DEPARTMENT OF MANAGEMENT

AFDELING FOR VIRKSOMHEDSLEDELSE

Working Paper 2006-5

A Quantitative Model of Dynamic
Customer Relationships

Niels Stender

UNIVERSITY OF AARHUS • DENMARK

ISSN 1398-6228

A Quantitative Model of Dynamic Customer Relationships

Niels Stender
University of Aarhus
niels@innohead.com

October 15, 2005

Abstract

A discrete semi-markov process for modeling the evolution of customer relationships is presented. The transition probability function is formulated conditionally on individual observed and unobserved heterogeneity, using a bayesian hierarchical formulation of the latter.

The model is used in a study of the customer transaction history of a telecommunications company to derive churn and revenue-forecasts on a customer level. Estimation is carried out using a Metropolis-Hastings MCMC method.

1 Introduction

The formation of an ongoing relationship between a consumer and a supplier is an integral feature of many products and services. The relationship can be dynamic in the sense that consumption intensity and type of product or service involved changes dynamically. Financial performance of such companies depends decisively on the distribution of type and intensity in the customer base, so an understanding of how and why the relationship change through time is an important contributor to informed decision making.

Properties of such customer relationships have been studied using a patchwork of marketing research methods such as customer satisfaction, segmentation and churn studies. Attempts to develop a quantitative, rather than qualitative, methodology for a unified framework of analysis have not gained attention until recently.

Schmittlein and Peterson (1994) presented one of the earliest models dealing with relationship duration and revenue, while Reinartz and Kumar (2003) and Donkers et al. (2003) are recent examples. The former paper also presents a brief overview of the literature on stochastic models of relationship durations. A related strain of research is the literature on purchase timing, originating from the challenges posed by retail sector scanner data, such as Manchanda et al. (1999) and Chib et al. (2002).

The purpose of this paper is to develop methodology aiding the empirical study of patterns in dynamic customer relationships and revenue, based on customer transaction data already existing in many companies' data archives.

Customer relationships are seen as discrete stochastic processes with multiple levels. The most basic construct is the notion of a state. At any given point in time, the bond between an agent and a company is thought to occupy exactly one state among a finite set of states. This set could contain states with labels such as "Not a customer", "Subscribing to service A", "Subscribing to service B" and "Subscribing to both A and B".

A probability law of transitions between states conditional on observed and unobserved heterogeneity are formulated in section 2. This serves as a foundation for individual level measures of the probability of a future deterioration or improvement of the relationship, simultaneously allowing the researcher to study antecedents of transitions.

The full maximum likelihood involves complicated integrals, hence an approach to estimation through MCMC methods of the Metropolis-Hastings type is presented in section 2.4.

An application of the ideas is carried out in section 3. The customer base of a telecommunications company provides the setting for this task, as customer relations can last for years and the range of services consumed easily change over time. Customer dynamics conditional on sociodemographic variables are explored, and finally, the out-of-sample predictive performance is studied.

1.1 A Note on the Theoretical Approach

Why present a statistical model framework with little embedded behavioral theory, instead of employing a model with a foundation on qualitative marketing theory or microeconomics? The selfimposed data limitation makes the proposition too ambitious. Lack of comprehensive data on the environment and individuals information sets, such as attributes of competitive offerings, the individual's knowledge of and preference for these attributes, knowledge of competitive product attributes, switching costs and so on induces a vast range of problems for scientific inference regarding the individuals decision process.

With respect to representation, even knowing how the individual makes his decisions a priori would not result in empirical models mathematically different from those based on pure probabilistic arguments. Interpretations of resulting parameter estimates would differ, however.

There are very few studies of consumer behavior where the researcher can claim to know most qualitatively important aspects of the information set that the agent is basing his decisions upon. Miravete and Palacios-Huerta (2002) is one of the rare exceptions.

Competitors simultaneously change the properties of their product offerings, new products arrive, and customers gradually learn about new offerings and may even change inherent preferences. For marketing as a science it is unfortunate that no single factor in the market can be held constant, impeding efforts to find useful laws.

1.2 Issues in the Modeling of Transition Data

In forming the probability law governing a discrete stochastic process, several areas from statistics and econometrics intersect. One is that of discrete panel data methods and discrete choice models, while the other is event history analysis with models of durations as a special case. Duration models are concerned with time dependence, specifically time spent occupying a given state and how that affects the probability of continually occupying the state.

Functional form issues aside, the key problems in the modeling of transition data are that of *state dependence* and *unobserved heterogeneity*.

If the distribution of the state of a service in period t+1 depends on the state in period t or earlier, some form of state-dependence is at play. In the absence of state dependence, the sequence of events (0, 1, 0, 1) for a given service would be just as likely as (1, 1, 0, 0). This will not hold in most realistic cases. For a customer having bought into a given service, financial switching costs, through fees required to terminate a contract or entering into new ones, can induce state dependence. Other possibilities include indirect costs such as time lost due to the actual transition, costs of search and information, psychological or rational habit formation, learning and perceived risks are other arguments for state dependence. Lack of state dependence should only be expected in the extreme case of zero transaction costs, a transparent market, identical products, a constant demand and a type of product that cannot be hoarded.

Unaccounted unobserved heterogeneity can lead to spurious state dependence being mistaken for true state dependence as argued by Heckman (1981), Lancaster (1990) and Hsiao (2003). That can in fact amount to fundamental errors of inference regarding the causality of heterogeneity and events; hence a realistic model of relationship transition data must be able to address state dependence and unobserved heterogeneity.

Allenby and Rossi (1999) and Rossi and Allenby (2000) outlines arguments in favor of Bayesian MCMC-methods when dealing with marketing applications. Semi-parametric and non-parametric methods have been increasingly popular in many social sciences, as they lessen the number of assumptions necessary to make on the data generating process. In a marketing context, however, they have two important drawbacks. Firstly, heterogeneity is seen as a nuisance parameter and is often conditioned away. In marketing, the anatomy of heterogeneity can have important policy implications and might even be the object of the study. Secondly the marketing analyst will frequently find individual-level parameter estimates useful. With the usual panel data case, with a large number of individuals and a small number of time periods. In Bayesian MCMC methods, information about individual level heterogeneity parameters and the population distribution of a given parameter is used simultaneously through “parameter shrinkage” to back out an individual estimate.

These advantages come at the price of likely misspecification. The misspecification issues are in practice checked through predictive validity and graphic diagnostics, as presented in Allenby and Rossi (1999), and by more flexible models.

2 A Model of Customer Relations

At any given point in time *t*, a relation to agent *k* will occupy exactly one state s_{kt} among a finite set of states ? . The *relation state space* ? will be enumerated by integers from 0 and up to just below total number of states for notational convenience. In the introductory example we would have ? ? $(0,1,2,3)$.

The meaning of the states in ? controls how detailed behaviors the model will be able to account for, so defining too small a state space could easily hide important insight. The data quality, time and computer power available will on the other hand put an upper limit to the size of the state space, as the number of parameters to be estimated grows in an order corresponding to the square of the number of states in the model. This growth can be curbed by imposing prior restrictions, such as disallowing some transitions and preventing some heterogeneity variables from influencing the probability law in some states.

2.1 A Taxonomy of the Relation State Space

A brief suggestion for a taxonomy of the structure of relation state spaces follows. Given a set of *K* services or products one could conceive of several methods of constructing a state space. One extreme would be a *horizontal* ordering, consisting of a full enumeration of all possible service configurations as done in the introductory example. That would yield a state space consisting of 2^K states, which would be unwieldy for all but very small *K*. The opposite extreme would be a *vertical* ordering, if there is a natural ranking of the services. That could be a *cumulative* ranking, as when subscribing to service B implies that one also subscribes to services A, but not the opposite way around. Alternatively it could be a *mutually exclusive* ranking, if it is not possible to subscribe to service A and B at the same time. For a given *K* the vertical construction schemes will yield smaller state spaces, but for some services it will not be meaningful. In some cases a hybrid of the horizontal and vertical scheme should be used, such as when service A and B are mutually exclusive, but service C is available in conjunction with either of them.

The type of model presented in this paper is most useful for scenarios that allow for vertical orderings of the state space, while cases with inherent horizontal orderings should be studied using the type of models suggested in the purchase timing and scanner data literature, such as the high-dimensional probit model that recently has seen methodological advances. Edwards and Allenby (2003) demonstrates several applications of the high-dimensional probit model.

2.2 A Probability Law of Transitions

Consumer behavior, in a limited and formalized fashion, can now be studied as transitions between states in the relationship state space. One of the most useful mathematical devices conceived for studying such transitions is that of a markov chain[1].

[1] For an introduction see Hoel, Paul G., Sidney C. Port, and Charles J. Stone (1987), Introduction to Stochastic Processes. Boston: Waveland Press, Inc.

Even if no data but state transitions and times were available, it would still be interesting for a decision maker to study these to gain insight into the aggregate dynamics of the customer base. See Morrison et al. (1982) for an early application and Donkers et al. (2003) for a current. Both papers implicitly assume homogenous transition probabilities in the markov chain over all agents. This cannot realistically be true, indeed useful knowledge is learned through the heterogeneity in those. A probability law is to be formulated, relating transitions to agent characteristics.

Let t_e denote the time of entry into a the state *s*. Let $x(t)$ designate a vector of individual specific information, the first element fixed to 1 as a scale normalization. In typical applications the individual specific information would be agent and service-choice characteristics such as demographics and price.

Any function that maps from the space of $x(t)$ to a suitable space of sets of probabilities could be used. A specific functional form is given by thinking of state transitions as a sequential series of continuous competing risk durations observed at discrete intervals. The advantage of this approach lies in having the hazard function as a point of departure for the analyst. This helps interpretation of parameters, later model extensions, comparisons to existing models in the statistical literature and even though data almost always is observed at discrete intervals, it is often easier to formulate behavioral relations in continuous time.

Let $\lambda_{sd}(t)$ denote the intensity of moving from state *i* to state *j*.

$$\lambda_{sd}\left(t\right) = e^{\beta_{sd}' x(t)},\ s \in \Omega,\ d \in \Omega \setminus \{s\}, \tag{1}$$

where β_{sd} is a vector of parameters controlling the influence of $x(t)$ on the hazard. The functional form guarantees a non-negative hazard and is a standard form for hazards. If $x(t)$ is constant over all t, eq. (1) conditional on β_{sd} results in exponentially distributed durations of stays in a given state, as no duration dependence is implied. The customer will be just as likely to leave the state in the beginning of a spell as in the end. When unobserved heterogeneity is included in section 2.3, a constant $x(t)$ will not cause duration dependence to disappear.

Conditional on unobserved heterogeneity, in a continuous setting, the survival times in a given state would be exponentially distributed. In this discrete setting, survival times would be distributed geometrically.

Eq. (1) is a simplistic hazard structure with a view to time-dependence, but in a marketing environment the individual heterogeneity can be expected to play a larger role than the inherent shape of the hazard. If duration dependence was a key element of the modeling effort, it could be incorporated using a parametric distribution such as the generalized gamma distribution. A simpler approach could be followed by including duration-specific dummies in $x(t)$. For discretely observed duration data, no hazard function more complicated than a step function can be identified anyway. With or without time-dummies, the hazard in (1) would fall in the class of proportional hazards.

Now partition the timeline into equally spaced intervals of length Δ. Assume $x(t)$ and state is observed at the endpoints of each of these intervals, such that

$$x(t) = \begin{cases} x_1 & \text{for } t \in [0, \Delta] \\ x_2 & \text{for } t \in (\Delta, 2\Delta] \\ \ldots \end{cases} \tag{2}$$

and let s_i denote the state occupied in interval *i*. Let $\lambda_i\left(t\right) = \sum_{j \neq i} \lambda_{ij}\left(t\right)$.

Now the functional form of the dependence between the discrete time transition probability and $x(t)$ is derived. Let $p_{sd}(m)$ denote the transition probability from state s to d in the interval of time $(\Delta(m-1); \Delta m]$, conditional on surviving in state s from t_e and up to time $\Delta(m\text{-}1)$. Formally,

$$
\begin{aligned}
p_{sd}(m) &= \int_{\Delta(m-1)}^{\Delta m} e^{-\int_{t_e}^{t}\lambda_s(u)du}\lambda_{sd}(t)dt \Big/ e^{-\int_{t_e}^{\Delta(m-1)}\lambda_s(u)du} \\
&= \left[-\frac{1}{\lambda_s(t)} e^{-\int_{\Delta(m-1)}^{t}\lambda_s(u)du}\lambda_{sd}(t)\right]_{\Delta(m-1)}^{\Delta m} \\
&= \left(1-e^{-\Delta\sum_{i\neq s}e^{\beta_{si}x_m}}\right)\frac{e^{\beta_{sd}x_m}}{\sum_{i\neq s}e^{\beta_{si}x_m}},
\end{aligned}
\tag{3}
$$

remembering that λ_{sd} is constant except at interval endpoints. Note how eq. (3) turns out to be independent of t_e. Note also that Δ can always be set to one by a suitable change of interval scale.

