when determining the qualitative significance of a particular covariate, it is not easy to interpret them in a quantitative sense. The true values of interest are the marginal effects on the probabilities and correlations themselves, and not the effects on the thresholds, or on some function of the correlation. A conversion is done by computing the first derivative of the probabilities with respect to each covariate. Recall that the joint probability of switching and claiming is the volume under the surface of a bivariate normal distribution, truncated by two orthogonal thresholds. The marginal probability of switching is the cumulative distribution function of a normal distribution, evaluated at the threshold μ_s (the same story applies for claiming). So, if x_{si} is the i^{th} element of x_s, and β_{si} is its corresponding coefficient, then

$$\Pr(s) = \Phi(\beta_s x_s),$$

and

$$\frac{\partial \Pr(s)}{\partial x_{si}} = \beta_{si}\phi(\beta_s x_s). \tag{2.3}$$

where $\Phi(\cdot)$ is the standard normal distribution function and $\phi(\cdot)$ is the standard normal density function. Similarly, the marginal effect of a covariate on the correlation is computed by differentiating (2.2). However, estimates of β_s, β_c and β_ρ reveal the sign and significance of the marginal effects. Another step is required to quantify these effects.

The relationship among switching and claiming probabilities, thresholds and correlations is illustrated in Figure 2.1. In this figure, I plot the contours of a bivariate normal distribution with mean at $(0,0)$ and a covariance of ρI_2 where $\rho = 0.5$. These contours represent the joint distribution of w_s and w_c. The perpendicular lines represent μ_s and

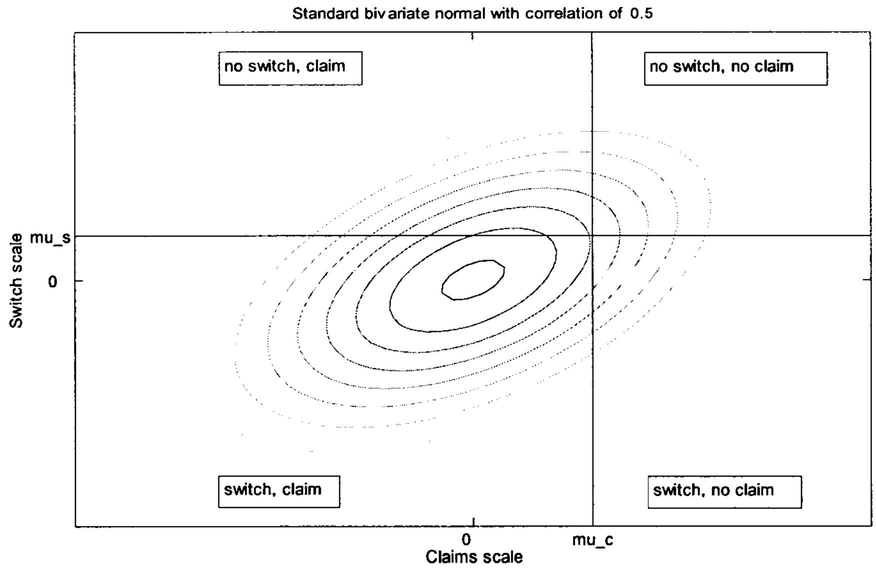


Figure 2.1: Graphical illustration of bivariate probit

μ_c, which are determined by functions of covariates and coefficients, as described above. When μ_c increases, the probability that the "claim" component of a random draw from this distribution is less than μ_c increases as well (the same holds for the "switch" component). This is why lower draws are more likely to indicate the occurrence of the event in question. The μ_s and μ_c lines divide the domain space into four quadrants, such that the probability of each event combination is equal to the volume under the joint distribution and over the quadrant in question. Each household, in each year, has a different probability of switching or claiming because each household, in each year, has different values for μ_s, μ_c and ρ.

2.3.2 Data

I estimate the model using the State Farm data that was described in section 1.3. From this data, I compute binary variables that indicate whether or not a customer increased his deductible from one year to another, and whether or not a claim was filed in a particular year. For each household, there are four observed pairs. For the first observation, "switch" is 1 if the customer increased his deductible at the start of the policy year that begins during calendar year 2000, and "claim" is 1 if the customer filed a claim during the 2000-2001 policy year. The second observation represents the switch decision in 2001 and whether a claim was filed in 2001-2002, and so forth. Naturally, since the data set runs to the end of 2004, the last observation must be a deductible decision in 2003 and the claim incidence in 2003-2004.

The mechanism of deductible choice that generates the switching variables merits some special attention. At the start of each policy year, the customer receives a renewal notice that includes his deductible and premium for his insurance contract. If the customer makes no action, the policy is renewed at the current deductible level. However, the customer also has the option of choosing a different deductible. Until 2003, the available deductibles were \$100, \$250, \$500, \$1000 and \$2000. In 2003, the \$100 and \$250 deductibles were discontinued, and customers with those deductibles were automatically switched to a \$500 deductible policy. Additionally, in 2003 State Farm introduced deductibles that were denominated as 1%, 2% or 3% of the coverage amount. For example, if a customer has a policy with a coverage limit of \$100,000, then a policy with a 1% deductible is equivalent to a policy with a \$1000 deductible (State Farm prices these two policies identically).

Covariates

In addition, the State Farm data includes characteristics about all of the policies that the customer could have chosen in each year, and, of course, the characteristics for the policy that he ultimately did choose. The policy options are characterized by the coverage limit, deductible amount and the premium; all other aspects of the terms of insurance are identical across deductible options and across households. Also, the data set contains certain customer-specific characteristics, such as claims and switching history. From this information, one can derive several different as covariates. For each of these variables, I define a short abbreviation, in parentheses, that I use in the tables and for future reference.

Coverage limit (COVERAGE) The coverage limit is the amount for which the customer's house is insured. It is used as a proxy for wealth, since it is correlated with the value of the home. When other covariates interact with coverage amount, this variable in transformed to take a value of 0 if the coverage is less than the mean and 1 if it is greater than the mean (this transformation makes it easier to interpret the coefficients on the interactions).

"1K/1K" premium (PREMIUM) The homeowner pays a premium each year in exchange for insurance coverage. The insurer sets this premium by considering multiple factors, including the coverage limit, the chosen deductible (policies with high deductibles cost less), and the underlying expected loss that the insurer faces under the policy (essentially the riskiness of the customer). I observe the first two directly, and I observe how the premium varies according to deductible (i.e., I know the formula that converts the premium on a policy with a \$500 deductible to the premium for a \$1000 deductible). Therefore, I use the premium, per \$1000 of coverage, that would have been paid had the customer chosen a deductible of \$1000. This normalized premium varies only with the risk level of the customer, as determined by factors such as claims history (experience rating), geography, home construction and so forth. Thus, this variable is a proxy for all of the information that the insurer observes when setting the price of the policy.

Minimum premium saved (PREMSAVE) Because I know how premiums vary with deductibles, I can compute the amount that a customer could save by increasing his deductible. This variable is smallest amount that a customer could save by increasing his deductible, and is equal to the premium associated with the next highest deductible if the customer is currently at a deductible level that is not the highest one available (PREMSAVE takes a value of zero otherwise). Since there may be more than one deductible that exceeds the current one, I use the difference in premiums for a policy under the current deductible and one with the next highest deductible available. For example, if a customer has a $500 deductible, he could potentially raise it to $1000 or $2000. PREMSAVE is the difference between the premium for a $500 deductible policy and the premium for a $1000 deductible policy.

Premium change (PREMCHANGE) This is the amount that the customer's premium would change, per $1000 of coverage, had the customer retained the same deductible as in the previous year. This change reflects not only changes in the insurer's beliefs about the customer's risk level (either through filed claims or changes in the risk profile of the customer's risk class), but also external factors such as changes in the regulatory environment or competitive pressures.

Deductible lag (DEDLAG) This is the size of the deductible that the customer chose in the previous period.

Switch in previous period (SWITCHLAG) This variables takes a value of 1 if the customer increased his deductible in the previous period, and 0 if he did not.

Claim in previous period (CLAIMLAG) This variables takes a value of 1 if the customer filed a claim in the previous period, and 0 if he did not.

Data Subsampling

In order to make the analysis both interpretable and computationally feasible, I estimate the model using a carefully constructed sample from the complete database. First, I restrict the sample to customers who were active for the entire period of 1999 to 2004, whose coverage limit is $100,000 or more, and whose deductibles exceed $500 in 2002 and thereafter. This yields observations for four policy periods (2001, 2002, 2003 and 2004), which is sufficient to conduct inference on within-household correlations. In 2003, State Farm instituted a minimum deductible of $500, so I use that same minimum deductible to filter out those customers whose deductible increases may not have been “voluntary.” By removing those customers with coverage limits of less than $100,000 (for whom the 1% deductible amount may have been between two deductible options that existed before 2003), those customers whose deductible increase in 2003 may have been solely because of the change in choice set, and not due to either observed cues or asymmetric information, are eliminated. For example, a customer with a coverage limit of $60,000 with a $500 deductible in 2002 might increase his deductible to 1% in 2003 simply because $600 is closer than $500 to his "optimal" deductible.

Second, the number of households in the working data set is reduced to 3,000 by taking a stratified systematic random sample from the entire population of eligible households. For the purposes of this random sample, each household is placed in one of five strata according to its observed binary pairs: did not file any claims and did not increase its deductible in any year; increased its deductible in at least one year, but did not file any claims; filed at least one claim, but did not increase its deductible; increased its deductible in at least one year, and filed at least one claim, but did not do so in the same year; or, in at least one year, the household increased its deductible and filed a claim. This stratification scheme ensures that there are enough households from each strata to allow testing of hypotheses of correlation between the switching of deductibles and the filing of claims. In subsequent analyses, the likelihood contributions for each customer are weighted to reflect their representativeness from the underlying population.

2.3.3 Estimation

Model parameters are estimated through maximum simulated likelihood (MSL) techniques. The likelihood contribution from each household in each year is the volume under the corresponding BVN distribution, and above the quadrant that defines the observed bivariate outcome (see section 2.3.1). If there were no correlation between w^s_{ht} and w^c_{ht}, then it would be straightforward to estimate β using standard maximum likelihood techniques. However, the only way to compute this volume for a BVN with non-zero correlation is through some kind of numerical integration, such as Monte Carlo integration. One way to simulate the probabilities is by using the "GHK simulator" (Geweke, 1991; Keane, 1994), which maintains a smooth likelihood surface (which is necessary when using Newton-based optimization methods) while estimating the volume under a truncated multivariate normal distribution. Although the GHK simulator is often used to estimate multinomial probit models (see Train, 2003, chap. 5), it is easily adapted to this bivariate probit setting.[2]

2.4 Results

In this section, I examine the estimates for the coefficient vectors β_c, β_s and β_ρ for two versions of the model. Both models include the same set of covariates that help determine the probabilities for claiming and switching. The difference between the models is that in Model 1, the within-household correlation is defined as a function of a single parameter, while for Model 2, the correlation is defined as a function on an intercept and covariates. The parameter estimates for Model 1 are summarized in Table 2.1, and the estimates for Model 2 are summarized in Table 2.2. Not only are the covariates for claiming and switching interesting in their own right, but they control for all observable behavior

[2]To reduce Monte Carlo error when simulating the probabilities, we use Halton sequences of pseudorandom variables, rather than truly random draws, to instill negative covariance among the draws (Train, 2003, chap. 9).

that the insurer uses to price the policies (Chiappori and Salanie, 2003). For example, the PREMIUM variable is correlated with all of the information that the insurer observes when setting the premium, as discussed in section 2.3.2. In addition, the intercepts and coefficients for some variables (e.g., PREMCHANGE 2001 and PREMCHANGE 2002).

I begin the analysis by examining the effects of the covariates on the marginal probabilities to switch and claim. Note that the marginal probabilities are not affected by any correlation effects, since the correlation coefficient of a BVN distribution does not influence the marginal distributions. For this analysis, I concentrate on the parameter estimates from Model 1. Then, I investigate the role of asymmetric information on switching and claiming behavior by comparing the estimates of β_ρ in Model 2. Since β_ρ for Model 1 is a scalar, it acts as a measure of overall asymmetric information. For Model 2, β_ρ consists of the coefficients for the set of covariates. Hence, one can examine how these covariates affect the correlation between switching and claiming.