The probability of staying in the same state is then

$$
\begin{aligned}
p_{ss}(m) &= 1-\sum_{i\neq s}p_{si}(m) \\
&= e^{-\Delta\sum_{i\neq s}e^{\beta_{si}x_m}}.
\end{aligned}
\tag{4}
$$

For the continuous model to be equivalent to the discrete model, it is necessary to restrict the number of allowed state changes within an interval to one. Several state changes taking place within an interval would be a possibility in some scenarios, but an interval length of relatively short length along with administrative or psychological costs of changing a state will reduce the relevance of accounting for it. This additional restriction is not treated any further in the continuous case and is inherent to the discrete representation.

2.2.1 Sampling

The model presented so far is implicitly contingent on a special sampling scheme. Randomly sampling people entering a state during an interval of time is called *flow sampling*, while randomly sampling people at a fixed point in time is called *stock sampling*. For a customer base the former scheme could be implemented by sampling among all customers being activated subsequent to a given date, while the latter would be to sample indiscriminately among all existing customers, ignoring their activation date. Selection is independent of the individual hazard only in the flow-sampling scheme and so avoids inducing a selection bias and the corresponding issue of correction. See Lancaster (1990), chapter 5 and 8 for a detailed examination of this issue.

To illustrate the idea of stock sampling further, consider figure 2.1. Here we assume the available observation window is the interval t_0 to t_1. Flow sampling implies that all cases activated after t_0 are included, so in the figure only case A and B qualifies for inclusion.

Since customers are unobservable on the right side of the observation window, customers activated precisely on the left edge of the observation window achieve the maximum potential event history length. In the case of no customer churn, one would still see observed event histories getting shorter as people are activated at a later date.

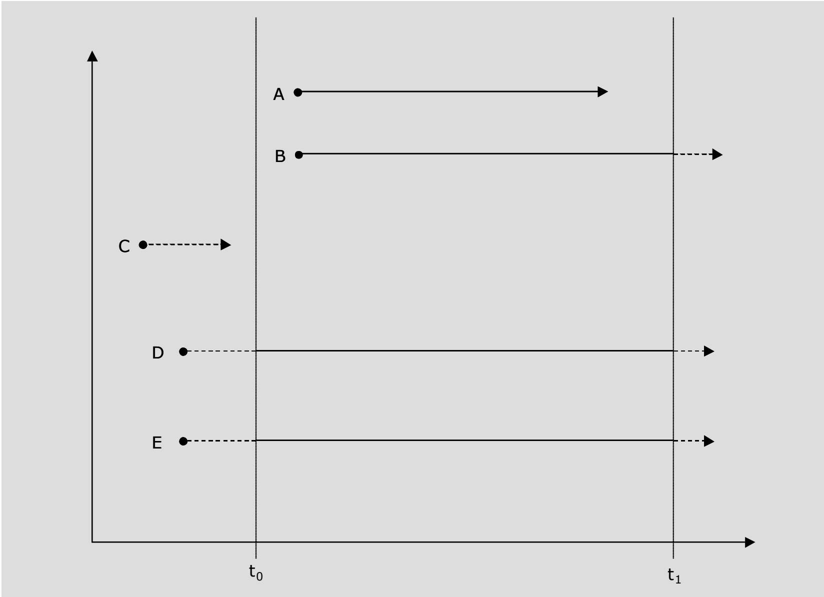


figure 2.1. Illustation of sampling scheme. Cases A and B are included. Cases C, D and E are excluded.

2.3 Unobserved Heterogeneity

Two customers with the same characteristics vector $x(t)$ might display systematically differing stochastic behavior. This could be due to some characteristics being unobserved or due to elements inherent to the laws governing behavior. It will thus be relevant to include model parameters that are individual to each agent, thereby allowing individuals to posses different hazard levels even when conditioning on $x(t)$.

In most panel data scenarios, the depth of individual event histories will be small relative to the number of cases. Basing any model parameters on individual event histories alone will usually be infeasible. One way of mitigating this small sample effect is to introduce a mixing distribution over the parameters. Using a discrete point mass heterogeneity distribution for the intercept term has been advocated by Heckman and Singer (1984) and has been the typical choice in econometric applications. In the marketing science literature, latent class models have been popular. Such models assume data is generated by agents with unobserved group membership, each group being homogenous with respect to some behavioral parameters. This yields a discrete heterogeneity distribution with mass at sets of parameters. A different approach is advocated by Allenby and Rossi (1999). They argue for a continuous mixing distribution instead, since the tails of the true heterogeneity distribution is unlikely to be well described by a discrete approach. Often the marketeer should be more interested in picking up likely extreme behavior than average behavior. Accurately capturing tail behavior of heterogeneity is essential in for instance churn studies.

Eq. (1) is modified to include an individual level intercept term. For each customer k define hazards

$$\lambda_{ksd}(t) = e^{\alpha_{ksd} + \beta_{sd}' x(t)}, k \in \{1, \ldots, N\}\ s \in \Omega, d \in \Omega \setminus \{s\}. \tag{5}$$

Even if $\lambda_{sd}(t)$ where constant for all t, the aggregate empirical hazard would still be allowed to take on any shape depending on the form of the mixture. Think of a high and a low hazard population, each having a constant hazard. Then the aggregate empirical hazard would be downward sloping as the high-hazard individuals left the population.

Specifying how to account for unobserved heterogeneity cannot by the nature of the problem be determined from data alone. A typical choice in the litterature is to try out a normal prior.

$$? _{ksd} \; ? \; N ?0, ? ^{2}_{sd} ?$$

$$1/? ^{2}_{sd} \sim Gamma ? \frac{v}{2}, v\, s^{2} ?. \qquad (6)$$

$? _{sd}$ is attached to an inverse gaussian prior, the conjugate distribution, where the prior parameters lend themselves to a special interpretation: s^2 is the prior variance and v a degrees-of-freedom measure of the strength of the prior belief.

Putting eq. (3) and (4), modified to match eq. (5), together, the probability for individual *k* of moving from state *s* to state *d* is displayed in eq. (7).

$$p_{ksd} ?m? ? \begin{cases} ?1 ? e^{? ? _{i?s} e^{? _{ksi} ? ? ?_{si} x_m}} ? \dfrac{e^{? _{ksd} ? ? ?_{sd} x_m}}{? ? _{i?s} e^{? _{ksi} ? ? ?_{si} x_m}} & \text{for } s ? d \\ e^{? ? _{i?s} e^{? _{ksi} ? ? ?_{si} x_m}} & \text{for } s ? d \end{cases} \qquad (7)$$

Let $s_{km} ? ?$ denote the state of agent *k* in period *m*. Given the state history of agent *k*, $H_k ? ?s_{im}?_{m?1}^{M_k}$, with M_i being the number of observed intervals and *N* the number of observed agents, the likelihood of all observations is then

$$\prod_{k?1}^{N} \prod_{m?1}^{M_i ?1} p_{s_{km} s_{k,m?1}} ?m?, \qquad (8)$$

when considering the first state as given and non-stochastic.

2.4 Estimation

The main difficulty in estimating the current model stems from the unobserved heterogeneity. Using maximum likelihood methods for estimating the parameters, one would have to solve the integral of eq. (7) with respect to the heterogeneity density of eq. (6). It has no closed form solution and must be solved numerically. Another issue would be the high number of individual heterogeneity parameters $? _{i??}$ that would be part of the estimation. Many optimization routines are not geared to solve for parameters numbering in the thousands.

Using simulation methods instead comes with several advantages. The full distribution of the parameters is a byproduct of the estimation and identification issues with some parameters will not prevent proper estimates of fully identified parameters. The full distribution of predictions is likewise available, as are functions of any parameters.

The popular Metropolis-Hastings algorithm is used for the estimation of model parameters. See Chib et. al (1996) or a Robert and Casella (1999) textbook. The algorithm generates a chain of values with a distribution that converges to the true distribution of the parameters. Statistics such as moments and quantiles are then calculated as functions of the chain.

To illustrate the methodology, observe how one step of the chain is generated.

Algorithm. Let $?^{?t?}$ denote the value of all parameters at time t except $?$. Let $?^{?0?}$ and $?^{?0?}$ contain initial values. Let $L??|??$ denote a function proportional to the likelihood of $?$ conditional on $?$.

1. Draw $? \sim N?0,?^{2}_{?}?$
2. Let $w ? ?^{?t?1?} ? ?$,
3. Draw $u \sim U?0,1?$
4. If $u ? \frac{L?w|?^{?t?1?}?}{L??^{?t?1?}|?^{?t?1?}?}$ then (ACCEPT) let $?^{?t?} ? w$
5. Otherwise (REJECT) let $?^{?t?} ? ?^{?t?1?}$

Then it goes on to the next parameter of interest and the steps are repeated. Note how w is always accepted if the associated likelihood is higher than that of the previous parameter value, while it may be accepted, but with a probability decreasing in the likelihood ratio. $?^{2}_{?}$ is a tuning parameter of the algorithm that is set during initial runs to obtain an acceptance rate of around 30%.

2.5 Customer Equity and Return on Investment

Quality measures of future customer value and behavior are essential for rational resource allocation in activities such as marketing campaign targeting, individual pricing and treatment, and in corporate policy development such as determining sales force salary incentives.

If we assume away word-of-mouth effects and any accounting difficulties of attributing cost and revenue to an individual customer, one customer is clearly more desirable than another if he earns the company more. The case is less clear when contrasting a customer that generates more revenue in a short time and then ceases to transact, against a loyal, lower revenue-generating individual. It is the problem of valuing the *customer equity*, the construct named by Blattberg and Deighton (1996).

If a per period monetary value is assigned to each state in the relationship state space, the scenario is similar to that of financial asset pricing in finite state models[2]. The difference lies in the absence of a market and market prices that rules out arbitrage, so probabilities of being in each state are needed to derive valuations.

The model presented in this paper can generate predictions of the relationship state n periods ahead. Furthermore, the simulation method introduced in section 2.4 can generate the *distribution* of state positions n periods ahead, incorporating the statistical variance of parameter estimates. Let p_{kst} denote the probability of agent k inhabiting state s in period t.

Theoretically, a lifetime value of the customer could be derived. In practice however, fitting a much shorter valuation horizon ranging in months or a year is advisable in most applications. Factors such as competition, new products, new customer preferences and learning will move the assumption of stationarity of the data generating process too far away from the truth.

Let $?_{s}$ denote per-period value generated by a customer relationship occupying state s. The n period ahead cumulative expected revenue $?_{kn}$ for customer k is then

[2] Most introductory textbooks to finance, such as Luenberger, David G. (1997), Investment Science: Oxford Press., section 9.9, explain the theory and hence no further explanation is given.

$$E(?_{kn}) = \sum_{t=1}^{n} \sum_{s=1}^{S} p_{kst} ?_{s}. \tag{9}$$

To assess the risk of revenue, some relevant distribution quantiles could be sampled during simulation.

If a marketing action variable x_{kit} for customer k is included as part of the estimation, the sensitivity of $?_{kn}$ with respect to a threshold change in x_{kit} can serve as a key effect measure. For the binary case of

$$x_{kit}^{?} = \begin{cases} 0 \text{ for } t ? 1 \\ 1 \text{ for t } ? 1 \end{cases} \tag{10}$$

with time being equal to zero at the time of prediction, the sensitivity of revenue $?_{kn}(x_{kit})$ can be studied as the simulated value of

$$?_{kn}(x_{kit}) = E(?_{kn} \mid x_{kit} = x_{kit}^{?}) - E(?_{kn} \mid x_{kit} = 0 \text{ for all } t). \tag{11}$$

If a per individual cost of initiating the marketing action variable is known, say c, then the n-period-ahead return on investment of initiating the action on individual k is

$$ROI_{n} = \frac{?_{kn}(x_{kit})}{c} - 1, \tag{12}$$

when ignoring the discounting issue altogether. Quantiles of *ROI* could be calculated and only customers with a significantly positive *ROI* exposed to the marketing action.

Even when no monetary value is assigned to each state, some simple but interesting measures regarding customer behaviour and customer equity can be reported, such as expected remaining duration in a given state and *n* periods ahead probability of deterioration or improvement.

3 Application

In the following section we shall study the key dynamics in a customer database from a telecommunications service provider. There are multiple products and the products can be consumed at varying intensities.

The service provider offers, in order of increasing technical sophistication and price, ordinary public switched telephone networking (PSTN) for analogue voice and data communications, integrated service digital networking (ISDN) for faster internet access, and asymmetric digital subscriber lines (ADSL) for broadband internet access. All service plans have a fixed monthly cost and a usage-dependent cost. For PSTN and ISDN, voice and data traffic are billed on a per-minute basis. For ADSL internet usage is free. Company policy requires customers to include a landline telephone in their subscription, so there is still a variable component in the bill from voice calls. It can be said that ADSL and ISDN subscriptions encompasses PSTN subscriptions.

Attention to customer behavior is divided into three distinct behavioral elements having an impact on revenue: *Churn*, *product choice* and *consumption level*. In the study we shall examine dynamics within and across these elements, conditioning on sociodemographic variables, time and individual heterogeneity. A few special questions shall also be considered, namely patterns in data related to customer churn, and patterns associated with customers choosing ADSL products.

3.1 Data

3.1.1 Event histories

9218 subscribers are flow sampled[3] by narrowing interest to customers *activated* in the period from January 2001 to April 2003. Product type and consumption levels are recorded on a monthly basis, so ? equals one month. No individual is observed past April 2003, even when he is still with the company.

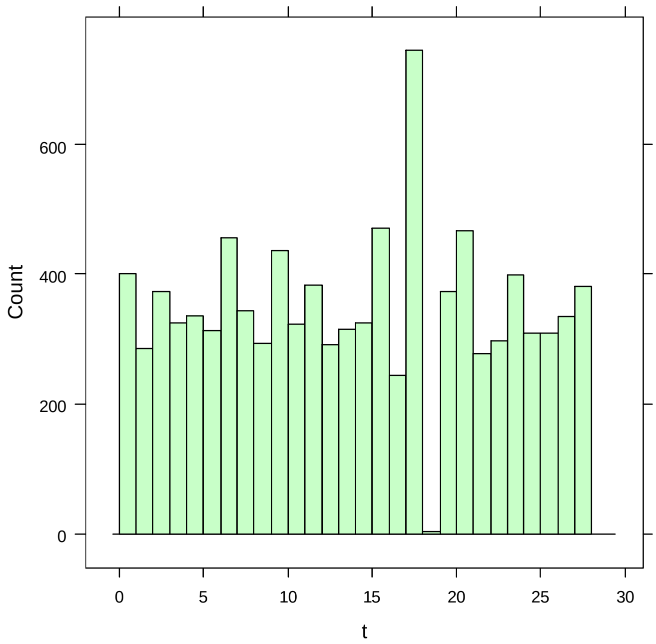


figure 3.1. Histogram of event history durations

---

[3] See the end of section 2.2

The median event history length is 15 months. Examining figure 3.1 event histories of length 18 and 19 months are of special concern, since both are extreme with counts of 744 and 3, respectively. Experts from the marketing intelligence department think it represents an unusual technical glitch: Customers arriving at t-19 months were not registered until t-18 months ago. The author could model this possibility explicitly, but choose to ignore it to avoid further complicating the setup.

Although it is outside the scope of this paper to model arrival rates, it is noted that these do not seem like draws from a poisson distribution. This will not influence the centrality of parameter estimates, but may give rise to varying variance on estimates at different time depths.

3.1.2 Sociodemographics

Sociodemographic variables from a commercial database are included in the analysis. Variables are measured on an aggregate level. The database partitions the geographic region into cells of varying size, depending on the legislative requirements. Some measures of income, for instance, require at least 150 households as a calculation base, while age measures are considered much less sensitive and may be based on a little as 20 households. This scheme undoubtly induces an additional level of noise in the data, but few companies have legal access to more detailed types of sociodemographics on a large scale.

Sociodemographic variables are grouped into five sets: Age, income, jobmarket status, education and houshold type. Variables in all groups, except income, are measured in proportions. Descriptive statistics can be seen in figure 5.1 (p. 58). Within these groups, proportions sum to one, so at least one variable must be excluded for analysis to commence. *Adult*, *jhi*, *enone* and *hs0* are excluded on the grounds of being highly correlated to variables *child* (-0.83), *elong* (0.80), *enone* (-0.80) and *hc2p* (-0.87) respectively.

As seen in figure 5.2 (p. 59), the variables are in general correlated. A high variance on individual parameter estimates must be expected. A general-to-specific variable elimination scheme will be utilized later on, but given the correlation between demographic variables, a cautious approach must be taken when interpreting variable coefficients.

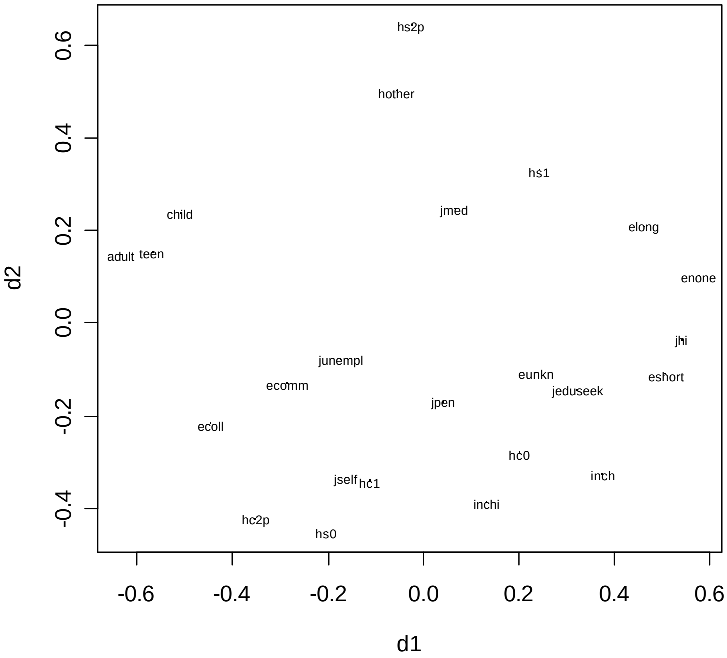


figure 3.2. Correlation structure of sociodemographic variables. Plot of two-component multidimensional scaling of absolute correlations.

3.2 Model

Each of the modeled behavioral elements - *churn*, *product choice* and *consumption level* - are characterized by a distinct discrete state. Churn is binary, so either you stay as a customer or you quit all engagements with the company. *Product choice* is limited to the three major products PSTN, ISDN and ADSL. The *consumption level* is measured as a grouped dollar value of voice and data access, including local and international calls, excluding any fixed plan-based fees.

Product choice and *consumption level* in period *t* are thought to influence each other and churn in period *t+1*. *Churn* is influencing *product choice* and *consumption level* through censoring alone. *Product choice* and *consumption level* are undefined for churned customers.

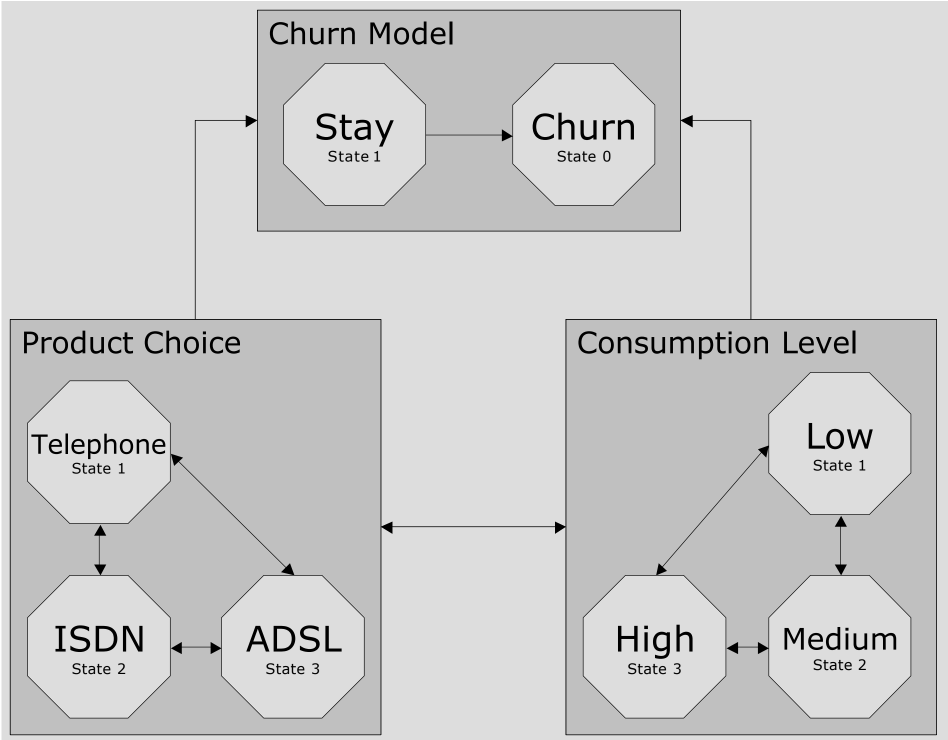


figure 3.3. Relation State Space

The model formulated in terms of hazards is seen in eq. (13). When comparing to eq. (5) an extra index representing the submodel in question appears on the variables.

$$
\begin{aligned}
&?_{kij0}\,?t?\;?\;\exp(?_{kij0}\;?\;?'_{ij0}x_{k[t]?}),\;\; ?_0\;?\;?0,1?\\
&?_{kij1}\,?t?\;?\;\exp(?_{kij1}\;?\;?'_{ij1}x_{k[t]?}),\;\; ?_1\;?\;?0,1,2,3?\;\;.\\
&?_{kij2}\,?t?\;?\;\exp(?_{kij2}\;?\;?'_{ij2}x_{k[t]?}),\;\; ?_2\;?\;?0,1,2,3?
\end{aligned}
\qquad (13)
$$

The submodels are labeled as {(0,Churn), (1,Product Choice), (2,Consumption Level)}. For the churn model the state space is labeled as {(0,churn), (1,stay)}. For the product choice model it is labeled as {(0,undefined), (1,PSTN), (2, ISDN), (3, ADSL)}, while the consumption level state space model is labeled as {(0,undefined), (1, low), (2, medium), (3, high)}.

To churn, by unsubscribing all products, and later returning as a customer is a rare event as seen in section 3.2.3. This has to do with the time horizon and the difficulty of tracking former customers. Even though the company is putting an effort into the area, it will continue to depend on inherently error-prone processes such as name matching. The non-customer state is therefore considered an absorbing state, a state from which a customer does not return.

In the product choice and consumption level models, transitions to the undefined state are treated as a censoring event.

Transition from ADSL to ISDN is so rare that it is excluded from the model. Such events are also treated as censoring events.

Each agent gives rise to twelve unobserved heterogeneity terms $?_{kijl}$, l being the submodel number. One term in the churn model, five in the product choice model and six in the consumption level model. The unobserved heterogeneity is estimated according to eq.(6) with prior parameters $s_1^2 \; ? \; 0.1$ and $v_1 \; ? \; 1$.

3.2.1 Grouping of Variable Consumption

The consumption level is continous by nature, but to simplify matters and study the effects of low and high consumption more clearly, it is discretized. Log(consumption+1) is discretized into three groups. These groups are chosen rather arbitrarly to include approximately 30%, 50% and 20% of the population, respectively. Variable revenue of less than 8% of the mean is denoted *low*, while revenue between 8% and 240% of the mean is denoted *medium*. Consumption higher than 240% of the mean is denoted *high*.

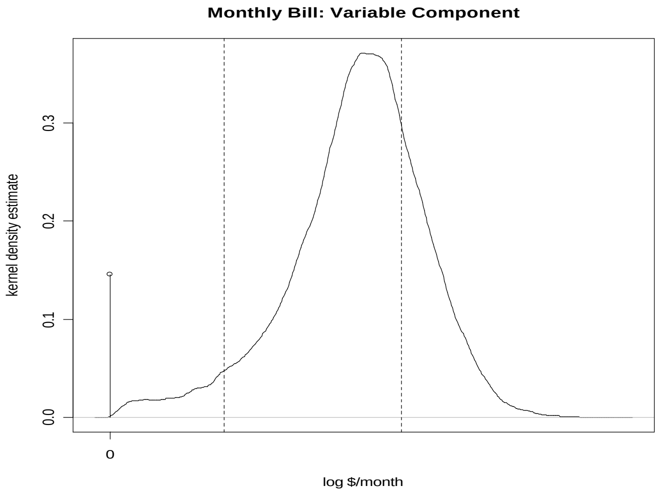


figure 3.4. Empirical density of monthly log(variable bill+1) for active customers. Around 15% of customers have a zero-level consumption as indicated by the vertical line. Actual bill levels are secret, hence the blanking of the x-axis.

3.2.2 Duration Dependence

Transitions in and out of states might relate to duration of stay. Discrete hazard curves are shown in figure 5.4 (p. 61). Several time-dummies are included to account for this possibility, permitting the level of the hazard to vary across time. The discrete hazard curve of each submodel is loosely examined for kinks and time-dummies are included according to the table below.