2.4.1 Effect on switching and claims

The results in Table 2.1 correspond to the estimated values of β_s and β_c, which determine the effects of the covariates on switching and claiming. They are interpreted as the marginal effect of covariates on μ_s and μ_c (the thresholds described in section 2.3.1), but since an increase in either threshold μ corresponds to an increase in the corresponding probability, the qualitative significance of the covariates is revealed through their coefficients. Several of these estimates are of note. First, note that PREMIUM (see section 2.3.2 for the definitions of these variables) is positively correlated with the incidence of claims. Recall that PREMIUM is a proxy for the insurer's estimate of the riskiness of policyholder, since insurers generally must collect more from customers, in the form of premium revenue, than the amount that they expect to pay in the form of claims. Hence, it is not surprising that customers with higher premia are more likely to file claims, and it validates the model that insurers use to set premia based on future expectations of risk. This effect is significant even after controlling for both the deductible in the previous period and the incidence of a claim being filed in the previous period.

Next, note that the coefficients for the PREMCHANGE variables are positive, indicating that the probability of switching increases when the customer is faced with a increase in premium. To understand why one might observe such an effect from this pricing cue, consider a customer whose renewal notice indicates an increase in premium from the previous year. This premium increase is based on a renewal of a policy with no change in the terms; the deductible amount remains the same from year to year. However, if the customer considers insurance premiums to be an expense, and that expense is going to increase if the customer retains his current deductible, then it is possible that the customer's underlying mental account for insurance expenses will exceed some psychological budgetary threshold. One way the customer could reduce his immediate insurance-related expenses is to increase his deductible for the next year. Thus, this result is consistent with the mental budgeting hypothesis proposed and tested by Heath and Soll (1996).

The effect of PREMCHANGE goes down between 2001 and 2002, increases substantially in 2003, and continues to increase in 2004. In 2003, State Farm placed a message in renewal notices that referenced the amount of money the customer could save by increasing his deductible. Thus, when a customer is cued by a premium increase, the effect of that cue increases in the year of the renewal notice message. However, one cannot immediately rule out that the effect is simply latent nonstationarity, rather than an effect of the message, since the effect continues to increase from 2003 to 2004, when no message was sent. PREMSAVE is also a significant covariate on the probability of switching, and the effects increase over time. However, one cannot distinguish between latent nonstationarity and the renewal notice effect as plausible explanations for this increase. As PREMSAVE goes up, the probability that the customer will take advantage of that opportunity goes up as well.

I tested a model that included an interaction between PREMSAVE and PREMCHANGE, and found it to be insignificant.

Additionally, these effects on switching behavior are different for customers with expensive homes (as observed in the PREMCHANGE x COVERAGE and PREMSAVE x COVERAGE interactions). Because the coefficients on these interactions are negative, one can infer that customers with more expensive homes are less likely to respond to savings on premia than poorer customers. This result provides additional support for the mental budgeting hypothesis, since customers of lower means may have a greater need to save money on premiums when the situation presents itself.

Of course, one can quantify all of these effects by computing the marginal effects of a unit change of these cues on the marginal probabilities that the switch or claim event occurs (see section 2.3.1). For example, suppose one wanted to quantify the marginal effect of a premium increase from 2002 to 2003. Note that this effect is nonlinear in the covariates, so the effect is estimated at the covariate's mean value. For this example, the mean premium increase is \$0.72 per \$1000 in coverage, and the mean home value is \$182,000, so the implied premium increase is \$131. Applying equation (2.3) with $\beta_{si} = 0.225$ (the PREMCHANGE 2003 parameter in Table 2.1), the marginal effect is 0.00889. This value is interpreted as the increase in the probability of switching if, for a \$182,000 home, the premium were to increase \$313 (based on a \$1.72 increase per \$1000 of coverage, instead of \$0.72). So if there were a 6% probability that this policyholder would switch after a \$131 increase, the probability increases to about 6.8% if the premium increase were \$313. While a 0.8% increase may not seem like a lot, it does represent a 13.3% increase in the number of switchers, which could have a noticeable impact on the profitability of these policies. If the coverage amount were above the mean (say, \$200,000), then the implied premium increase (again, at \$0.72 per \$1000 in coverage) is \$144. In this case, the marginal effect from PREMCHANGE 2003 and PREMCHANGE x COVERAGE

COVARIATES AFFECTING CLAIMS

	parameter	s.e.
INTERCEPT 2001	-0.993	0.013
INTERCEPT 2002	-1.330	0.015
INTERCEPT 2003	-1.717	0.018
INTERCEPT 2004	-1.842	0.019
PREMCHANGE 2001	-0.067	0.053
PREMCHANGE 2002	0.054	0.044
PREMCHANGE 2003	-0.260	0.042
PREMCHANGE 2004	0.122	0.042
PREMSAVE 2001	0.211	0.102
PREMSAVE 2002	-0.076	0.109
PREMSAVE 2003	0.178	0.116
PREMSAVE 2004	-0.077	0.067
COVERAGE	0.001	0.000
PREMIUM	0.115	0.013
PREMIUM X COVERAGE	-0.035	0.024
PREMCHANGE X COVERAGE	0.012	0.048
PREMSAVE X COVERAGE	0.256	0.136
DEDLAG	0.000	0.000
CLAIMLAG	0.065	0.024

COVARIATES AFFECTING DEDUCTIBLE SWITCHING

	parameter	s.e.
INTERCEPT 2001	-2.289	0.020
INTERCEPT 2002	-2.408	0.023
INTERCEPT 2003	-1.230	0.023
INTERCEPT 2004	-1.073	0.022
PREMCHANGE 2001	0.136	0.050
PREMCHANGE 2002	0.033	0.047
PREMCHANGE 2003	0.225	0.032
PREMCHANGE 2004	0.356	0.030
PREMSAVE 2001	0.778	0.122
PREMSAVE 2002	1.353	0.128
PREMSAVE 2003	2.522	0.095
PREMSAVE 2004	3.371	0.063
SWITCHLAG	-0.068	0.023
COVERAGE	0.003	0.000
PREMIUM	0.416	0.013
PREMIUM X COVERAGE	0.265	0.022
PREMCHANGE X COVERAGE	-0.027	0.042
PREMSAVE X COVERAGE	-2.354	0.130
DEDLAG	-0.007	0.000
CLAIMLAG	-0.165	0.028

COVARIATES AFFECTING CORRELATION

	parameter	s.e.
INTERCEPT	-0.104	0.017

Table 2.1: Parameter estimates for Model 1

interaction must be considered. Thus, the marginal effect is $0.00889 - 0.00157 = 0.00732$. If the premium were to increase \$344, the 6% probability of switching would only increase to 6.7%.

2.4.2 Asymmetric information

I now turn to the question of the presence of asymmetric information. Informational asymmetries can be measured by looking at the correlation coefficient of the latent bivariate normal distribution that determines the switching and claiming probabilities. A negative correlation coefficient means that as the conditional switching propensity increases, the conditional claim propensity decreases. In other words, a negative correlation is a signal that more conditionally low-risk customers are increasing their deductibles, relative to the number of conditionally high-risk customers. This is what economic theories of adverse selection and moral hazard predict, and it is what Chiappori and Salanie (2000) observed in their study for automobile insurance using a common correlation across households.

In Model 1, this correlation is determined by a single parameter, which is then converted to a correlation coefficient using equation (2.2). This "intercept" parameter, the last entry in Table 2.1, is estimated to be -0.104. To convert this value to a correlation coefficient, apply equation (2.2) to get $\rho = -0.052$.

The main question, however, is how cues influence this correlation. In Model 2, the correlation is a function of both time-varying intercepts and a set of covariates. If the coefficient on a covariate is negative, then an increase in the level of the corresponding cue increases the effect of asymmetric information. The reason for that is due to one of the two economic stories of asymmetric information. If customers behave according to adverse selection, then customers who believe they are low risk will increase their deductibles, while if the story is one of moral hazard, then customers who increase their deductibles will behave in a way that leads to fewer claims. In either case, if a pricing cue that increases the propensity of a customer to increase his deductible also makes the correlation between switching and claiming more negative, then that cue induces more conditionally low-risk customers to switch, relative to the number of conditionally high-risk customers. Hence, negative coefficients on cues indicate more asymmetric information in the system when the intensity of that cue is increased.

The parameter estimates for Model 2 are summarized in Table 2.2. There are three sections in this table: one each for the estimates of the effects on claiming, switching and the correlation. The estimates for the switching and claiming parameters in Model 2 (the first two sections in Table 2.2) are qualitatively similar to those in Model 1. A notable exception is that after controlling for the effect of the premium cues on the correlation, the effect of PREMSAVE on switching is clearly more significant in 2003 than in adjacent years. This result makes makes it more likely that the increase in the effect of premium savings in 2003 is due to the interaction with the message, and cannot be attributable solely to latent nonstationarity.

Next, consider the effect of covariates on the correlation term. The first observation from the third section in Table 2.2 is that PREMSAVE is large and positive in all four years, suggesting that the more a customer can save by increasing his deductible to the next highest level, the more likely it is that such a customer is conditionally high-risk. However, this effect is mitigated for customers with high coverage limits (as evidence by the PREMSAVE x COVERAGE interaction). In addition, the same kind of effect exists for PREMCHANGE in 2001 and 2002, but the opposite effect in 2003 and 2004. One possible reason for the change in sign is due to the message that was included in 2003 renewal notices (see section 2.1). This interaction indicates that the message triggers more conditionally low-risk customers to switch. The trend in these coefficients is likely to be due to such an interaction, rather than latent nonstationarity, because the coefficient increases again in 2004, the year after the renewal notices were introduced. Even though

COVARIATES AFFECTING CLAIMS

	parameter	s.e.
INTERCEPT 2001	-0.883	0.012
INTERCEPT 2002	-1.245	0.015
INTERCEPT 2003	-1.815	0.017
INTERCEPT 2004	-1.806	0.018
PREMCHANGE 2001	-0.151	0.047
PREMCHANGE 2002	0.046	0.037
PREMCHANGE 2003	-0.359	0.033
PREMCHANGE 2004	0.327	0.025
PREMSAVE 2001	1.945	0.046
PREMSAVE 2002	1.566	0.053
PREMSAVE 2003	1.173	0.091
PREMSAVE 2004	-0.097	0.060
COVERAGE	0.001	0.000
PREMIUM	0.052	0.009
PREMIUM X COVERAGE	-0.022	0.016
PREMCHANGE X COVERAGE	0.101	0.042
PREMSAVE X COVERAGE	0.493	0.080
DEDLAG	0.000	0.000
CLAIMLAG	0.107	0.021

COVARIATES AFFECTING DEDUCTIBLE SWITCHING

	parameter	s.e.
INTERCEPT 2001	-2.042	0.023
INTERCEPT 2002	-2.215	0.027
INTERCEPT 2003	-1.091	0.026
INTERCEPT 2004	-1.279	0.022
PREMCHANGE 2001	0.320	0.056
PREMCHANGE 2002	0.241	0.038
PREMCHANGE 2003	0.100	0.039
PREMCHANGE 2004	0.321	0.022
PREMSAVE 2001	0.647	0.058
PREMSAVE 2002	0.094	0.068
PREMSAVE 2003	1.289	0.123
PREMSAVE 2004	-0.022	0.071
SWITCHLAG	-0.073	0.019
COVERAGE	0.002	0.000
PREMIUM	0.011	0.010
PREMIUM X COVERAGE	-0.008	0.019
PREMCHANGE X COVERAGE	-0.224	0.047
PREMSAVE X COVERAGE	-0.264	0.083
DEDLAG	-0.005	0.000
CLAIMLAG	-0.154	0.030

COVARIATES AFFECTING CORRELATION

	parameter	s.e.
INTERCEPT 2001	0.017	0.033
INTERCEPT 2002	-0.190	0.037
INTERCEPT 2003	-0.442	0.038
INTERCEPT 2004	-1.870	0.030
PREMCHANGE 2001	0.537	0.085
PREMCHANGE 2002	0.393	0.069
PREMCHANGE 2003	-0.184	0.050
PREMCHANGE 2004	-0.083	0.027
PREMSAVE 2001	17.824	0.468
PREMSAVE 2002	18.176	0.449
PREMSAVE 2003	17.997	0.381
PREMSAVE 2004	13.552	0.287
COVERAGE	0.001	0.000
PREMIUM	-2.363	0.047
PREMIUM X COVERAGE	0.304	0.040
PREMCHANGE X COVERAGE	-0.344	0.064
PREMSAVE X COVERAGE	-2.163	0.278
DEDLAG	0.002	0.000
CLAIMLAG	-0.037	0.041

Table 2.2: Parameter estimates for Model 2

there was no message in 2004, there could still be some lingering effect from the message in the previous year.