Submodel	Timepoints
Churn	1,4,12
Product Choice	1,3,12,16
Consumption Level	1,2,3,5,7

The dummies are cumulative, in the sense that a time dummy setup at time t_1 is set as in eq. (14).

$$s_{t_1,t} ? \begin{cases} 1 \text{ for } & t ? t_1 \\ 0 \text{ for } & t ? t_1 \end{cases} \qquad (14)$$

3.2.3 Transitions

The customer agent can move from one state to the next each month. Statistics on raw transition frequencies are shown in

figure 3.5.

Churn Model

Transition Frequency

	Destination	
Source	Churn	Stay
Churn	8382	43
Stay	924	55433

Markov Transition Probabilities

	Destination	
Source	Churn	Stay
Churn	0.9949	0.0051
Stay	0.0164	0.9836

Product Choice

Transition Frequency

	Destination		
Source	PSTN	ISDN	ADSL
PSTN	41743	57	259
ISDN	33	5070	106
ADSL	66	15	8084

Markov Transition Probabilities

	Destination		
Source	PSTN	ISDN	ADSL
PSTN	0.9925	0.0014	0.0062
ISDN	0.0063	0.9733	0.0203
ADSL	0.0081	0.0018	0.9901

Consumption Level

Transition Frequency

	Destination		
Source	LO	MED	HI
LO	4819	1231	239
MED	1443	31347	3020
HI	213	3146	9975

Markov Transition Probabilities

	Destination		
Source	LO	MED	HI
LO	0.7663	0.1957	0.0380
MED	0.0403	0.8754	0.0843
HI	0.0160	0.2359	0.7481

figure 3.5. Monthly submodel transition frequency and markov probabilities.

It is clear from

figure 3.5 that product choice transitions are rare, while consumption level changes are a common phenomena. This might lead to more variable estimates for the product choice model.

3.2.4 Submodel Interaction Variables

The submodels influence each other through a dummy-variable mechanism. The *product choice* state (PC) is allowed to influence the *churn* state and the *consumption level* state. The variable *internet* is included in the *churn model* according to eq. (15).

$$internet_t = \begin{cases} 0 \text{ if} & PC_{t-1} = PSTN \\ 1 \text{ if} & PC_{t-1} \in \{ISDN, ADSL\} \end{cases} \qquad (15)$$

In a similar way the variable *isdn* and *adsl* is included in the *consumption level model*, being equal to one if PC entered state ISDN and ADSL, respectively.

Customers once subscribing to ISDN or ADSL and then rejecting either might be more or less prone to subscribe again. The dummy variable *inettch* ("internet touch") is equal to one, if the customer doesn't currently subscribe to ISDN or ADSL, but once did. The variable is included in all three models.

The *consumption level* state influences the *churn* state and *product choice* state through variables *pstnlo* and *pstnhi*. If *consumption level* is in state *low* then *pstnlo* equals one, while a state of *high* results in *pstnhi* set to one. Both are set to zero otherwise.

3.3 Other Studies

With regards to sociodemographics, it is useful to form some expectations of patterns in behaviour. Anderson et al. (1999) study patterns of sociodemographics and consumption of communication services, so their findings can serve as a benchmark. Some relevant main conclusions are loosely summarized in table 3-1.

Dimension	Regular Phone Use Higher for…	Computer Use Higher for…
Social class / income	Higher social class	Higher social classes, as they are more likely to own a computer. Conditional on ownership of a computer, lower social class correlates with more time spend using a computer, except for the bottom 20%.
Age	Young and old	Younger people. Conditional on ownership, older people used computers more overall, especially for work related projects.
Family Structure	Weakly higher use for couples with children	Couples with children, young singles without children. Least likely for singles with children.

table 3-1. Overview of patterns in consumption as presented in Anderson et al. (1999).

3.4 Results

Estimation. 20.000 iterations were spent calibrating the control parameters of the Metropolis-Hastings algorithm, another 20.000 on burn-in. One variable in each state in each submodel is eliminated using a general-to-specific principle. Then a new cycle of elimination started using 2000 iterations for tune-in and 2000 for parameter estimates. The process was repeated until no insignificant variables remained at the 10% level.

Model fit. How does one judge the predictive qualitiy of a model? One measure is that of model lift, where the model's predictive ability is compared to that of a completely random prediction. Given a sample of size *n* binary customer events $y_i \in \{stay, churn\}$, related characteristics x_i and a churn score function $f(x)$, we wish to calculate decision lift. If the sample contains n_{churn} churners, we extract n_{churn} customers according to a ranking generated by $f(x)$. Let $n_{correct}$ designate the number of correct predictions. The lift is the ratio of $n_{correct}/n_{churn}$ to n_{churn}/n.

Lift measures for the model's ability to predict churn and predict if a customer buys ADSL are reported in section 3.4.1 and 3.4.2.

Parameter estimates for the reduced model are reported in figure 6.1 (p. 63). Statistics on the individual level heterogeneity parameters are not reported, as they are too numerous, but their empirical distribution are shown in figure 6.2 (p. 64). The unobserved heterogeneity parameter for the churn model is nearly rejected on a 1% level, questioning the choice of prior for this particular parameter. The empirical distribution seems left-skewed and this is confirmed by a D'Agostino test for skewness, so a better prior should accommodate this. Parametric alternatives could for instance be a skew-normal distribution.

Interpretation. It is challenging to interpret coefficient estimates in this complex model. One-period ahead transition sensitivities to variables can be understood directly from coefficient signs, but multiple periods complicate matters by allowing variables to have a direct and indirect effect on the outcome of interest. A further complication is variables exhibiting some degree of multicolinearity, unobserved heterogeneity and states pushing customers into new states with varying degrees of inherent force.

In the following sections the outcome of interest is the probability p_i of a submodel state variable equaling s^* at least once within n periods of time. This is investigated by simulating the state path n periods ahead for each agent, observing if s occurs. Model parameters are evolved simultaneously using the Metropolis-Hastings algorithm, so statistics based on the simulation is not conditional on average model parameter estimates, but on their full distribution.

Let $\tilde{s}_{i,k,t}$ denote the simulated state in period t for customer i in simulation run numer k. Given K simulation runs, the estimated probability $\hat{p}_i$ of customer i touching state s^* is given in eq. (16).

$$\hat{p}_i = \frac{\#\{k \mid \exists j \in \{1,\ldots,n\} : \tilde{s}_{i,t+j,k} = s^*\}}{K} \tag{16}$$

Eq. (16) is the basis for lift calculations and for sensitivity analysis. Sensitivity analysis, with respect to sociodemographic variables, is carried out by estimating the average p_i conditional on one variable x_j at a time. A local regression method, *locfit* in the statistical system R, does that for us. To get a rough idea of the sensitivities, the conditional average is tagged at the first, second and third quantile of each variable x_j.

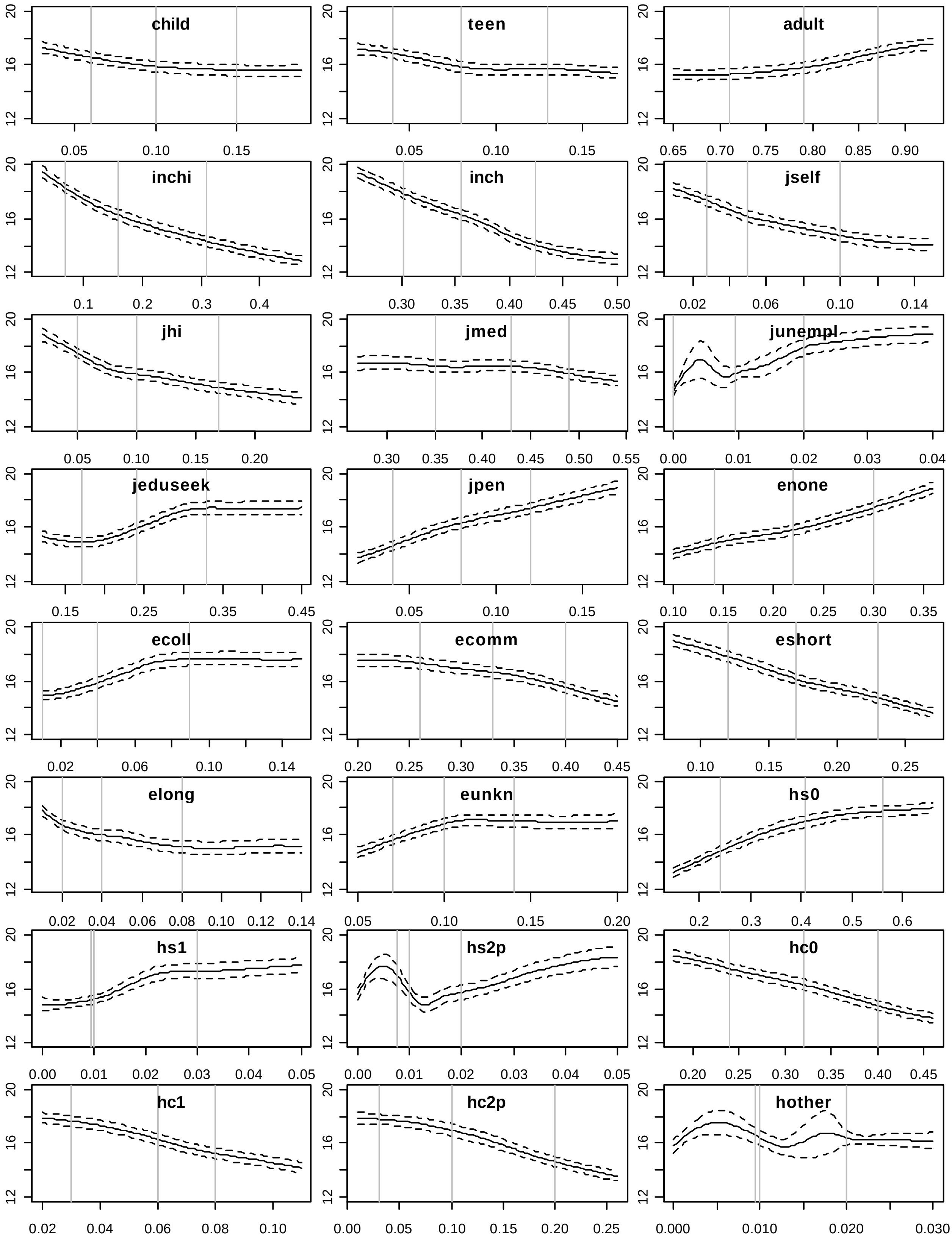


figure 3.6. Percentage probability of churn in twelve months conditional on a sociodemographic variable. Black line: Average. Dashed lines: 95% confidence band. Grey vertical lines: First quantile, median and third quantile.

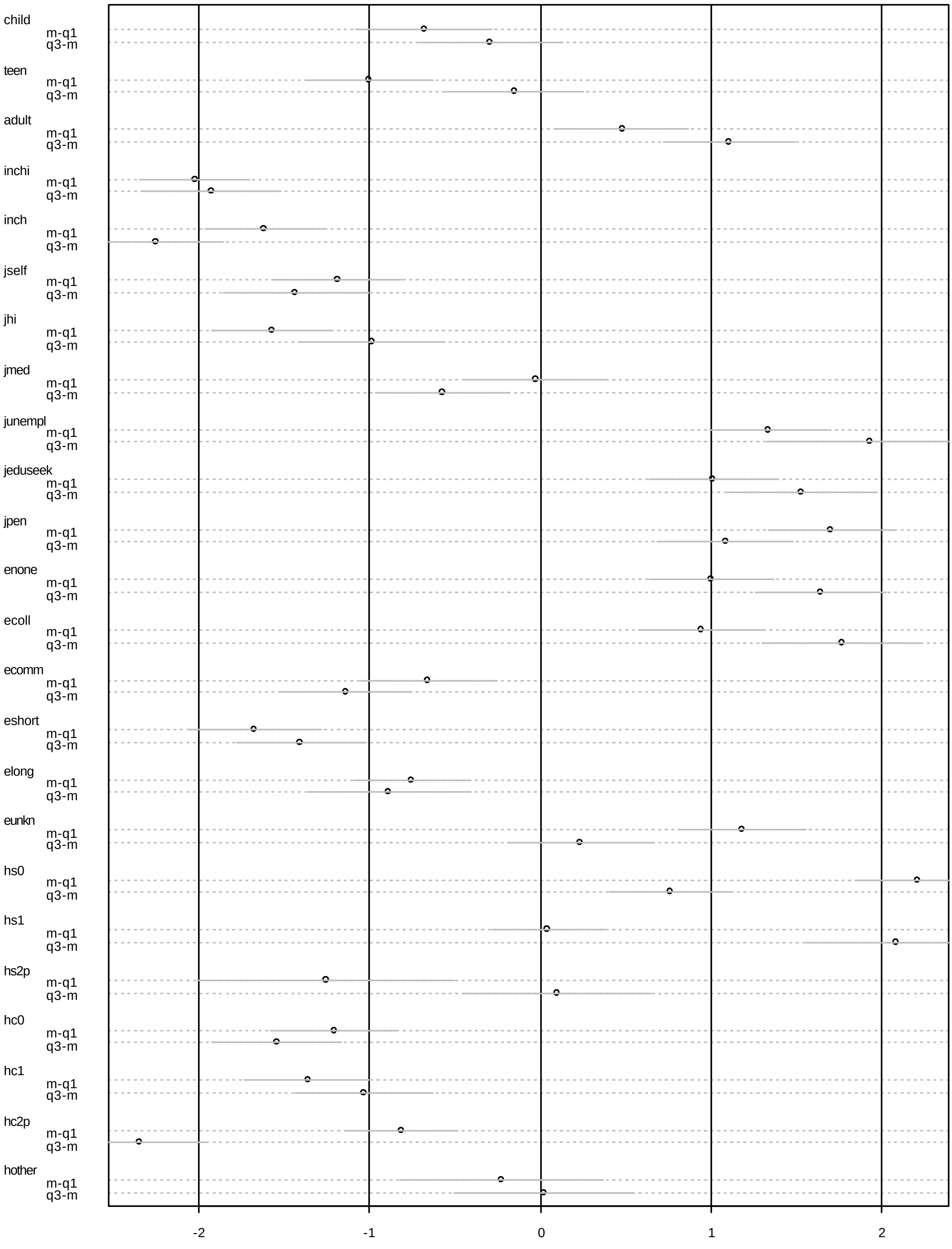


figure 3.7. Churn sensitivity. Predicted 12 months ahead churn percentage sensitivity to sociodemographic variables. “m-q1”: change in average churn from first quantile to median of variable. “q3-m”: change in average churn from median to third quantile of variable. Point: Average change. Grey line: 95% confidence band.

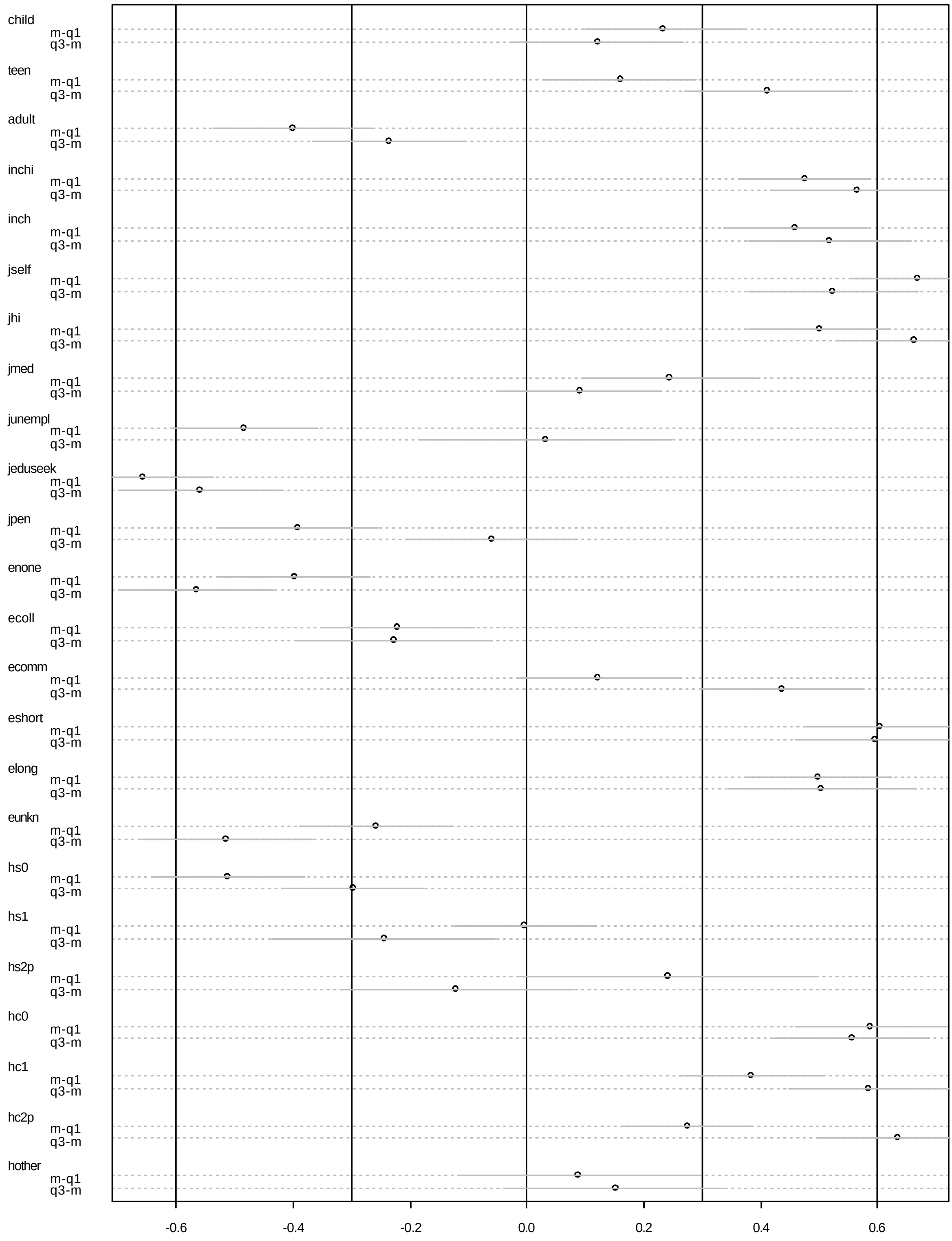


figure 3.8. Transitions from PSTN to ADSL. Sensitivity. Predicted 12 months-ahead probability of transition from PSTN to ADSL sensitivity to sociodemographic variables. "m-q1": change in average churn from first quantile to median of variable. "q3-m": change in average churn from median to third quantile of variable. Point: Average change. Grey line: 95% confidence band.

3.4.1 Exit Behavior

Predictions are out of sample in more than one sense. Firstly, the predicted sample did not contribute to the likelihood of any aggregate parameter values. Secondly, individual heterogeneity parameter estimates are based on a truncated version of the sample. The agent event history is modified, for an *n* period ahead prediction, by excluding the remaining *n* observations from estimation. Individuals with a transaction history length below size *n* are excluded from comparisons.

horizon	total	actual	predicted	ratio correct	ratio random	lift
1	3602	92	15	0.1630	0.0255	6.4
2	3494	144	25	0.1736	0.0412	4.2
3	3402	197	35	0.1777	0.0579	3.1
6	2990	293	52	0.1775	0.0980	1.8
12	2233	424	143	0.3373	0.1899	1.8

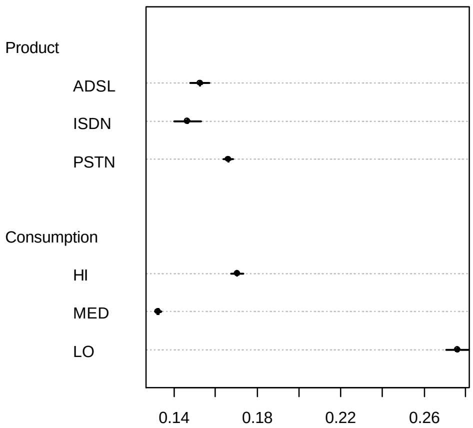
figure 3.9. Lift for out-of-sample predicted churn.

In general, lower risk of churn is associated with areas of higher income, higher proportions of self-employed, high-ranking job positions, educated people living in households as couples or as singles with two or more children.

Higher risk of churn is associated with areas of lower income, higher proportions of people with no job market affiliation, little or no formal education, and living in single households.

Among the sociodemographic groups, age composition seems to carry little significance.

Two effects are highly non-linear. The proportion of single households is associated with higher risk for proportions below the median of 0.42. For values above the median, the association is reversed, though less steep. A similar relation holds for couple households with two or more children.



	Mean	sd
Consumption Level		
ADSL	0.153	0.002
ISDN	0.147	0.003
PSTN	0.166	0.001
Product Choice		
LO	0.276	0.003
MED	0.133	0.001
HI	0.170	0.001

figure 3.10. Predicted 12 months ahead churn conditional on submodel state variables. Point: Proportion estimate. Grey line: 95% confidence bands.

Customers with low and high telephone consumption are more likely to churn than medium level consumers, low state consumers being much more likely to churn.

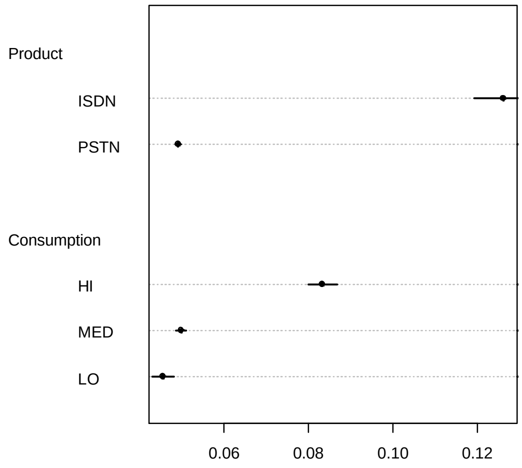
Product choice does not seem to induce significant differences in churn, even though possession of ADSL or ISDN lowers risk directly: In figure 6.1 p. 63 the *internet10* coefficient is significant and negative. It is seemingly paradoxical that an effect is directly significant, but results in a prediction that says otherwise. Studying the consumption level submodel, it turns out that ISDN and ADSL plays an important role. Leaving the *low* state is significantly less likely if an internet product is present. Entering state *low* from state *medium* is significantly more likely in the presence of an internet product. Entering state *low* from state *high* is more likely for ADSL customers. Put differently, assume we are in period *t*. A customer with consumption level in state *medium* and product choice in state ISDN/ADSL is less likely to churn than other customers in period *t+1*. He is, however, more likely to move his consumption level to state *low* or *high*, thereby increasing the risk of churn in period *t+2*. Considering that state *low* is much riskier than state *medium* and *high*, the indirect higher risk absorbs the direct effect.

3.4.2 Who Buys ADSL?

horizon	total	actual	predicted	ratio correct	ratio random	lift
1	2602	11	0	0.0000	0.0042	0.0
2	2541	26	0	0.0000	0.0102	0.0
3	2513	55	2	0.0364	0.0219	1.7
6	2273	86	10	0.1163	0.0378	3.1
12	1788	125	18	0.1440	0.0699	2.1

figure 3.11. Lift for out-of-sample predicted PSTN to ADSL transitions.

Few sociodemographic characteristics have a direct significant impact on the probability of transitions to the ADSL state. Average houshold income, *inch*, and the proportion of teenagers in a given area, *teen*, are the only two directly significant sociodemographic variables as seen in figure 6.1 (p. 63). For ISDN customers, only income is directly significant. However, due to the high correlation in sociodemographic variables, there is an observed significance in several variables correlated with income and teenagers. Conditional on PSTN customers, it is clear from figure 3.8 (p. 52) that higher proportions of children, self-employed, couple households and unclassified households are positively related to PSTN to ADSL transitions. Higher proportions of unemployed, people seeking education, uneducated and single housholds without children are negatively related.



	Mean	sd
Product		
ISDN	0.126	0.004
PSTN	0.049	0.000
Consumption Level		
HI	0.083	0.002
MED	0.050	0.001
LO	0.045	0.001

figure 3.12. Predicted 12 months ahead probability of an ADSL transition conditional on submodel state variables. Point: Proportion estimate. Grey line: 95% confidence bands. "Con.lvl": Consumption Level.

PSTN customers with high consumption are more likely to get ADSL, while medium and low consumption display little difference. ISDN customers are much more likely to get ADSL.

No time-dummies survive variable elimination, so duration of stay seems to carry no information.

3.4.3 Importance of Individual Unobserved Heterogeneity

A second run of the model is done excluding individual unobserved heterogeneity, to see if the complication adds anything to the predictive ability of the model.

Horizon	Lift 1	Lift 2	Ratio.
1	6.4	4.9	0.76
2	4.2	3.4	0.81
3	3.1	2.8	0.91
6	1.8	1.9	1.03
12	1.8	1.8	0.99

figure 3.13. Lift on 12 months ahead churn predictions. "Lift 1": Full model. "Lift 2": A model excluding individual unobserved heterogeneity.