Altogether, these results may indicate a budgetary motivation, either fiscal or mental, for customer behavior. From the previous section, customers with smaller homes are more sensitive to pricing cues when deciding whether or not to increase deductibles. Now, these same cues trigger high-risk customers to switch, and that this leverage is even more forceful for these poorer customers. It should not be surprising that individuals that face tight budget constraints are more sensitive to premium increases or premium savings than other customers if these customers are the poorer members of the customer population. Thus, conditional on all of the information the insurer uses to price the policies, to the extent that these pricing cues induce customers to increase their deductibles, it is the conditionally high-risk customers who do the switching. Interestingly, while the main effect of the amount of coverage is statistically significant, the magnitude of the coefficient is small (a $1000 increase in coverage increases the correlation coefficient by only 0.00011). Hence, the wealth effect acts only through the cues.

Another interesting observation is the effect of premium levels themselves (as opposed to the change in premium) on asymmetric information. The premium level is set by the insurer (exogenously for present purposes, since it is unlikely that a single customer would influence overall premium levels) based on the expected rate of claims for the coming policy year. Thus, the premium is a proxy for the insurer's perception of the customer's risk. Since the effect of premium on the correlation is negative, high-risk customers exploit asymmetric information more than low risk customers. Again, this effect is stronger for customers with less expensive houses. An inference from this result is that even after all observable risk has been accounted for, the poor, high-risk customers will still choose (or have chosen) not to increase their deductibles.

	Calibration		Holdout	
	Model 1	Model 2	Model 1	Model 2
log likelihood	37593	36447	39655	39277
BIC	37788	36729	39850	39559

Table 2.3: Statistical comparison of models using BIC

2.4.3 Model assessment

Although many covariates in these models are significant, it is not immediately clear that the model represents the observed data well. One can assess model performance based on either relative or absolute criteria, and then compare the model fit on both the calibration data set, as well as a holdout data set. The log likelihoods and BIC (Bayesian information criterion, a penalized measure of model performance) for both models are summarized in Table 2.3. For both models, the BIC for Model 2 is better than for Model 1, indicating that even after taking the additional number of parameters in Model 2 into account, Model 2 is still a superior fit. One can also use a likelihood ratio test (LRT) to compare the models. The LRT test statistics are 2292 for the calibration data set and 756 for the holdout data set, which are distributed according to a χ^2 distribution with 18 degrees of freedom (the difference in the number of parameters) and are significant at any reasonable level of significance.

So, even though Model 2 is statistically "better" than Model 1, this relative model assessment says nothing about how well Model 2 explains the observed customer behavior in an absolute sense. One way to assess model fit is to compare the expected aggregated incidence of the behavior of interest with the observed incidence of that behavior. Table 2.4 summarizes the actual and expected percentages of customers who file claims in each year, switch their deductibles in each year, and both claim and switch in the same year. The model captures this behavior reasonably well for both the calibration and holdout samples (the holdout sample was constructed in a manner identical to the calibration sample, as

	Calibration		Holdout	
Pct switching	Actual	Predicted	Actual	Predicted
2001	0.07	0.06	0.07	0.06
2002	0.06	0.05	0.07	0.05
2003	0.23	0.23	0.23	0.24
2004	0.21	0.20	0.22	0.20
Pct claiming	Actual	Predicted	Actual	Predicted
2001	0.16	0.19	0.16	0.19
2002	0.09	0.11	0.09	0.11
2003	0.04	0.04	0.05	0.04
2004	0.03	0.04	0.04	0.04
Pct both	Actual	Predicted	Actual	Predicted
2001	0.01	0.02	0.01	0.02
2002	0.01	0.01	0.01	0.01
2003	0.01	0.01	0.01	0.01
2004	0.01	0.01	0.01	0.01

Table 2.4: Comparison of the predicted and actual aggregated probabilities of switching and claiming.

described in section 2.3.2).

However, comparing aggregated predictions does not reveal the performance of the model at the individual level. One way to do this is to examine how well the model estimates the probabilities of the occurrence of individual events. In this case, there are four possible events for each household-year: switch and claim, switch and no claim, no switch and claim, and no switch and no claim. Following a modification of a method proposed by Barnett et al. (1981), each household-year is assigned to one of these four events, and matched to the outcome to the probability of that event occurring for that household-year. These household-years are then grouped into three bins: those that are estimated to have a high probability for their assigned event, those with a medium estimated probability, and those with a low estimated probability. If the model performs well at the individual event level, then for both the calibration and holdout samples, events that are predicted to have a high probability of occurring should happen often, and events that are predicted to have a low probability should happen rarely. If the predicted probabilities do not match the observed frequencies within each group, even though they do match across the entire sample, then the aggregate measure of model fit is simply an artifact of the overestimation from one group offsetting the underestimation in another.

The predicted probabilities and observed frequencies for each of the three bins are summarized in Table 2.5. Within the group of events that the model says should occur with high probability (i.e., predicted to occur with a probability above 0.67), the model predicts a frequency of about 82 percent. In fact, these same, individual events occur about 79 percent of the time in both the calibration and holdout data sets, indicating that the model does a reasonably good, but not perfect, job of predicting high probability events. At the other end of the scale, across all of those events that the model predicts should happen rarely, the estimated frequency is about 6.5 percent, while the observed frequency is about a percentage point higher. The model does worse among those events that it predicts should occur with moderate frequency (between 33% and 67%), but the difference between the estimated and actual frequencies is still about 7 percent for both the calibration and holdout samples. This discrepancy could be because only about 5.5 percent of all events are predicted to have moderate probabilities to begin with, so the estimation of the model parameters attempts to maximize the likelihood of the occurrence of the events at the extremes. What is important here is that the model is not offsetting overprediction of less probable events with underprediction of more probable ones, or vice-versa. Therefore, one can say that the model does reasonably well at the individual event level as well as the aggregate level.

	Calibration		Holdout	
Bin	Predicted	Observed	Predicted	Observed
.67-1.00	0.82	0.79	0.82	0.79
.33-.67	0.55	0.48	0.54	0.48
0-.33	0.07	0.08	0.07	0.08

Table 2.5: Model fit at the individual level

2.4.4 Unobserved heterogeneity

The individual-level heterogeneity that is incorporated in the model is limited to that which is observed through covariates. It is possible that there are differences among households that are not observed. I tried several different approaches to add unobserved heterogeneity into the model. First, I modeled the switch and claim incidences in each year as being generated by two independent Bernoulli trials. If one wanted to incorporate heterogeneity into this model, one might allow each of the two probabilities for switching and claiming to vary across the population according to their own beta distribution. Each beta distribution had two parameters, but since only a single outcome for each Bernoulli trial in this year is observed, these parameters cannot be identified in this model.

The next attempt was to model the switching and claiming behavior for all four years a beta-binomial model. This model can be identified. However, the estimated parameters implied a degenerate beta distribution that would ordinarily reflect a lack of any unobserved heterogeneity. However, the reason these beta distributions were spiked is that there were no households that either switched three or four times, or filed three or four claims. Hence, all of the mass of the beta distributions were concentrated on the left tail in order to keep the right tails extremely light. Furthermore, the beta-binomial distribution assumes that the probabilities of switching or claiming are stationary, which appears to be inappropriate for this data set.

I then decided to incorporate unobserved heterogeneity by adding multiple latent classes to the model. At first, I estimated latent class models without any covariates: the switching and claiming thresholds, plus the correlation coefficients, were expressed as single parameters. Estimates of the parameters suggested another degenerate solution, where all customers are believed to be members of the same class, suggesting, again, the absence of unobserved heterogeneity.

Finally, I estimated the model allowing for two and three latent classes of parameters for the full set of covariates. For L classes and k parameters per class, the total number of parameters is $Lk + L - 1$. Since $k = 58$, I am not including the coefficient estimates for the latent class models, but none of the key parameters on covariates of interest vary sufficiently that the underlying story changes. Furthermore, while model fit (as measured by BIC) improves as additional classes are added, any improvement on absolute measures were hardly distinguishable. Therefore, I can continue to tell a credible central story of asymmetric information by examining the estimates of coefficients from a model that assumes customer homogeneity.

2.5 Policy implications and future research

In this paper I report four patterns in the behavior of customers of homeowner's insurance:

1. the probability that a customer raises his deductible will go up with the amount of the increase in premiums that the customer faces, and with the amount of money that the customer could save by making the deductible switch;

2. the overall correlation between incidence of switching and incidence of claiming is negative, suggesting that, on average, more conditionally low-risk customers are the ones who are increasing their deductibles;

3. marketing cues, such as increases in premiums and messages about the potential for saving on premiums, make the correlation between switching and claiming more

positive, indicating that it is the high-risk customers that switch in response to those cues; and

4. customers with high premiums (whom the insurer believes are higher risk) exploit asymmetric information by increasing deductibles in years in which they do not file claims (a behavior predicted for conditionally low-risk customers).

These findings should be of interest to insurers who may be thinking of inducing customers to increase their deductibles. At first glance, one might think that higher deductibles would be desirable for the insurer. Not only would the expected indemnity for a policy with a high deductible be lower (since losses with severities below this higher threshold would not be eligible for reimbursement), but the company would save on the servicing and processing costs associated with handling these claims. But insurance policies are profitable only if they are priced correctly for the customers what ultimately purchase them. For example, in the Rothschild and Stiglitz (1976) model, high-deductible policies are priced at the break-even level under the assumption that only low-risk customers choose them; if high-risk customers take them, the revenue from these policies would not be able to offset the increase in paid indemnities. In a real-world market, one does not know which policies, if any, are profitable for any particular mixture of customer types. For example, regulatory restrictions could force insurers to price policies for specific groups of customers at rates that deviate from the competitive equilibrium rates. But it is entirely possible that there are some customers for whom the insurer would prefer not to choose policies with certain deductibles, either because the indemnities will exceed the premiums or because overall premium revenue would decline.

So when an insurer decides to increase the premiums of a customer, adjust the premium gap between otherwise identical policies that have different deductibles, or even send messages to customers asking them to switch, the results may be unpredictable. It should not be surprising that, on average, the marginal probability of increasing one's deductible goes up when the customer is faced with a premium increase, or could save a lot of money on premiums by taking the higher deductible, and that the effect is weaker for customers with more expensive homes (and ostensibly more overall resources). If the customer has categorized expenses related to insurance premia into a separate mental account, increasing his deductible provides an opportunity to maintain those expenses below a budgetary threshold.

While this is a nice behavioral result, the real motivation is whether or not the policy remains profitable (or, at least, as profitable) for the insurer if all customers were to switch to this policy. Even if the competitive equilibrium assumptions of RS are violated (e.g., through regulation, search costs, multiperiod learning or multiple customer types), the insurer receives less revenue for policies with higher deductibles, so a particular policy type can only be profitable if enough low-risk customers are in the mixture of customers who choose that policy.[3] Overall, it is the conditionally low-risk customers making the switch, as result that is consistent with some previous research on the subject for personal property/casualty insurance lines (Chiappori and Salanie, 2003). But customers whose switching is triggered by pricing cues are more likely to be conditionally high-risk customers. Thus, insurers must be careful about the tools they use if they want certain customers to increase their deductibles. Even internal measures of risk assessment, such as those that determine the baseline premium, may not be sufficient to target the "right" kind of customers to switch, since these customer are the mostly likely to increase their deductibles and then, ironically, not file claims at all.

[3]If too many high-risk customers, relative to the number of low-risk customers, also increase their deductibles, the subsequent volume of claims on that policy will make the policy less profitable. The term "enough" in the sentence in the main text is purposely vague here. If a policy is priced such that the premiums are extremely high, it might take a large number of customers who are extremely high-risk to make the policy unprofitable. In any event, such a policy becomes *less* profitable.