It is clear that individual unobserved heterogeneity improves the lift of the model, but the relative gain decreases as the forecast horizon grows. The gain disappears when moving beyond a forecast horizon of three months.

The method used to compare out of sample lift measures is bound to disfavor a model including individual unobserved heterogeneity. As the forecast horizon grows, the individual event history shrinks in line with the out of sample principle. The unobserved heterogeneity estimates being very dependent on individual event histories will thus gradually loose valuable information. The extent to which this distortion influences the comparison is unknown, but we can rest assured that the method makes it unlikely to overestimate the value of unobserved heterogeneity as a modelling element.

3.4.4 Customer Value

We will now calculate the present value of customer revenue streams five years into the future, applying a 10% per year discount rate. Each state is assigned a value, not exactly equal to the true value due to publication restrictions, but enough for the big picture to stay the same.

	$/month
Product	
PSTN	20
ISDN	30
ADSL	40
Consumption	
LO	0
MED	10
HIGH	60

figure 3.14. Revenue per month derived from customer relation states.

When studying figure 3.15 it can be concluded that the pattern of relations is similar to that of churning customers. Well-off, better-educated customers in couple households are more valuable, as these customers also display lower churn levels. However, many variables associated with significant changes in churn are not significant in association with the 5-year value prediction. Areas with above-average unemployment carry higher churn risk, but not a significantly lower value. Similarly, groups with above-average rates of single households with one child usually go with higher churn rates, but not significantly lower customer value.

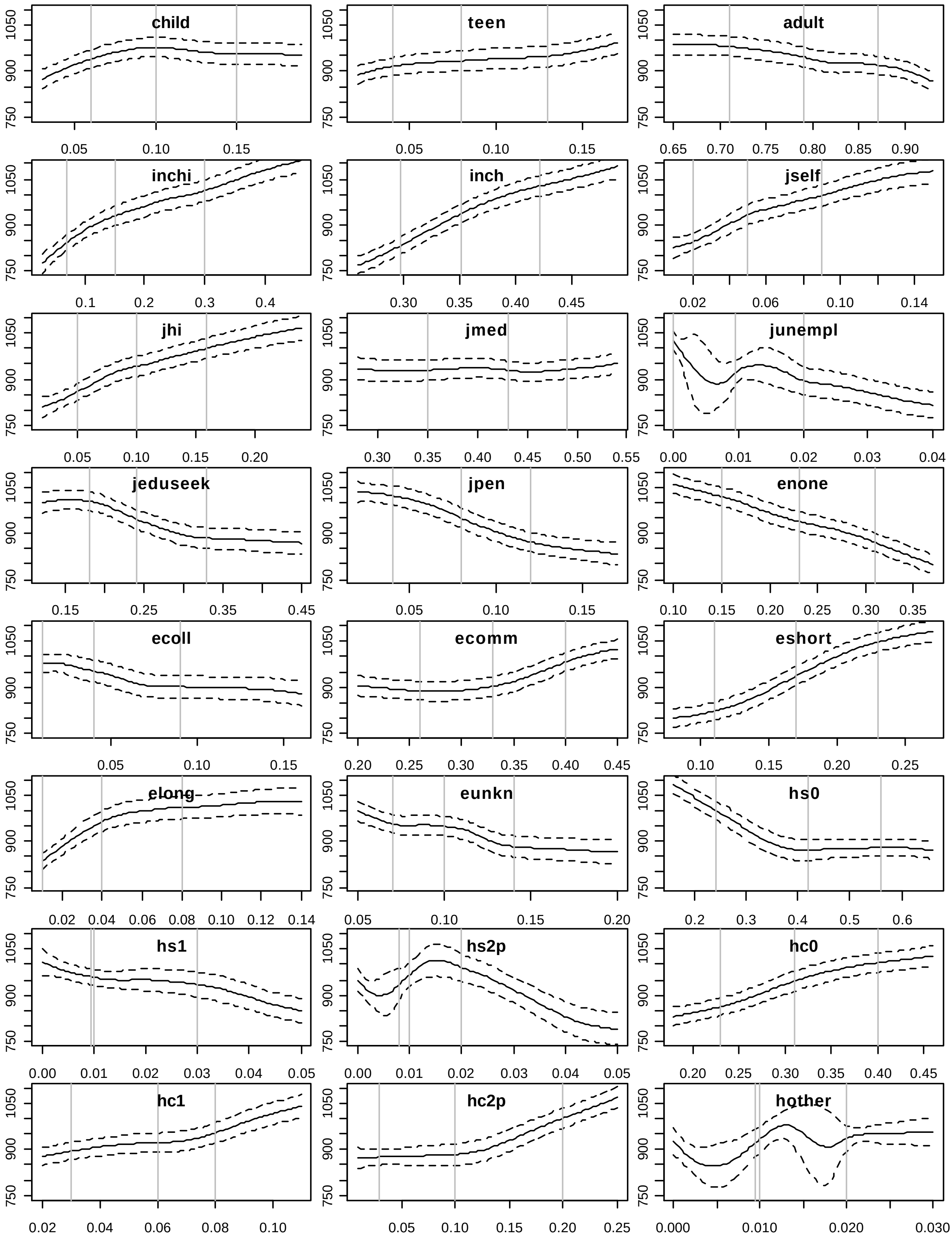


figure 3.15. Forecasted 5 years ahead average present value of customers, discounted at 10% per year.

4 Conclusion

A comprehensive model of customer value is a natural ingredient in the analysis of a very large range of management problems. Accounting and budgeting can profit from revenue forecasts, financial analysts and risk managers have an interest in customer base pricing. Marketing resource allocation decisions should also to a higher degree be linked to expected changes in customer value.

Imagine a simplified budget decision scenario regarding a marketing campaign designed to win new customers. The revenue $?$ from a given campaign is a function of response-rate per person *r*, number of potential customer touched *n*, per-person cost of the campaign *c* and customer present value *v*.

$$? \; ? \; rnv \, ? \, cn \tag{17}$$

v is a stochastic variable, while *c* and *n* can be treated as constants. We might have no knowledge of *r*, but we can at least find a lower bound that must be met.

$$\begin{aligned} E(?) \quad &? \quad 0 \\ r \; ? \; &\frac{c}{E?v?} \end{aligned} \tag{18}$$

Using numbers from figure 3.15 we could even derive different bounds for different sociodemographic groups. Using only household income *inch* we exchange the unconditional expectation $E(v)$ for a conditional $E(v \mid inch)$. The expected 5-year present value of a region earning $30,000/year, the first quantile, is $825, while an area earning around $42,000/year, the third quantile, has a present value of $1050. The minimum *r* for the former region is then 6.1%, and 4,8% for the latter. A campaign expected to yield a response rate between these two levels should obviously exclude the former group from the campaign.

4.1.1 Future Research

As an alternative to discretizing the continous consumption variable, one could extend the statespace and introduce a model of variable consumption, letting relevant parameters depend on state. Also, a model of inflow of new customers would be a useful addition for a company to get a more complete picture of effects of marketing actions.

The presence of large datasets and the lack of knowledge about functional form of the transition probability and observable heterogeneity-relation, makes nonparametric approaches attractive. Few papers exist dealing with nonparametric models and state space representations.

Given observed marketing action variables, such as records about customers exposed to various marketing treatments, an obvious next step for this statespace representation is to test richer consumer behavior and marketing theory.

5 Appendix: Descriptive Statistics

Title	Description	Min	Q1	Median	Q3	Max
Age Groups						
child	Proportion aged <10	0.00	0.06	0.10	0.15	0.44
teen	Proportion aged 10-19	0.00	0.04	0.08	0.13	0.90
adult	Proportion aged >=20	0.09	0.71	0.79	0.87	1.00
Income						
inchi	Proportion of households earning >$100k/yr	0.00	0.07	0.16	0.31	0.86
inch	Average household income, normed $100k	0.08	0.30	0.35	0.42	1.29
Jobmarket Status						
jself	Proportion self-employed	0.00	0.02	0.05	0.10	0.46
jhi	Proportion managers and workers on the highest level	0.00	0.05	0.10	0.16	0.62
jmed	Proportion being workers on average pay and lower	0.00	0.35	0.42	0.49	0.73
junempl	Proportion unemployed	0.00	0.00	0.01	0.02	0.46
jeduseek	Proportion currently seeking an education	0.00	0.17	0.24	0.33	0.99
jpen	Proportion outside the jobmarket	0.00	0.04	0.08	0.13	0.57
Education						
enone	Proportion with no formal education	0.01	0.14	0.23	0.31	0.57
ecoll	Proportion with the equivalent of a college degree	0.00	0.01	0.04	0.09	0.64
ecomm	Proportion with a commercial background	0.02	0.26	0.33	0.40	0.65
eshort	Proportion with a short to medium education	0.01	0.12	0.17	0.23	0.68
elong	Proportion with the equivalent of a university degree	0.00	0.01	0.04	0.08	0.60
eunkn	Proportion with an unknown educational background	0.01	0.07	0.10	0.14	0.76
Household Type						
hs0	Proportion of singles with no children	0.01	0.24	0.42	0.56	0.98
hs1	Proportion of singles with one child	0.00	0.01	0.01	0.03	0.20
hs2p	Proportion of singles with two children or more	0.00	0.01	0.01	0.02	0.25
hc0	Proportion of couples with no children	0.01	0.24	0.31	0.40	0.65
hc1	Proportion of couples with one child	0.00	0.03	0.06	0.08	0.23
hc2p	Proportion of couples with two children or more	0.00	0.04	0.10	0.20	0.51
hother	Proportion of unclassified households	0.00	0.01	0.01	0.02	0.16

figure 5.1. Overview of sociodemographic variables. Min: Minimum value. Q1: 25% percentile. Q3: 75% percentile. Max: Maximum value. Names in *italics* indicate exclusion from estimation due to perfect multicolinearity.

	child	teen	adult	inchi	inch	jself	jhi	jmed	junempl	jeduseek	jpen	enone
child	1.00	0.36	-0.83	0.29	0.20	0.23	0.10	0.10	-0.12	-0.20	-0.01	0.00
teen	0.36	1.00	-0.82	0.28	0.16	0.27	-0.02	0.11	-0.07	-0.10	-0.12	0.12
adult	-0.83	-0.82	1.00	-0.35	-0.22	-0.30	-0.05	-0.13	0.12	0.19	0.08	-0.07
inchi	0.29	0.28	-0.35	1.00	0.87	0.53	0.66	0.07	-0.30	-0.37	-0.55	-0.58
inch	0.20	0.16	-0.22	0.87	1.00	0.44	0.66	-0.02	-0.30	-0.32	-0.44	-0.57
jself	0.23	0.27	-0.30	0.53	0.44	1.00	0.24	-0.01	-0.24	-0.31	-0.41	-0.22
Jhi	0.10	-0.02	-0.05	0.66	0.66	0.24	1.00	-0.10	-0.08	-0.50	-0.30	-0.74
jmed	0.10	0.11	-0.13	0.07	-0.02	-0.01	-0.10	1.00	-0.03	-0.62	-0.13	-0.08
junempl	-0.12	-0.07	0.12	-0.30	-0.30	-0.24	-0.08	-0.03	1.00	-0.21	0.40	0.01
jeduseek	-0.20	-0.10	0.19	-0.37	-0.32	-0.31	-0.50	-0.62	-0.21	1.00	-0.09	0.47
jpen	-0.01	-0.12	0.08	-0.55	-0.44	-0.41	-0.30	-0.13	0.40	-0.09	1.00	0.31
enone	0.00	0.12	-0.07	-0.58	-0.57	-0.22	-0.74	-0.08	0.01	0.47	0.31	1.00
ecoll	-0.30	-0.39	0.41	-0.37	-0.25	-0.33	0.08	0.06	0.57	-0.29	0.45	-0.23
ecomm	0.18	0.30	-0.29	0.34	0.17	0.30	-0.14	0.49	-0.35	-0.11	-0.47	-0.03
eshort	0.10	0.01	-0.07	0.60	0.54	0.30	0.70	0.14	-0.08	-0.50	-0.34	-0.80
elong	0.05	-0.10	0.03	0.48	0.58	0.16	0.80	-0.27	-0.01	-0.32	-0.09	-0.67
eunkn	-0.12	-0.13	0.15	-0.35	-0.28	-0.24	-0.31	-0.51	0.01	0.58	0.19	0.27
hs0	-0.49	-0.48	0.59	-0.74	-0.51	-0.61	-0.39	-0.29	0.30	0.44	0.46	0.32
hs1	0.19	0.09	-0.17	-0.36	-0.28	-0.34	-0.27	0.00	0.08	0.08	0.43	0.33
hs2p	0.36	0.32	-0.41	-0.23	-0.22	-0.21	-0.26	0.01	0.08	0.04	0.38	0.33
hc0	0.10	0.15	-0.15	0.65	0.46	0.56	0.43	0.26	-0.26	-0.39	-0.56	-0.44
hc1	0.42	0.30	-0.44	0.58	0.43	0.40	0.40	0.23	-0.21	-0.41	-0.30	-0.33
hc2p	0.54	0.54	-0.66	0.69	0.46	0.58	0.31	0.19	-0.26	-0.32	-0.41	-0.23
hother	0.17	0.23	-0.24	0.07	0.06	0.12	-0.04	0.03	-0.02	-0.06	0.04	0.07