Here is where some future research on separating the effects of adverse selection and moral hazard may be helpful. For example, consider a customer who is conditionally high-risk, but doesn't know it. Suppose the insurer tells this customer that his premiums will increase in the next year. If he is budget constrained (mentally or otherwise), he may increase his deductible customer to relieve that constraint. Now suppose that he is less likely to make investments in his home to improve safety than other customers who also increased their deductibles, and thus he files a claim on his policy. The source of the asymmetric information for this customer is moral hazard, but not adverse selection. If the insurer knew whether a customer responded because of moral hazard or adverse selection, it might be able to influence behavior specifically. For instance, the insurer might decide to provide less of a reduction in premium for the increase in deductible, and then add an additional credit for subsequent investments in safety. Hence, the customer alleviates the budget constraint as before, but also has an incentive not to act as a moral hazard. Unfortunately, the model cannot differentiate between moral hazard and adverse selection, and answering such questions are notoriously difficult. Recently, several researchers have attempted to determine the effect of one or the other in empirical data (Abbring et al., 2003a,b; Dionne et al., 2004). An appropriate follow-up would be to evaluate the effect of cues on the effects of either moral hazard or adverse selection.

Chapter 3

Modeling the "Pseudodeductible" in Insurance Claims Decisions

3.1 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 chapter, I 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. This 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 chapter 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 less than $1,000) and to understand why the average size of claims increases (within household) from claim to claim, even after adjusting for inflation. I present an explanation for both of these phenomena that not only explains 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, I 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 that 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 chapter 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. While 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, I 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.

The 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, since neither the unclaimed losses nor the pseudodeductible threshold are observed directly. By exploiting a rich proprietary source of household-level data, I can (and do) infer a relationship between a customer's rate of losses (both claimed and unclaimed) and the pseudodeductible. I 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 or not 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 and which ones are not. By decomposing the generating process of the observed data into these three subprocesses, one can identify relationships between the rate at which all losses occur and the size of the difference between the policy deductible and the pseudodeductible.

The 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. My 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 this sample, 18 percent of households filed at least one claim of less than $1000 during the six-year observation period); and

2. the percentage of households whose claim severity increases from claim to claim (53.1 percent in this sample).

In other words, in developing a model that allows for the possibility that customers leave money on the table in the short term, one can make better predictions about many different features of the observed customer activity. Furthermore, the model allows one 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 my 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 have been several streams of research that touch on many of the concepts that I use in this chapter. 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. Yet, for all of the interest in this topic, there has been no attempt (as far as I know) to estimate the claims "rule" that individuals actually apply. Newhouse et al. (1980) comes closest, recognizing that the effective deductible (what I 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, since 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 (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 3.2 formally presents the pseudodeductible model in which the pseudodeductible is assumed to be stationary. Sections 3.3 and 3.4 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 section 3.5, as a demonstration of the power of the model, I segment the customer base according to the posterior probabilities of having particular combinations of loss rates and pseudodeductibles. Section 3.6 refines the model by using nonstationary pseudodeductibles to explain the phenomenon of increasing claim severities. Finally, in section 3.7, I discuss the results, extend insights outside of the insurance field and propose topics for future related research.

3.2 The Model

I 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, I begin by constructing the likelihood of the observed data by integrating together the distributions of all of the data. I then introduce a semiparametric prior distribution on the parameters of the likelihood.

3.2.1 Notation

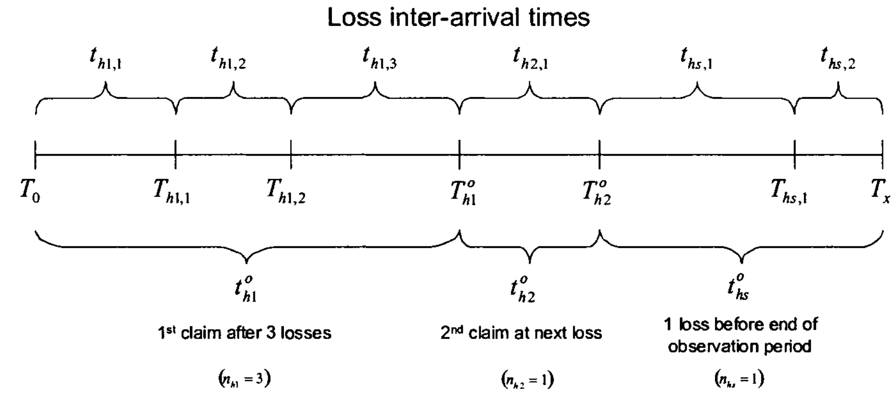
Let $f^o(t^o, y^o|\theta)$ be the probability density function for the timing and severity of observed claims for all H households, such that

$$f^o(t^o, t^o_s, y^o|\theta) = \prod_{h=1}^{H} f^o_h(t^o_h, t^o_{hs}, y^o_h|\theta_h), \tag{3.1}$$

where $f^o_h(\cdot)$ is the observed data likelihood for household h, $t^o_h = \left(t^o_{h1} \dots t^o_{hK_h}\right)$ is the vector of claim inter-arrival times for the K_h claims that are filed by household h during the observation period, t^o_{hs} is the "survival time" between the K^{th}_h filed claim and the end of the observation period, and $y^o_h = \left(y^o_{h1} \dots y^o_{hK_h}\right)$ 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

[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

timeline



2.pdf

Figure 3.1: Sample loss and claim arrival process

(the contents of θ_h will be defined in more detail as I develop the model). By definition, $T^o_h = \left(T^o_{h1} \ldots T^o_{hK_k}\right)$ is the vector of claim arrival times. Additionally, T^o_{h0} is the beginning of the observation period for household h, T^o_x is the end of the observation period, $t^o_{hk} = T^o_{hk} - T^o_{hk-1}$ for all $k = 1 \ldots K_h$ and $t^o_{hs} = T^o_x - T^o_{hK_h}$.

In contrast to T^o_h, 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^u_h is the vector of arrival times of *unobserved* (unclaimed) losses, then $T_h = (T^o_h, T^u_h)$. The loss inter-arrival 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 i^{th} loss, T_{hi}, as $T_{hk,j}$, the time of the j^{th} loss that occurs between the arrival times of the $k-1^{th}$ and k^{th} claims. Using this notation, the inter-arrival 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^o_h vector is empty, and $t^o_{hs} = T_x - T_{h0}$. A graphical representation of this arrival process for a hypothetical household with two observed claims is presented in Figure 3.1.

In addition, define D_{hk} as the amount of the policy deductible in force at the time

may be unobserved.

of claim k. Thus, the the amount of the indemnity received by the policyholder is $\breve{y}_{hk} = y^o_{hk} - D_{hk}$, and y^o_{hk}, 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^o_h is an empty vector [2]. The vector y_h is the set of severities for all I_h losses, both observed and unobserved. Since the timing of unobserved losses (if there are any) is unknown, I have to make some assumptions about the deductible levels that are in force at the times of these losses. 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 k^{th} and $k-1^{th}$ claims, and $\bar{D}_{hs}$ as the average deductible in force during the period between the k^{th} 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 $\Psi\left(D_{hk}, \psi_h\right)$, where $\Psi\left(\cdot\right)$ is a function and ψ_h is a latent, household-specific parameter that determines how much larger the pseudodeductible is over the policy deductible. $\Psi\left(\bar{D}_{hk}, \psi_h\right)$ is the pseudodeductible in force during the period between the arrival times of claim $k-1$ and claim k. The function $\Psi\left(\cdot\right)$ potentially could take an infinite number of functional forms, but in this chapter I consider three:

1. Identity model: $\Psi\left(D_{hk}, \psi_h\right) = D_{hk}$;

2. Additive model: $\Psi\left(D_{hk}, \psi_h\right) = D_{hk} + \psi_h$; or

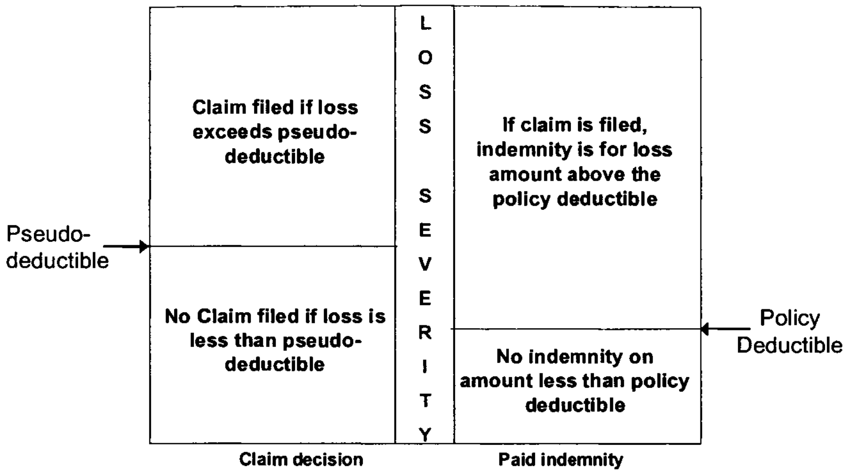
3. Multiplicative model: $\Psi\left(D_{hk}, \psi_h\right) = D_{hk}\left(1 + \psi_h\right)$

Stated explicitly, the pseudodeductible, $\Psi\left(D_{hk}, \psi_h\right)$, is the threshold that determines whether or not a loss is large enough to be claimed, while the pseudodeductible *factor*,

[2] I 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.

composition

box



3.pdf

Figure 3.2: Relationship between policy deductible and pseudodeductible

$\psi_h \geq 0$, controls the relationship between the policy deductible and the pseudodeductible. Under the Identity model (the 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 3.2 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, I 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. Since 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, while 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 \$1000 deductible instead, and faced a \$1200 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 \$1000. 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 \$1000.

3.2.2 Deriving the Likelihood

The likelihood for the observed claims data, conditional on the parameters, can be factored into distributions for the severity vector, the inter-claim arrival times and the survival time. Thus, from (3.1),

$$f_h^o\left(t_h^o, t_{hs}^o, y_h^o|\theta_h\right) = f_y^o\left(y_h^o|t_h^o, t_{hs}^o, \theta_h\right) f_s^o\left(t_{hs}^o|t_h^o, \theta_h\right) f_t^o\left(t_h^o|\theta_h\right) \tag{3.2}$$

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

The Choice and Severity Processes

Let $F_y(y_{hi}|\theta_h)$ be the cumulative distribution function for a single loss severity, y_{hi}, on the domain $(0, \infty)$, and let $f_y(y_{hi}|\theta_h)$ be its density. By definition, the pseudodeductible is the latent threshold that determines whether or not 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 $y_{hk}^o < \Psi(D_{hk}, \psi_h)$. Since I 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

$$f_y^o\left(y_{hk}^o|\theta_h\right) = \frac{f_y\left(y_{hk}^o|\theta_h\right)}{1 - F_y\left(\Psi\left(D_{hk}, \psi_h\right)|\theta_h\right)} \cdot \mathbf{1}\left\{y_{hk}^o \geq \Psi\left(D_{hk}, \psi_h\right)\right\} \tag{3.3}$$

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 one chooses for $F_y(y_{hi}|\theta_h)$. I 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, I assume that μ_{hi} is distributed across all losses according to a gamma distribution with shape parameter r and scale parameter a. Hence,

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

and

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

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

$$\begin{aligned} F\left(y_{hi}|r, c, a\right) &= \int_0^\infty F_y\left(y_{hi}|\mu_{hi}\right) g_\mu\left(\mu_{hi}\right) d\mu_{hi} \\ &= 1 - \left(\frac{a}{a + y_{hi}^c}\right)^r \end{aligned} \tag{3.4}$$

and the density function is

$$f\left(y_{hi}|r, c, a\right) = \frac{rcy_{hi}^{c-1}}{a + y_{hi}^c} \tag{3.5}$$

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). I 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, I 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 an extremely 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 (3.4) and (3.5) into (3.3) and multiplying across claims such that

$$f_{hy}^o(y_h|r,c,a) = \begin{cases} \prod_{k=1}^{K_h} f_y^o(y_{hk}^o|r,c,a) & \text{if } K_h > 0 \\ 1 & \text{if } K_h = 0. \end{cases} \tag{3.6}$$

Timing and Survival Likelihoods

The distributions for $f_t^o(\cdot)$ and $f_s^o(\cdot)$ are based on $f_t(t_{hi}|\theta_h)$, the probability density function of a single loss inter-arrival 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 as 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 inter-arrival times between t_{hk}^o and t_{hk-1}^o. Hence, the distribution of $f_t^o(t_{hk}^o|\theta_h)$, the claim inter-arrival time for a single claim, is equivalent to the n_{hk}-fold convolution of $f_t(t|\theta_h)$.

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

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

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 inter-arrival times would be

$$f_t^o\left(t_{hk}^o | n_{hk}, \lambda_h\right) = \frac{\lambda_h^{n_{hk}} t_{hk}^{o\, n_{hk}-1} \exp\left[-\lambda_h t_{hk}^o\right]}{(n_{hk}-1)!}. \tag{3.7}$$

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

$$p_{hk} = 1 - F_y\left(\Psi\left(\bar{D}_{hk}, \psi_h\right) | \theta_h\right) \tag{3.8}$$

(I use the average policy deductible $\bar{D}_{hk}$ since I are interested in p_{hk} for the time interval between two observed claims) . Because (3.7) is the distribution of the claim inter-arrival times conditional on n_{hi}, I uncondition (3.7) by summing over the distribution of n_{hk}. This operation yields the distribution of claim inter-arrival times of

$$\begin{aligned} f_t^o\left(t_{hk}^o | \theta_h\right) &= \sum_{m=1}^{\infty} \frac{\lambda_h^m t_{hk}^{o\, m-1} \exp\left[-\lambda_h t_{hk}^o\right]}{\Gamma(m)} \cdot (1-p_{hk})^m p_{hk} \\ &= \lambda_h p_{hk} \exp\left[-\lambda_h p_{hk} t_{hk}^o\right] \end{aligned} \tag{3.9}$$

(Kulkarni, 1995).

Notice that (3.9) 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 yields

$$f_s^o\left(t_{hs}^o | \theta_h\right) = 1 - \exp\left(-\lambda_h p_{hs} t_{hs}\right) \tag{3.10}$$

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

As in (3.6), the likelihood vector for the inter-claim times is

$$f_t^o\left(t_h^o|\theta_h\right) = \prod_{i=1}^{K_h} f_{hk}^o\left(t_{hk}^o|\theta_h\right). \tag{3.11}$$

By substituting (3.6), (3.10) and (3.11) into (3.2), and in turn substituting (3.2) into (3.1), one gets the likelihood of the complete observed data set in (3.1).

3.2.3 Heterogeneity Across Households

With the data likelihood established, I now turn to the question of modeling heterogeneity across households by specifying a class of mixing distributions for household-specific parameters. From (3.9), (3.10) and (3.11), notice that the entire parameter space for the model is $(\lambda_1 \ldots \lambda_H, \psi_1 \ldots \psi_H, r, c, a)$. In section 3.2.2, I 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 remains on 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, I 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 a 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 this case, I use the common kernel choice of the bivariate normal distribution. The advantage of using a MDP is that any distribution can be approximated by incorporating a sufficient number of these kernels without resorting to a specific parametric form (Kim, Menzefricke, and Feinberg, 2004). Furthermore, a 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, Menzefricke, and Feinberg (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, I call that component "empty," and thus L is an upper bound on the number of non-empty components in the mixture (which may be as high as the number of households). I 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, this prior on $(\log \lambda_h, \log \psi_h)$ is denoted as $MDP\,(G_0, \alpha)$, where G_0 (the "kernel" distribution) is the bivariate normal hyperprior on $(\log \lambda_h, \log \psi_h)$. By constructing the prior on (λ_h, ψ_h) in this way, the estimation process (discussed below) simulates from the marginal posterior distributions of λ_l, ψ_l and π_l for all l. Since 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. 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] α has no clear interpretation other than a determinant of the level of smoothing in the model (Walker et al., 1999).

3.3 Estimation

The goal is to provide inference regarding the posterior distribution

$$g(\theta|t^o, y^o) \propto f^o(t^o, y^o|\theta)\, g(\theta) \tag{3.12}$$

where $f^o(t^o, y^o|\theta)$ is distributed according to (3.1) and $g(\theta)$ is the full hierarchical prior for $\theta = (\lambda_1 \ldots \lambda_L, \psi_1 \ldots \psi_L, r, c, a)$:

$$\begin{aligned} (\log \lambda_l, \log \psi_l) &\sim MDP(\alpha, G_0) \\ G_0 &\sim MVN(\beta, \tau) \\ \beta &\sim MVN(\beta_0, \Omega) \end{aligned}$$

I 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 section 3.3.2. Next, I describe an application of this pseudodeductible model to a data set of insurance claims.

3.3.1 The Data

I used the State Farm data set, described in section 1.3, to estimate this model. To make the analysis computationally feasible, I selected a subset of households from the comprehensive data set to create a smaller working data set. I generated two systematic random samples, each containing 3,000 homes, for calibration and holdout data sets respectively. A graphical summary of this data is in Figure 3.3. Note that nearly 60 percent of the households have filed no claims at all during this period. Even though the data is sparse, a Bayesian inferential approach that shares information across subjects can draw within-subject inferences about household-level propensities, even when households do not file

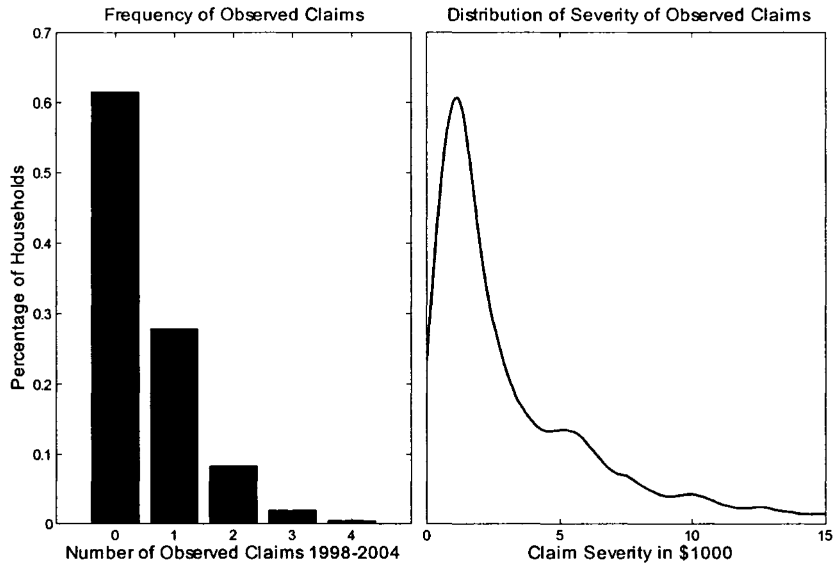


Figure 3.3: Distributions of claim frequencies and severities for observed data.

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) 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, 2004. 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).

3.3.2 Estimation Method

The inferential approach is to simulate draws from the marginal posterior distribution defined by (3.12) using Markov chain Monte Carlo (MCMC) methods. In particular, I used the freely available WinBUGS Bayesian modeling software package (Spiegelhalter et al., 1996). The WinBUGS code for the implementing the MDP prior was adapted from code described in Congdon (2001). Potential label-switching was addressed by post-processing MCMC draws according to the algorithm proposed by Stephens (2000). After several trial runs with L (the maximum number of components) set very high, I observed that many of the posterior estimates for π_l were indistinguishable from zero. Those components are interpreted as being "empty," and eventually limited the number of classes under consideration to $L = 10$. This choice of L dramatically reduces computation time, and I confirmed that a higher maximum value for L would have no impact on the results. The three tested models correspond to the three alternative definitions of the pseudodeductible listed in section 3.2.1: Identity, Additive and Multiplicative. I ran two independent Markov chains for each model and, after a burn-in period, I selected the final 6,000 draws from each chain, for a total MCMC sample of 12,000 draws per pseudodeductible model.[4]

3.3.3 Alternative Approaches

I considered several other inferential approaches to this problem. Instead of using a semiparametric Bayesian method, one might have tried to find an extremum estimator for the parameters of a mixing distribution on λ and ψ. There are several problems with this approach. First, one cannot derive a marginal distribution with which one could estimate the parameters of a mixing distribution that exists in closed form. Hence, I would have use some kind of simulation-based estimation methodology, such as the method of maximum simulated likelihood (MSL). Simulation does not rule out likelihood approaches, as long as one can apply Monte Carlo integration to estimate the marginal distributions, but the number of simulations required to ensure consistency of the estimator is large (Train, 2003). Furthermore, the discontinuity of the likelihood function, which is introduced by

[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.

(3.3), means that traditional "hill-climbing" algorithms like Newton-Raphson would be ineffective. Other extremum estimators, such as indirect inference and moment matching (Gourieroux et al., 1993; Gallant and Tauchen, 1996), can be estimated with fewer simulation draws, but still suffer from the problems caused by the discontinuous objective function. Optimization algorithms that do not require continuous or differentiable objective functions exist, such as pattern search and genetic algorithm methods, but the number of functional evaluations that these methods require to converge to a solution makes them infeasible for this particular problem.

So that leaves Bayesian inference using Markov chain Monte Carlo methods. But why use the MDP as the mixture distribution instead of a parametric form? To see this, think about how one might estimate ψ if it were homogeneous across the customer base. Because (3.3) requires that no observation can have a loss severity that is less than the pseudodeductible, so ψ must be low enough to ensure that this is the case. Hence, a degenerate solution occurs where ψ is set such that it makes the common pseudodeductible just less than the loss implied by the smallest observed claim. If ψ were any higher, then one would never have observed some of the claims that are observed, which is an inherent contradiction. Now, if I placed a parametric mixing distribution on ψ, there would still be a non-zero probability that some homes would violate the restriction from (3.3). But if one assigns households to different classes, as I do through the MDP, I can allow for zero probability that certain households are in classes with high ψ.

3.4 Results

In this section I 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, I will examine the parameter estimates for that model's loss frequency and pseudodeductible estimates. In addition, I will show how the Multiplicative model explains the observed data well in an absolute sense.

3.4.1 Log Likelihood Comparison

As a comparative global measure of model fit, I 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 3.1. 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, since the logs of the Bayes factors comparing it to the Additive and Identity models are over 3000 and 8000, 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]

	Log Marginal Likelihood	
Model	In-Sample	Holdout
Identity	-9590	-9616
Additive	-8639	-8681
Multiplicative	-8051	-8116

Table 3.1: Log Marginal Likelihoods

[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 I am 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, chap. 6).

3.4.2 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. I focus on the median of the loss severity distribution, and present the quantiles for the posterior distributions of the median loss severities in Table 3.2. These quantiles are computed from the MCMC samples described in section 3.3.[6]

	estimated median loss (1000US$)		
Model	10%	50%	90%
Identity	0.809	0.980	1.148
Additive	0.151	0.292	0.450
Multiplicative	0.130	0.249	0.400

Table 3.2: Estimated Median Loss Severities

The median of the median loss distribution for the Identity model is $980, while the estimated medians for the Additive and Multiplicative models are significantly lower: $292 and $249, 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 non-pseudodeductible Identity model. Thus, the Identity model overestimates the median loss severity. I 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.

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

3.4.3 Timing and Pseudodeductible Estimates

The Multiplicative model is the best-fitting model, as indicated in Table 3.1, and therefore all subsequent reported results are based on it. 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).

The posterior distribution for λ_l, ψ_l and π_l is characterized by plotting the medians of the component distributions in Figure 3.4, and by summarizing the quantiles in Table 3.3. 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 $\frac{1}{\lambda_l}$ is the expected time between losses. Furthermore, ψ_l is the percentage *above the policy deductible* that characterizes the pseudo-deductible 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 0.37 to 0.65, which during the seven-year observation period translates to 2.6 to 4.6 losses during that period. The range for ψ is much greater; the median pseudodeductibles range from 103% to 834% of the policy deductible, with a weighted average (including type 1) of about 270%. The relationship between λ and ψ, as illustrated in Figure 3.4, 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.