	ecoll	ecomm	eshort	elong	eunkn	hs0	hs1	hs2p	hc0	hc1	hc2p	hother
child	-0.30	0.18	0.10	0.05	-0.12	-0.49	0.19	0.36	0.10	0.42	0.54	0.17
teen	-0.39	0.30	0.01	-0.10	-0.13	-0.48	0.09	0.32	0.15	0.30	0.54	0.23
adult	0.41	-0.29	-0.07	0.03	0.15	0.59	-0.17	-0.41	-0.15	-0.44	-0.66	-0.24
inchi	-0.37	0.34	0.60	0.48	-0.35	-0.74	-0.36	-0.23	0.65	0.58	0.69	0.07
inch	-0.25	0.17	0.54	0.58	-0.28	-0.51	-0.28	-0.22	0.46	0.43	0.46	0.06
jself	-0.33	0.30	0.30	0.16	-0.24	-0.61	-0.34	-0.21	0.56	0.40	0.58	0.12
jhi	0.08	-0.14	0.70	0.80	-0.31	-0.39	-0.27	-0.26	0.43	0.40	0.31	-0.04
jmed	0.06	0.49	0.14	-0.27	-0.51	-0.29	0.00	0.01	0.26	0.23	0.19	0.03
junempl	0.57	-0.35	-0.08	-0.01	0.01	0.30	0.08	0.08	-0.26	-0.21	-0.26	-0.02
jeduseek	-0.29	-0.11	-0.50	-0.32	0.58	0.44	0.08	0.04	-0.39	-0.41	-0.32	-0.06
jpen	0.45	-0.47	-0.34	-0.09	0.19	0.46	0.43	0.38	-0.56	-0.30	-0.41	0.04
enone	-0.23	-0.03	-0.80	-0.67	0.27	0.32	0.33	0.33	-0.44	-0.33	-0.23	0.07
ecoll	1.00	-0.55	0.06	0.16	-0.09	0.49	0.05	-0.08	-0.28	-0.29	-0.52	-0.16
ecomm	-0.55	1.00	0.04	-0.39	-0.46	-0.53	-0.15	-0.06	0.47	0.32	0.46	0.08
eshort	0.06	0.04	1.00	0.53	-0.49	-0.44	-0.28	-0.27	0.50	0.39	0.34	-0.04
elong	0.16	-0.39	0.53	1.00	-0.16	-0.17	-0.16	-0.17	0.20	0.23	0.13	-0.02
eunkn	-0.09	-0.46	-0.49	-0.16	1.00	0.42	0.11	0.13	-0.46	-0.30	-0.27	-0.01
hs0	0.49	-0.53	-0.44	-0.17	0.42	1.00	0.15	-0.06	-0.77	-0.70	-0.87	-0.22
hs1	0.05	-0.15	-0.28	-0.16	0.11	0.15	1.00	0.55	-0.45	-0.10	-0.21	0.14
hs2p	-0.08	-0.06	-0.27	-0.17	0.13	-0.06	0.55	1.00	-0.36	0.02	0.07	0.20
hc0	-0.28	0.47	0.50	0.20	-0.46	-0.77	-0.45	-0.36	1.00	0.41	0.49	-0.02
hc1	-0.29	0.32	0.39	0.23	-0.30	-0.70	-0.10	0.02	0.41	1.00	0.58	0.13
hc2p	-0.52	0.46	0.34	0.13	-0.27	-0.87	-0.21	0.07	0.49	0.58	1.00	0.16
hother	-0.16	0.08	-0.04	-0.02	-0.01	-0.22	0.14	0.20	-0.02	0.13	0.16	1.00

figure 5.2. Correlation table for sociodemographic variables

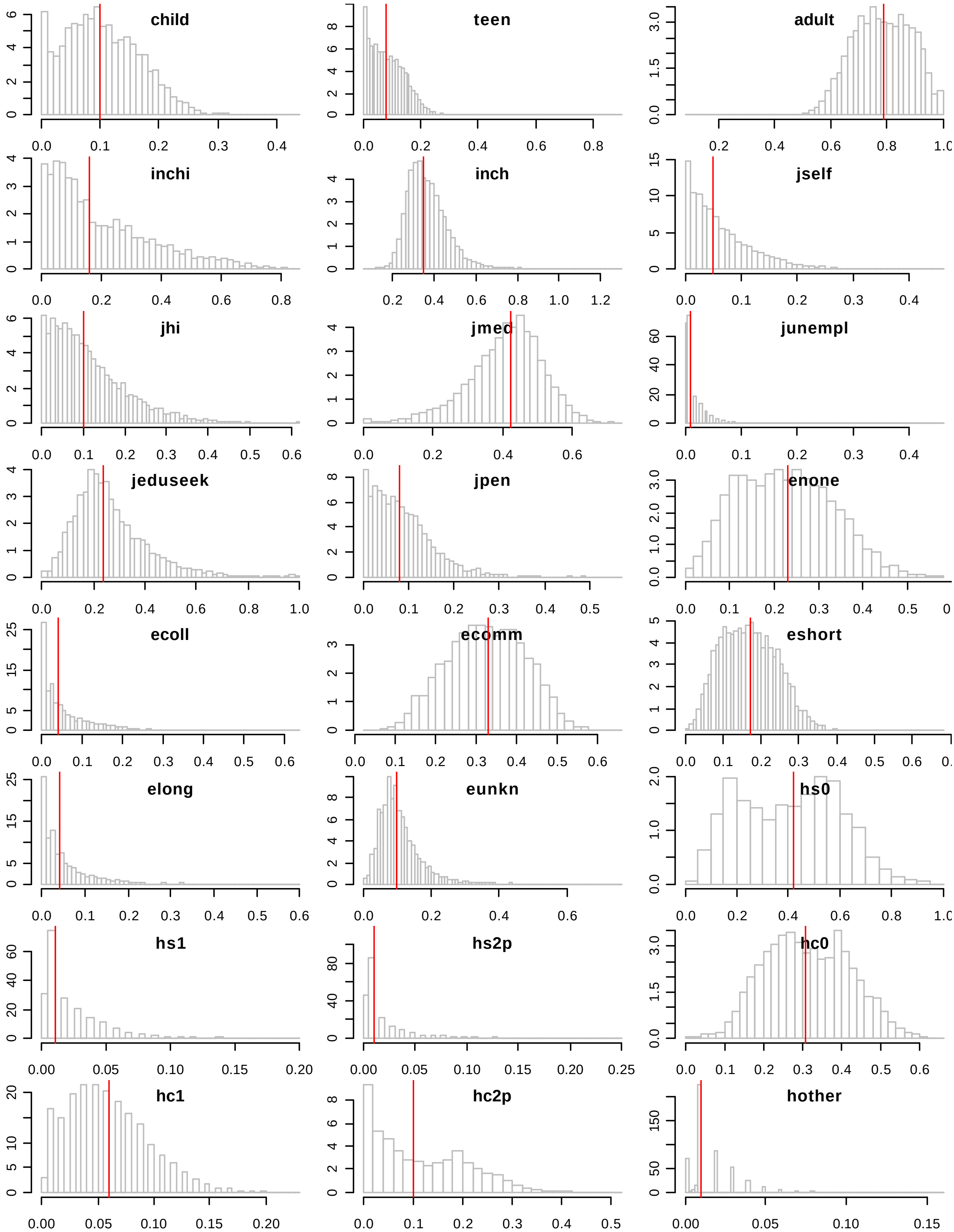


figure 5.3. Histogram for sociodemographic variables. A vertical line marks the median.

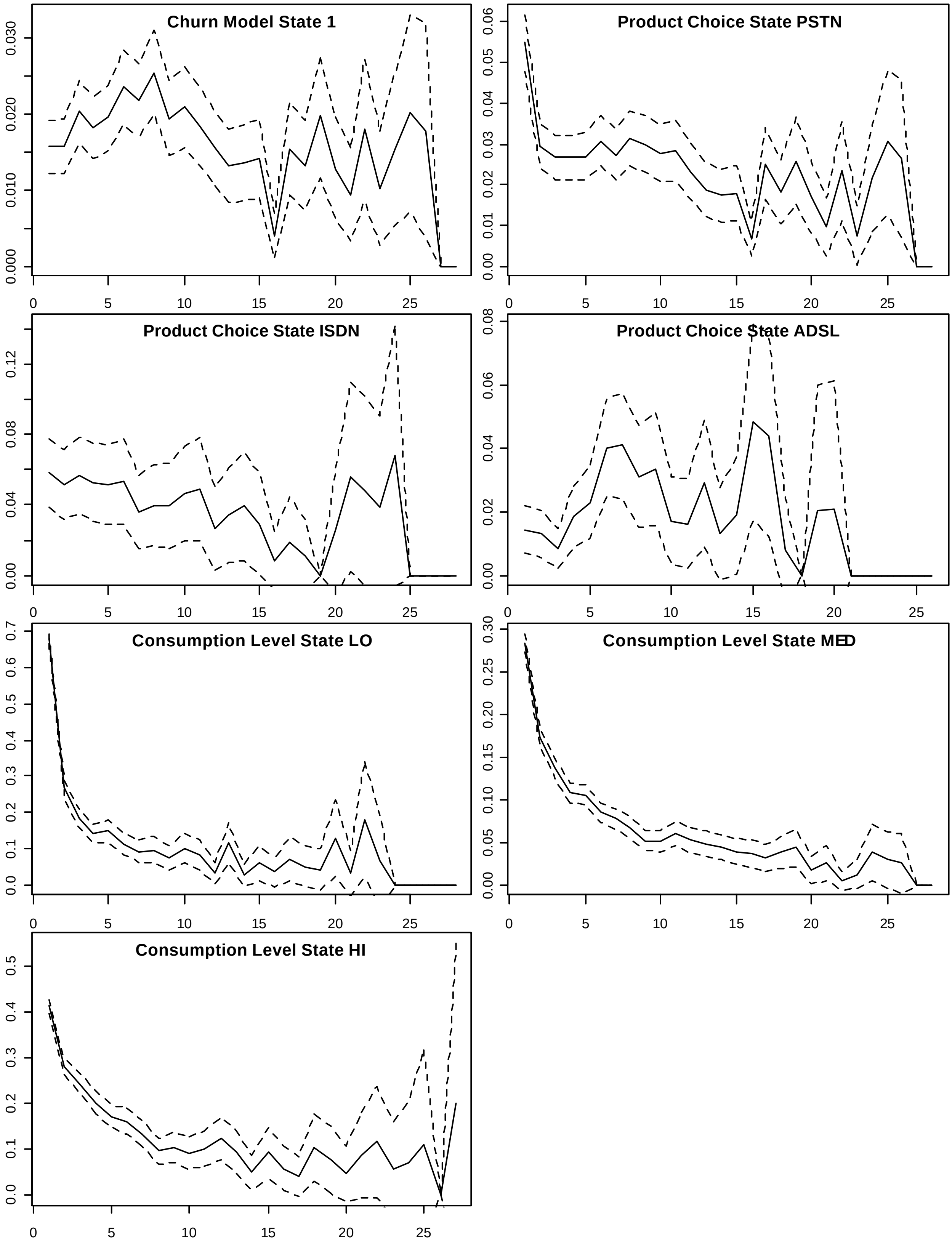


figure 5.4. Discrete Hazard Curves. Average probability of a transition out of a state. Grey lines: 95% confidence bands.

6 Appendix: Results

A display of common parameter estimates can be found below in . *jhi10* is the statistics for the variable *jhi* in the transition from state 1 (PSTN LO) to state 0 (non-customer).