If one believes 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

Type	Quantile	lambda	psi	proportion
	10%	0.26	0.02	0.07
1	**50%**	**0.37**	**0.03**	**0.13**
	90%	0.63	3.10	0.20
	10%	0.28	0.33	0.08
2	**50%**	**0.38**	**0.37**	**0.24**
	90%	0.67	8.61	0.34
	10%	0.23	0.02	<0.005
3	**50%**	**0.47**	**0.75**	**0.04**
	90%	1.04	7.89	0.14
	10%	0.20	0.03	<0.005
4	**50%**	**0.50**	**0.95**	**0.01**
	90%	1.25	9.73	0.05
	10%	0.31	0.81	0.06
5	**50%**	**0.46**	**0.99**	**0.13**
	90%	0.70	2.00	0.21
	10%	0.20	0.06	<0.005
6	**50%**	**0.51**	**1.05**	**<0.005**
	90%	1.37	14.19	0.01
	10%	0.20	0.06	<0.005
7	**50%**	**0.52**	**1.14**	**<0.005**
	90%	1.36	13.75	0.02
	10%	0.32	0.37	0.07
8	**50%**	**0.55**	**2.84**	**0.14**
	90%	0.89	4.72	0.23
	10%	0.37	0.13	0.04
9	**50%**	**0.57**	**6.33**	**0.09**
	90%	0.94	16.69	0.16
	10%	0.30	2.36	0.10
10	**50%**	**0.65**	**8.34**	**0.17**
	90%	1.27	14.43	0.24

Table 3.3: Posterior quantiles for lambda (loss rate) and psi (pct difference between pseudo-deductible and policy deductible) for each type, and the proportion of households in each type.

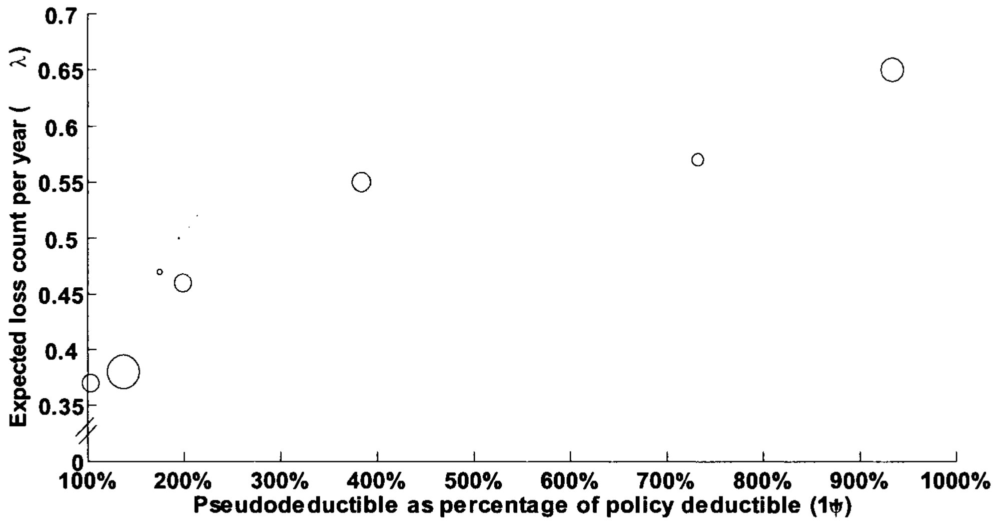


Figure 3.4: Estimates of the loss rate and pseudodeductible factor

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. While this model suggests that this relationship is static, I will show in section 3.6 that the pseudodeductible actually evolves in magnitude from claim to claim.

3.4.4 Posterior Prediction

With simulated draws from the posterior distribution in hand, I now show how the pseudodeductible model can explain those features of the data, of greatest interest, that were presented in the introduction. The method to accomplish this is the posterior predictive check (Rubin, 1984; Gelman, Meng, and Stern, 1996). The idea behind using a posterior predictive check (PPC) is that a model fits well if simulated datas ets 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 are most important. If the simulated test statistics appear to be generated from a model that is consistent with the observed ones, then one 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 3.5. However, one of the additional features of the data in which insurers are interested, as described in the introduction, is the proportion of households with at least one small claim. I 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 3.6 plots these PPCs. From these plots, the Identity model overestimates the percentage of households that file small claims, while 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 the central story–that the size of the pseudodeductible and the propensity to file small claims are intertwined. Including a pseudodeductible in the model permits replication of this important feature of the data almost exactly.

3.5 Inference

3.5.1 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

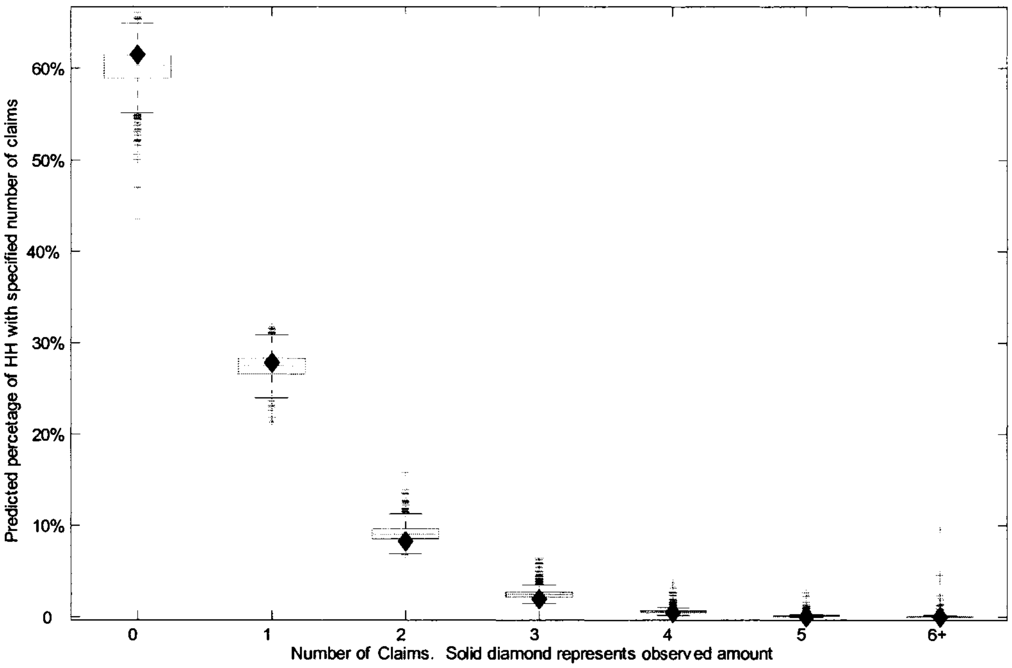


Figure 3.5: Posterior predictive intervals for number of claims

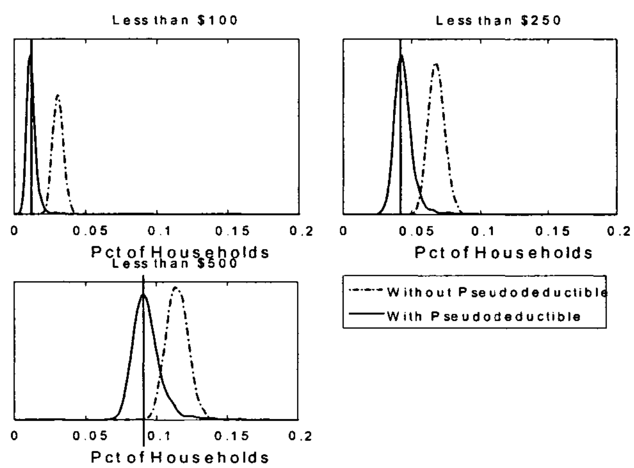


Figure 3.6: Posterior predictive intervals for percentage of households filing at least one claim

household h is drawn from component $l = 1 \dots L$ of the posterior mixing distribution. Applying Bayes's Theorem

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

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 one takes 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 3.4, and the means and 10% and 90% quantiles of the posterior distributions of π_l^h, the posterior classification probabilities, are presented in Table 3.5. Household A filed no claims during the observation period. There are two stories that might 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 one were to assume that because household A has no claims its loss rate must also be low, then one would be ignoring the 37% total posterior probability that A is in one of the three "high (λ, ψ)" types, as described in Table 3.5. But even if A were from one of these types, it is still possible that either no losses occurred or that there were some 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 consider two households with exactly one claim, B and C. If one were to consider

	Severity of Claim			
HH	claim 1	claim 2	claim 3	claim 4
A				
B	177			
C	35,120			
D	343	25,400		
E	350	482		
F	12,750	12,680		
G	5,587	7,010	7,255	
H	948	1,262	1,522	2,834

Table 3.4: Claim severity data for selected households

		1	2	3	4	5	6	7	8	9	10
Proportion		0.13	0.24	0.04	0.01	0.13	<0.5	<0.5	0.14	0.09	0.17
Exp. Losses		0.37	0.38	0.47	0.5	0.46	0.51	0.52	0.55	0.57	0.65
PsDed Factor		103%	137%	175%	195%	199%	205%	214%	384%	733%	934%
	10%	6.1%	6.2%	0.3%	<0.5%	6.6%	<0.5%	<0.5%	6.3%	2.6%	6.2%
A	**mean**	**15%**	**24%**	**6%**	**2%**	**14%**	**1%**	**1%**	**13%**	**10%**	**14%**
	90%	22%	35%	14%	7%	22%	1%	3%	21%	22%	22%
	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
B	**mean**	**26%**	**44%**	**7%**	**3%**	**2%**	**1%**	**1%**	**7%**	**7%**	**0%**
	90%	43%	71%	23%	7%	0%	1%	2%	35%	29%	3%
	10%	6%	9%	<0.5%	<0.5%	6%	0%	0%	8%	4%	8%
C	**mean**	**11%**	**19%**	**6%**	**2%**	**12%**	**1%**	**1%**	**15%**	**11%**	**23%**
	90%	17%	28%	14%	6%	18%	1%	3%	22%	20%	38%
	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D	**mean**	**64%**	**2%**	**11%**	**4%**	**3%**	**1%**	**1%**	**2%**	**8%**	**4%**
	90%	100.0%	8%	44%	9%	5%	0%	3%	8%	23%	12%
	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
E	**mean**	**66%**	**14%**	**0%**	**0%**	**5%**	**0%**	**1%**	**1%**	**10%**	**0%**
	90%	100%	43%	2%	3%	13%	1%	1%	2%	46%	4%
	10%	2%	5%	0%	0%	4%	<0.5%	<0.5%	5%	2%	5%
F	**mean**	**8%**	**14%**	**6%**	**3%**	**10%**	**1%**	**1%**	**15%**	**13%**	**30%**
	90%	14%	25%	16%	8%	18%	1%	3%	26%	28%	61%
	10%	1%	2%	0%	<0.05%	2%	<0.5%	<0.5%	2%	1%	2%
G	**mean**	**6%**	**11%**	**6%**	**3%**	**8%**	**1%**	**2%**	**14%**	**14%**	**34%**
	90%	11%	25%	19%	8%	16%	1%	3%	31%	35%	77%
	10%	<0.05%	0%	0%	0%	2%	0%	0%	0%	0%	0%
H	**mean**	**7%**	**13%**	**9%**	**4%**	**17%**	**1%**	**3%**	**32%**	**10%**	**5%**
	90%	15%	27%	26%	12%	37%	1%	5%	67%	40%	15%

Table 3.5: Posterior means of classification probabilities for selected households

the number of claims alone, one would predict identical loss rates for both of these households. But since the size of the claim for household B is only \$177, 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 one looks 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 D 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. Household F, on the other hand, did not file small claims, so there is a non-zero probability that this household is 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, while the story of few losses in the first place" is more plausible for E. It is the size of the *smallest* claim that informs which story applies to which household. These patterns apply to households with three or four claims (households G and H) as well.

3.5.2 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 I described in section 3.5.1, one 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 claim tend to have lower expected pseudodeductibles than those with no claims (since a relatively small claim could make higher pseudodeductibles impossible). For example, for household C in Table 3.5, the posterior expected ψ is 4.30. That household has a \$500 policy deductible, so its pseudodeductible is about \$2,150. Since household C would not file a claim on a loss below \$2,150 anyway, it could have saved money on premiums by taking a policy deductible of \$1,000 or \$2,000. In fact, 52 percent of all households with at least one claim (and eighty percent of all households in the 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.

3.6 Incorporating Nonstationarity

While the static multiplicative pseudodeductible version of the model explains the observed distributions of claim counts and severities, as well as the percentage of households filing small claims, it does not explain well the remaining two test statistics of interest: 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, the average increase from claim 1 to claim 2 is \$114. Also, 53 percent 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, I let each ψ_h adjust to a new value after each claim. Thus, a pseudodeductible factor, ψ_{hk}, is the pseudodeductible in force for household h immediately *before* the k^{th} claim. So, from the start of the observation period until the time of the first claim, the pseudodeductible is $\Psi\left(\bar{D}_{h1}, \psi_{h1}\right)$; between claims 1 and 2, the pseudodeductible is $\Psi\left(\bar{D}_{h2}, \psi_{h2}\right)$; and so forth.

To determine ψ_{hk+1}, multiply 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. Since 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 web site visits.

In this 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. Heterogeneity on b_h is incorporated 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. Since each v_{lk} is then drawn from a lognormal distribution with parameters b_l and σ^2, each household of type l draws a random vector $(\lambda_l, \psi_l, b_l, v_{l1} \ldots v_{lK_h})$.

This nonstationary model is estimated 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 non-empty components and was chosen in the same manner as described in section 3.3.2). The log marginal likelihood of the nonstationary multiplicative model is -6278, a dramatic improvement over that of the static model (see Table 3.1). The quantiles of these components and proportions are described in Table 3.6. 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, pseudodeductibles generally increase from claim to claim (this is clearly not true for every household or every claim, since those draws of v_{li} that are less than one will trigger a decrease in the pseudodeductible). Pseudodeductibles should increase for the same reasons that I discussed in section 3.4.3–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 3.7 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. Note that the stationary models are miscalibrated for these inherently nonstationary test statistics, and that the miscalibration is corrected when nonstationarity is added to the model.

This is a particularly important result, since 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 and more claims (or become more brazen in filing large

Type	Quantile	lambda	psi	median v	proportion
	10%	0.33	0.001	1.06	0.04
1	**50%**	**0.52**	**0.001**	**4.90**	**0.07**
	90%	0.77	0.001	31.19	0.11
	10%	0.19	<0.01	0.62	0.05
2	**50%**	**0.32**	**0.08**	**3.03**	**0.12**
	90%	0.52	1.68	15.03	0.17
	10%	0.18	0.93	0.79	0.07
3	**50%**	**0.30**	**1.75**	**3.53**	**0.13**
	90%	0.51	3.04	15.80	0.22
	10%	0.37	13.39	0.50	0.16
4	**50%**	**0.51**	**14.12**	**2.51**	**0.26**
	90%	0.73	14.94	11.59	0.34
	10%	0.32	6.65	0.49	0.08
5	**50%**	**0.45**	**7.43**	**2.05**	**0.16**
	90%	0.62	8.32	8.67	0.25
	10%	0.31	0.03	0.73	0.03
6	**50%**	**0.46**	**0.06**	**3.19**	**0.07**
	90%	0.70	1.04	15.49	0.11
	10%	0.22	0.04	0.57	0.02
7	**50%**	**0.40**	**0.82**	**2.94**	**0.05**
	90%	0.76	9.07	16.28	0.11
	10%	0.37	0.17	1.36	0.05
8	**50%**	**0.52**	**0.27**	**5.93**	**0.10**
	90%	0.70	0.54	30.88	0.14
	10%	0.23	0.05	0.47	<0.01
9	**50%**	**0.58**	**0.93**	**2.72**	**0.01**
	90%	1.21	19.04	13.60	0.04
	10%	0.17	0.12	0.46	<0.01
10	**50%**	**0.46**	**3.38**	**2.69**	**<0.01**
	90%	1.25	18.13	14.30	0.02
	10%	0.18	0.04	0.55	<0.01
11	**50%**	**0.50**	**0.99**	**3.06**	**<0.01**
	90%	1.43	11.54	17.29	0.01

Table 3.6: Posterior quantiles for parameter estimates in nonstationary model

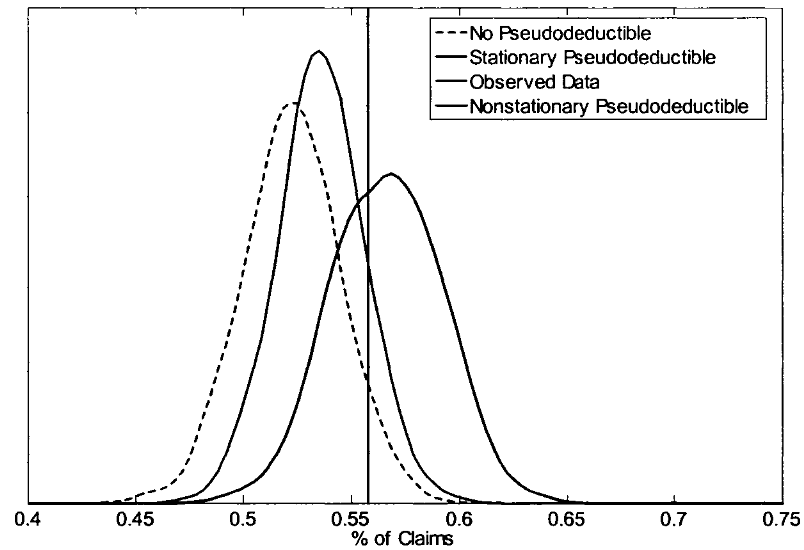


Figure 3.7: Posterior predictive interval for nonstationary model

claims). Instead, these increases may be caused by customers who become more and 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 \$1200 and \$2500. If one looks only at the claims, one might think that the household is becoming more risk prone, since the severity of the claims is going up. But now suppose that after the first claim, the pseudodeductible increases from \$800 to \$1500, and that the household experiences a loss of \$1000 sometime between the two claims. The \$1000 loss is unobserved, since it is less than the new \$1500 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.

3.7 Discussion

In this chapter, I 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. I model observed consumer behavior as the output of multiple error-laden stochastic processes, permitting the understanding of individual behavior despite its sometimes random and unpredictable nature. I 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 web site *visits* in a model that predicts web site *purchases*.

Given that I 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, 2004). The pseudodeductible may also play a part in the revenue-side of CLV calculations. In section 3.5.2, I mentioned that while customers with high pseudodeductibles could benefit by increasing their policy deductibles, it is unclear whether such a move is in an insurer's interest, since 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 the 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, since 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, I am interested in whether or not 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 understanding about why insurance customers behave the way they do. I did not set out to predict *ex-ante* which customers have the larger pseudodeductibles, but one could adapt the model to make the pseudodeductible a function of explanatory variables. For example, it might be interesting to see if whether or not the cause of a loss (e.g., natural factors or customer carelessness) alters the propensity of the customer to file a claim. I envision this chapter as the start of a potentially fruitful stream of research into better understanding about how and why customers "leave money on the table."

Chapter 4

Conclusion

In this dissertation I have investigated the connection between the choice of deductibles and the filing of claims from two different perspectives. First, I showed that customers' private information about their own propensity to files claims affects deductible choice, and that insurers can influence the degree to which this information plays a role. I then demonstrated that customers may decide not to file claims on some losses that exceed their chosen policy deductibles. Considered together, there is now deeper insight into how customer decisions on the cost side of the profit equation (the amount of money that is paid out in claims) influence, and are influenced by, customer decisions on the revenue side (the choice of deductible). In this chapter I discuss some opportunities for future research, and then propose some prescriptive recommendations, beyond those discussed in the previous chapters, that may aid in the decision-making processes for both insurers and customers.

Although the research in this dissertation examines some of the factors that induce homeowners to increase their deductibles, the model is clearly not a comprehensive predictive or descriptive model of deductible choice. Many factors affect deductible choice, such as inertia, mental accounting, budget constraints, but there are many others. An example is the treatment of deductibles that are denominated in terms of a percentage of coverage, as opposed to a fixed dollar amount. A complete choice model should take into account how customers perceive deductible alternatives when some are dollar-denominated and others are percentage-denominated. In a single period, there should be no difference on a $100,000 policy between a 1% deductible and $1000 deductible. But over multiple periods, the dollar value of the 1% deductible varies as the coverage amount varies. The insurer will want to know how customers process the similarities and differences of all of the available deductible choices. But these issues present challenges for modelers whose data is insufficient to identify parameters that might characterize customers' underlying preferences. Experimental research may be appropriate for this issue. In particular, conjoint analysis might be a useful way to assess the impact that the framing of a deductible (in either dollar or percentage terms) has on a customer's preferences for policies.

Developing a integrated model of deductible choice and claiming behavior is an important next step in this research stream. Suppose one wanted to learn more about the effects of asymmetric information on deductible switching and claims filing. In chapter 2, I did not differentiate between covered losses that are claimed from those that are unclaimed. But we know from chapter 3 that customers do, in fact, forgo reimbursement for losses whose severities are less than the pseudodeductible. An integrated model might be able to determine whether the "riskiness" that is correlated with deductible choice is due to underlying propensities to experience losses, or the selectivity of the customer when filing claims on those losses.

These two separate mechanisms, with different behavioral implications, are indistinguishable when examining claims history alone. Consequently, insurers who want to influence the degree of asymmetric information in a system might choose different tactics if the correlations are due to customers investment in safety (which is related to the rate of underlying losses), instead of the customer's propensity to file claims (which is related to the pseudodeductible). For example, if the insurer wanted to mitigate the mental budget pressures from a premium increase without triggering an increase in risky behavior, the insurer might offer rebates to customers for investment in risk mitigation. This could be a potentially good deal for the insurer, since the customer might be less likely to choose a less expensive high-deductible policy, while simultaneously reducing the expected indemnity to be paid on that policy. Similarly, if the insurer wanted to deter customers from filing small claims, he might offer a bonus or rebate in each year that a small claim is not filed. In this case, the insurer would retain the premium revenue associated with the lower deductible, while discouraging the filing of claims that are disproportionately expensive to service.

Within insurance companies, premiums are often set by actuarial teams who estimate the amount of future claims that are expected to be paid out. Marketing and sales organizations then use these premiums when writing policies for customers. This practice fails to take into account the effect that changes in premiums have on customers' subsequent choices. From chapter 2, customers may change their policy deductibles in response to changes in premiums. By making that switch, a customer is sending a signal to the insurer about his future propensity to file claims. Thus, deductible choice influences not only the lowest possible severity of loss that could be filed as a claim, but may also be correlated with the rate of occurrences of losses themselves. At the same time, if one believes that a customer's pseudodeductible is proportional to the policy deductible, then the increase in deductible could have a multiplicative effect in decreasing claims in the future. Hence, the research in this dissertation offers insurers an additional tool with which to assess the riskiness of individual customers. If insurers were to take this information into account, then customers have an added benefit of increasing their deductibles. In addition to saving money on premiums by avoiding low deductible surcharges, they may also save money on the risk-based portion of their premiums as well.

Not all customers who increase their deductibles are necessarily high-risk types. But a customer who does increase his deductible is more likely to be a high-risk type if he switches in response to marketing or pricing cues. And since this probability is even higher for customers with low coverage amounts, insurers could influence the proportion of high-risk customers at each deductible level by targeting who receives the cues. If one were able to assess the effect of these cues on either loss frequency or claims selectivity, the opportunities for targeting customers based on pseudodeductible become even more powerful.

Integrated research offers opportunities for customers to improve their own decision-making as well. For example, careful consideration of deductible choices could reveal the preferability of some deductibles over others. From chapter 3, note that the majority of customers in the sample have expected posterior pseudodeductibles that exceed the next highest level of deductible that is available. This customer might be able to save on premium expenditures by increases his policy deductible without decreasing the amount of indemnity he receives for a loss. The reason for this is that if the customer will absorb a certain amount of a loss anyway, he is better off choosing the highest deductible that is less than that threshold. The exact amount of this savings is difficult to measure, because under a multiplicative model of pseudodeductibles, the pseudodeductible should increase even more after a deductible increase. But if this were to happen, there is even more of an opportunity for savings. Hence, the qualitative result remains. The only way in which a deductible increase could cost the customer would be if the pseudodeductible *level* were a decreasing function of the policy deductible on some part of the domain of available deductibles. It is hard to envision a plausible story for such a model.

Nevertheless, an exciting facet of this dissertation is the confirmation that there is so much more that needs to be discovered through observational and empirical research into insurance decisions. A researcher could take an economic approach to the problem, as I did in chapter 2, or a more model-based approach, as I did in chapter 3. Each approach can reveal different aspects of the decision-making process. This dissertation is just a starting point.

Bibliography

Abbring, Jaap H., Pierre-Andre Chiappori, James J. Heckman, and Jean Pinquet. 2003a. Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish? *Journal of the European Economic Association* 1(2-3):512–521.

Abbring, Jaap H., Pierre-Andre Chiappori, and Jean Pinquet. 2003b. Moral hazard and dynamic insurance data. *Journal of the European Economic Association* 1(4):767–820.

Antoniak, C.E. 1974. Mixtures of dirichlet processes with applications to nonparametric problems. *Annals of Statistics* 2(4):1152–1174.

Barnett, A. I., Cynthia J. Rutherford, Jane F. Desforges, Nancy Gutensohn, and B. Davies. 1981. Evaluating the validity of a bayesian program for predicting stage in hodgkin's disease. *Methods of Information in Medicine* 20(1):174–178.

Bell, David E. 1982. Regret in decision making under uncertainty. *Operations Research* 30(5):961–981.

Berger, Paul D., and Nada I. Nasr. 1998. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing* 12(1):17–30.

Bernardo, Jose M., and Adrian F. M. Smith. 2000. *Bayesian theory*. 2nd ed. Chichester: John Wiley and Sons.

Braun, Michael, and Alexander Muermann. 2004. The impact of regret on the demand for insurance. *Journal of Risk and Insurance* 71(4):737–767.

Bureau of Labor Statistics. 2004. Consumer price index. Web Page. www.bls.gov/cpi, series ID cuur0200sa0,cuus0200sa0.

Cawley, John, and Tomas Philipson. 1999. An empirical examination of information barriers to trade in insurance. *American Economic Review* 89(4):827–846.

Chiappori, Pierre-Andre. 2000. Econometric models of insurance under asymmetric information. In *Handbook of insurance*, ed. Georges Dionne, 365–393. Boston: Kluwer Academic Publishers.

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. Testing for asymmetric information in insurance markets. *Journal of Political Economy* 108(1):56–78.

———. 2003. Testing contract theory: A survey of some recent work. In *Advances in economics and econometrics–theory and applications, eighth world congress*, ed. M. Dewatripoint, L. Hansen, and P. Turnovsky, 115–149. Cambridge: Cambridge University Press.

Chib, Siddhartha, and Edward Greenberg. 1998. Analysis of multivariate probit models. *Biometrika* 85(2):347–361.

Choi, James J., David Laibson, and Brigitte C. Madrian. 2005. 100 dollar bills on the sidewalk: Violations of no-arbitrage in 401-k accounts. NBER Working Paper No. 11554.

Congdon, Peter. 2001. *Bayesian statistical modelling*. Wiley Series in Probability and Statistics, Chichester: John Wiley and Sons.

De Leve, G., and P. J. Weeda. 1968. Driving with markov-programming. *ASTIN Bulletin* 5(1):62–86.

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

Dionne, Georges, Christian Gourieroux, and Charles Vanasse. 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. *Journal of Political Economy* 109(2):444–453.

Dionne, Georges, Pierre-Carl Michaud, and Maki Dahchour. 2004. Separating moral hazard from adverse selection in automobile insurance: Longitudinal evidence from france. Working Paper 04-05, HEC Montreal.

Doherty, Neil A., and Louis Eeckhoudt. 1995. Optimal insurance without expected utility: The dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty* 10(2):157–179.

Doherty, Neil A., and Harris Schlesinger. 1990. Rational insurance purchasing: Consideration of contract nonperformance. *Quarterly Journal of Economics* 105(1):243–253.

Draper, David. 1999. Discussion on "Bayesian Nonparametric Inference for Random Distributions and Related Functions" . *Journal of the Royal Statistical Society, Series B* 61(3):510–513.

Edwards, Yancy D., and Greg M. Allenby. 2003. Multivariate analysis of multiple response data. *Journal of Marketing Research* 40(3):321–334.

Escobar, Michael D., and Mike West. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90(430):577–588.

Fader, Peter S., and Bruce G. S. Hardie. 2000. A note on modelling underreported poisson counts. *Journal of Applied Statistics* 27(8):953–964.

———. 2004. Calculating customer lifetime value in a contractual setting. Working paper, marketing department, Wharton School, University of Pennsylvania.

Ferguson, T.S. 1983. Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, ed. H. Rizvi and J. Rustagi, 287–302. New York: Academic Press.

Finkelstein, Amy, and Kathleen McGarry. 2003. Private information and its effect on market equilibrium: New evidence from long-term care insurance. NBER Working Paper No. 9957.

Finkelstein, Amy, and James Poterba. 2004. Adverse selection in insurance markets: Policyholder evidence from the u.k. annuity market. *Journal of Political Economy* 112(1): 183–208.

Gallant, A. Ronald, and George Tauchen. 1996. Which moments to match? *Econometric Theory* 12(4):657–681.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian data analysis*. Boca Raton, Fla.: Chapman and Hall.

Gelman, Andrew, Xiao-Li Meng, and Hal Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6(4):733–807.

Geweke, John F. 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computer science and statistics: Proceedings of the twenty-third symposium on the interface*, 571–578. Alexandria, Va.: American Statistical Association.

———. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics*, ed. Jose M. Bernardo, James O. Berger,

A. Philip Dawid, and Adrian F. M. Smith, vol. 4, 169–194. Oxford: Oxford University Press.

Gourieroux, Christian, A. Monfort, and E. Renault. 1993. Indirect inference. *Journal of Applied Econometrics* 8:S85–S118.

Grace, Martin F., Robert W. Klein, Paul R. Kleindorfer, and Michael R. Murray. 2003. *Catastrophe insurance: Consumer demand, markets and regulation.* Boston: Kluwer Academic Publishers.

Grossman, Herschel I. 1979. Adverse selection, dissembling, and competitive equilibrium. *The Bell Journal of Economics* 10(1):336–343.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart. 2004. Valuing customers. *Journal of Marketing Research* 41(1):7–18.

Hardie, Bruce G. S., and Peter S. Fader. 2005. Applied probability models with applications in marketing. Preprint.

Hastings, N. A. J. 1976. Optimal claiming on vehicle insurance. *Operational Research Quarterly* 27(4, Part 1):805–813.

Heath, Chip, and Jack B. Soll. 1996. Mental budgeting and consumer decisions. *Journal of Consumer Research* 23(1):40–52.

Hsee, Christopher K., and Howard Kunreuther. 2000. The affection effect in insurance decisions. *Journal of Risk and Uncertainty* 20(2):141–159.

Israel, Mark. 2004. Do we drive more safely when accidents are more expensive? identifying moral hazard from experience rating schemes. Working Paper, Northwestern University.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther. 1993. Framing, probability distortions and insurance decisions. *Journal of Risk and Uncertainty* 7: 35–51.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*, vol. 1. New York: John Wiley and Sons, Inc.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy* 98(4): 1325–1348.

———. 1991. Anomolies: The endowment effect, loss aversion and status quo bias. *Journal of Economic Perspectives* 5(1):193–206.

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

———. 1984. Choices, values and frames. *American Psychologist* 39(4):341–350.

Keane, Michael P. 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62(1):95–116.

Kim, Jin Gyo, Ulrich Menzefricke, and Fred M. Feinberg. 2004. Assessing heterogeneity in discrete choice models using a dirichlet process prior. *Review of Marketing Science* 2(1):1–39.

Kleindorfer, Paul R., and Howard Kunreuther. 1983. Misinformation and equilibrium in insurance markets. In *Economic analysis of regulatory markets*, ed. Joerg Finsinger, 67–90. New York: St. Martin's Press.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 1998. *Loss models: From data to decisions*. Wiley Series in Probability and Statistics, New York: John Wiley and Sons.

Kulkarni, Vidyadhar G. 1995. *Modeling and analysis of stochastic systems*. London: Chapman and Hall.

Kunreuther, Howard, and Mark Pauly. 2005. Insurance decision making and market behavior Working paper, Wharton School, University of Pennsylvania.

Haehling von Lanzenauer, Christoph. 1974. Optimal claim decisions by policyholders in automobile insurance with merit-rating structures. *Operations Research* 22(5):979–990.

Lemaire, Jean. 1995. *Bonus-malus systems in automobile insurance*. Huebner International Series on Risk, Insurance and Economic Security, Dordrecht, Netherlands: Kluwer Academic Publishers.

Loomes, Graham, and Robert Sugden. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal* 92(368):805–824.

Machina, Mark J. 2000. Non-expected utility and the robustness of the classical insurance paradigm. In *Handbook of insurance*, ed. Georges Dionne, 37–96. Huebner International Series on Risk, Insurance and Economic Security, Boston: Kluwer Academic Publishers.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta. 1999. The shopping basket: A model for multicategory purchase incidence decisions. *Marketing Science* 18:95–114.

Moe, Wendy W., and Peter S. Fader. 2004a. Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing* 18(1):5–19.

———. 2004b. Dynamic conversion at e-commerce sites. *Management Science* 50(3): 326–335.

Moffitt, Robert. 1983. An economic model of welfare stigma. *American Economic Review* 73(5):1023–1035.

Mossin, Jan. 1968. Aspects of rational insurance pricing. *Journal of Political Economy* 76(4-1):553–568.

Newhouse, Joseph P., John E. Rolph, Bryant Mori, and Maureen Murphy. 1980. The effect of deductibles on the demand for medical care services. *Journal of the American Statistical Association* 75(371):525–533.

Newton, Michael A., and Adrian E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B* 56(1):3–48.

Norman, J.M., and D.C.S. Shearn. 1980. Optimal claiming on vehicle insurance revisited. *The Journal of the Operational Research Society* 31(2):181–186.

Pashigian, B. Peter, Lawrence L. Schkade, and George H. Menefee. 1966. The selection of an optimal deductible for a given insurance policy. *Journal of Business* 39(1):35–44.

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

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

Puelz, Robert, and Arthur Snow. 1994. Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy* 102(2):236–257.

Rothschild, Michael, and Joseph E. Stiglitz. 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* 90(4):629–649.

Rubin, Donald B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* 12(4):1151–1172.

Samuelson, William, and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1(1):7–59.

Schlesinger, Harris. 1981. The optimal level of deductibility in insurance contracts. *Journal of Risk and Insurance* 48(3):465–481.

Schmittlein, David C., Albert C. Bemmaor, and Donald G. Morrison. 1985. Why does the nbd model work? robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Science* 4(3):255–266.

Sethuraman, J. 1994. A constructive definition of dirichlet priors. *Statistica Sinica* 4(2): 639–650.

Slovic, Paul, Baruch Fischoff, Sarah Lichtenstein, Bernard Corrigan., and Barbara Combs. 1977. Preference for insuring against probable small losses: Insurance implications. *Journal of Risk and Insurance* 44(2):237–258.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Gilks. 1996. BUGS: Bayesian Inference Using Gibbs Sampling. Computer Software, Version 1.4, www.mrc-bsu.cam.ac.uk/bugs/.

Stephens, Matthew. 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B* 62(4):795–809.

Thaler, Richard H. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1(1):39–60.

———. 1999. Mental accounting matters. *Journal of Behavioral Decision-Making* 12: 183–206. Reprinted in K-T 2000, Ch. 14.

Train, Kenneth E. 2003. *Discrete choice methods with simulation.* Cambridge: Cambridge University Press.

Venezia, Itshak. 1984. Aspects of optimal automobile insurance. *Journal of Risk and Insurance* 51(1):63–79.

Venezia, Itzhak, and Haim Levy. 1980. Optimal claims in automobile insurance. *Review of Economic Studies* 47(3):539–549.

Walker, Stephen G., Paul Damien, Purushottam W. Laud, and Adrian F. M. Smith. 1999. Bayesian nonparametric inference for random distributions and related functions. *Journal of the Royal Statistical Society, Series B* 61(3):485–527.

Wilson, Charles. 1977. A model of insurance markets with incomplete information. *Journal of Economic Theory* 16(2):167–207.

Yaari, M. E. 1987. The dual theory of choice under risk. *Econometrica* 55(1):95–115.