Model for Churn (0, churn),(1, stay)								
Title	Avg	SD	Sig	Q0.5	Q0.025	Q0.05	Q0.95	Q0.975
sigma10	0.4985	0.0780	**	0.4898	0.3854	0.3968	0.6542	0.6898
s12_10	0.3345	0.0817	**	0.3338	0.1872	0.2091	0.4720	0.4931
pstnlo10	1.4436	0.0898	**	1.4417	1.2771	1.2956	1.5915	1.6226
pstnhi10	0.6948	0.0874	**	0.6954	0.5232	0.5449	0.8370	0.8584
internet10	-0.2849	0.0940	**	-0.2842	-0.4647	-0.4389	-0.1429	-0.1074
mu10	-3.8968	0.2577	**	-3.8625	-4.4438	-4.3596	-3.5131	-3.4308
Teen10	1.2677	0.5676	**	1.2333	0.0934	0.3570	2.2022	2.4470
jmed10	0.8048	0.3645	**	0.8014	0.1629	0.2370	1.4075	1.5219
ecomm10	-1.5456	0.5063	**	-1.5595	-2.6212	-2.4313	-0.7822	-0.5926
eshort10	-2.0976	0.5704	**	-2.0853	-3.1628	-3.0467	-1.2135	-1.0323
eunkn10	-1.5766	0.6570	**	-1.6201	-2.7138	-2.5856	-0.5049	-0.2644
inch10	-0.9122	0.4459	**	-0.9275	-1.7056	-1.5824	-0.1318	-0.0389

Model for Product Choice (1,PSTN),(2,ISDN),(3,ADSL)								
Title	Avg	SD	Sig	Q0.5	Q0.025	Q0.05	Q0.95	Q0.975
sigma12	3.8453	0.4748	**	3.7995	3.0861	3.1515	4.6200	4.7297
sigma13	1.9199	0.1780	**	1.9016	1.5936	1.6592	2.2424	2.2962
sigma21	1.1396	0.2467	**	1.1165	0.7664	0.7869	1.5852	1.6735
sigma23	1.9150	0.2735	**	1.9186	1.4258	1.4725	2.3588	2.4193
sigma31	1.1212	0.2905	**	1.1414	0.5160	0.6136	1.5706	1.7193
mu12	-14.7535	1.5472	**	-14.8811	-17.4563	-17.2028	-12.0182	-11.8988
teen12	6.7159	3.7733	*	6.5287	-0.1194	0.8553	13.1094	13.7290
inch12	6.0041	1.8473	**	6.1692	2.2538	2.8238	8.7652	9.0687
pstnhi13	0.6280	0.1520	**	0.6204	0.3415	0.3828	0.8688	0.9131
inettch13	0.9399	0.3792	**	0.9570	0.1868	0.3062	1.5663	1.6448
mu13	-7.4215	0.3700	**	-7.3776	-8.1827	-8.0590	-6.9260	-6.8228
inch13	2.1206	0.7945	**	2.2071	0.4225	0.8043	3.2755	3.4872
mu21	-5.6423	0.2764	**	-5.6335	-6.2283	-6.1138	-5.2292	-5.1775
mu23	-5.5955	0.3970	**	-5.5947	-6.3822	-6.2781	-4.9223	-4.8316
child23	6.3864	2.1342	**	6.3848	2.4604	3.1741	10.1244	10.8509
pstnlo31	-0.9310	0.4582	**	-0.9063	-1.9040	-1.7511	-0.2687	-0.1520
mu31	-3.5775	0.5610	**	-3.5224	-4.7046	-4.5593	-2.7100	-2.5835
inch31	-4.7501	1.4793	**	-4.8405	-7.2523	-7.0198	-2.2055	-1.8201

Model for Consumption Level (1,low),(2,medium),(3,high)								
Title	Avg	SD	Sig	Q0.5	Q0.025	Q0.05	Q0.95	Q0.975
sigma12	0.7328	0.0840	**	0.7307	0.5718	0.5924	0.8744	0.9075
sigma13	1.5936	0.2266	**	1.5890	1.1983	1.2336	2.0352	2.1623
sigma21	1.7985	0.0848	**	1.8014	1.6411	1.6658	1.9390	1.9666
sigma23	1.2674	0.0485	**	1.2694	1.1655	1.1826	1.3486	1.3584
sigma31	1.3713	0.3232	**	1.3190	0.8367	0.8850	2.0209	2.1107
sigma32	0.7255	0.0504	**	0.7224	0.6369	0.6516	0.8158	0.8349
s1_12	0.6454	0.0904	**	0.6421	0.4655	0.4931	0.7922	0.8119
s2_12	0.3223	0.1300	**	0.3118	0.0964	0.1254	0.5666	0.5918
s3_12	0.5216	0.1126	**	0.5209	0.3062	0.3428	0.7153	0.7490
adsl12	-0.4471	0.1001	**	-0.4442	-0.6561	-0.6233	-0.2758	-0.2555
mu12	-2.2348	0.1106	**	-2.2315	-2.4565	-2.4222	-2.0559	-2.0260
teen12	-2.4502	0.6524	**	-2.4517	-3.7968	-3.4847	-1.3905	-1.2070
s1_13	1.2705	0.2373	**	1.2780	0.7818	0.8538	1.6490	1.6967

s2_13	0.9129	0.5680	-	0.8612	-0.1874	-0.0046	1.9328	2.0486
s3_13	-0.9464	0.5190	**	-0.9100	-2.0462	-1.8466	-0.1491	-0.0484
s5_13	1.0918	0.4617	**	1.0596	0.3226	0.4030	1.9252	2.1757
isdn13	-0.5515	0.3334	*	-0.5501	-1.2384	-1.1107	-0.0211	0.0843
adsl13	-1.6051	0.3482	**	-1.5997	-2.3487	-2.2089	-1.0631	-0.9724
mu13	-5.3318	0.4549	**	-5.3073	-6.4524	-6.1831	-4.6774	-4.6178
s1_21	0.3508	0.0686	**	0.3484	0.2143	0.2400	0.4640	0.4866
s5_21	0.3033	0.0952	**	0.3050	0.1094	0.1374	0.4553	0.4819
isdn21	0.4407	0.1576	**	0.4412	0.1220	0.1744	0.7063	0.7627
adsl21	0.5046	0.1210	**	0.5062	0.2704	0.3022	0.6997	0.7393
inettch21	-1.0653	0.5716	**	-1.0380	-2.2450	-2.0726	-0.2515	-0.0579
mu21	-3.9129	0.2094	**	-3.9045	-4.3519	-4.3077	-3.5695	-3.5299
jmed21	-1.5004	0.4308	**	-1.5224	-2.2811	-2.1707	-0.7062	-0.5881
hs2p21	4.1931	1.8749	**	4.1284	0.6622	1.2982	7.3462	7.7055
s1_23	0.3040	0.0570	**	0.3073	0.1890	0.2024	0.3932	0.4084
s2_23	0.1447	0.0724	**	0.1376	0.0148	0.0318	0.2653	0.3059
s3_23	0.1728	0.0706	**	0.1781	0.0180	0.0488	0.2868	0.2962
s7_23	0.4411	0.0768	**	0.4486	0.3037	0.3147	0.5665	0.6011
isdn23	0.5180	0.0992	**	0.5212	0.3301	0.3548	0.6719	0.7144
mu23	-4.2892	0.1457	**	-4.2970	-4.5216	-4.5009	-4.0504	-4.0022
teen23	2.2014	0.5052	**	2.1994	1.1854	1.3822	3.0568	3.1808
pen23	2.5148	0.5216	**	2.5166	1.4741	1.6618	3.4371	3.5567
inch23	1.0736	0.3007	**	1.1047	0.4781	0.5509	1.5104	1.5548
s1_31	0.1436	0.1659	-	0.1488	-0.2073	-0.1274	0.4047	0.4697
s5_31	0.4666	0.2610	*	0.4595	-0.0668	0.0382	0.8836	0.9601
adsl31	0.4134	0.2208	*	0.4155	-0.0093	0.0608	0.7877	0.8566
mu31	-8.4085	0.5420	**	-8.3677	-9.6074	-9.3699	-7.5660	-7.4272
jmed31	3.7239	0.8604	**	3.6425	2.1574	2.3982	5.3874	5.5916
edu_seek31	3.2287	0.7342	**	3.1677	1.8890	2.1388	4.5172	4.7266
pen31	7.5359	1.2379	**	7.4957	5.1729	5.5743	9.6053	10.0598
s1_32	0.3433	0.0484	**	0.3450	0.2527	0.2652	0.4200	0.4378
s3_32	0.3038	0.0670	**	0.3030	0.1627	0.1933	0.4133	0.4305
s7_32	0.6097	0.1069	**	0.6114	0.4051	0.4297	0.7916	0.8169
Isdn32	-0.4690	0.0844	**	-0.4691	-0.6334	-0.6097	-0.3329	-0.3058
mu32	-2.0982	0.1212	**	-2.0969	-2.3274	-2.3002	-1.9083	-1.8893
teen32	-1.4361	0.4694	**	-1.4558	-2.3157	-2.1780	-0.6740	-0.5654
i10032	-0.2773	0.1698	*	-0.2777	-0.6020	-0.5644	-0.0130	0.0370
hc132	2.0332	0.8390	**	2.0345	0.4256	0.6922	3.3167	3.7523

figure 6.1. Parameter Estimates. Avg: average, SD: standard deviation of parameter. Sig: Significance i.e. is zero outside the confidence interval? “” Yes, outside [2.5%, 97.5%], “*” Outside [5%, 95%], “-“ Inside, not significant. Q*x*: *x*-quantile.**

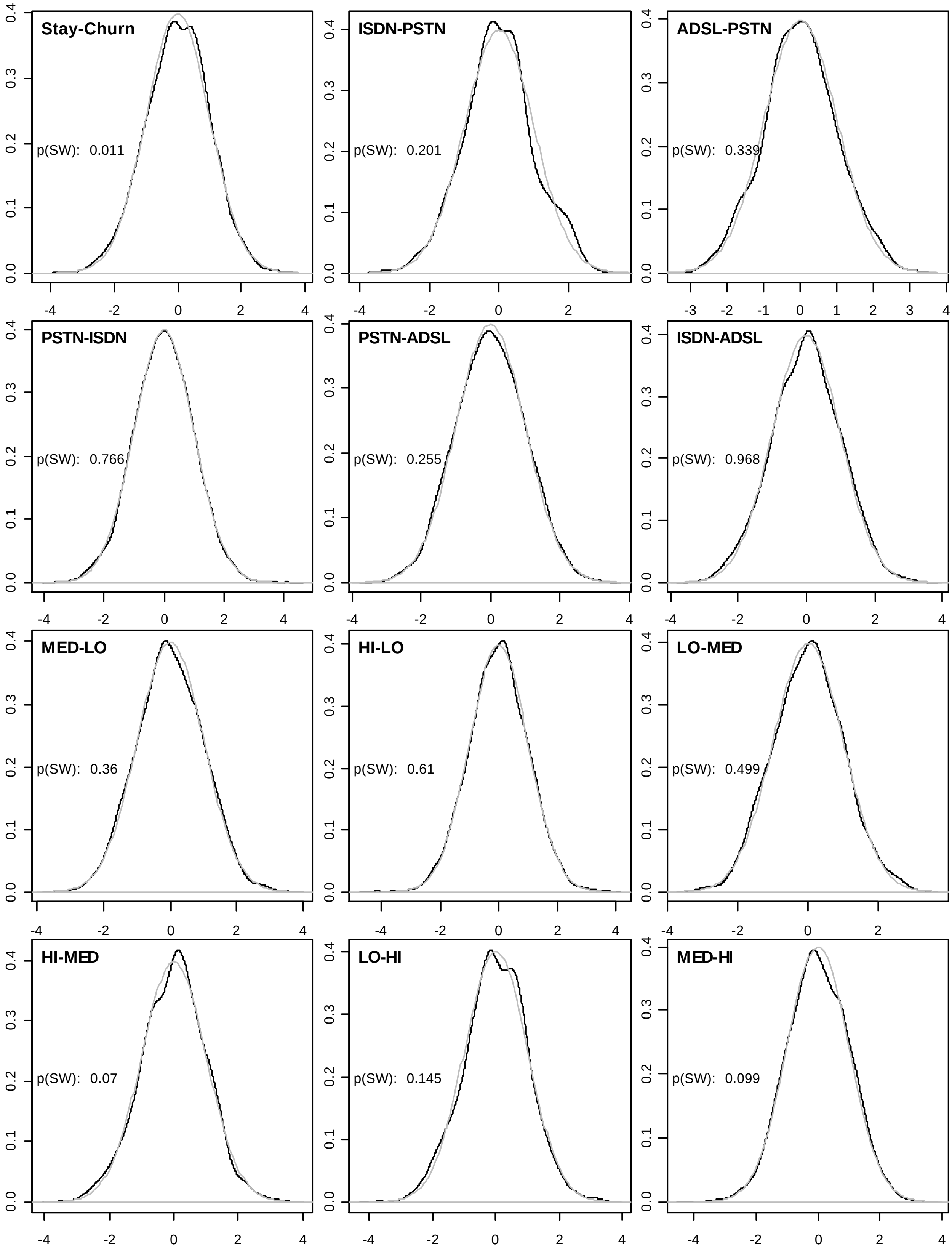


figure 6.2. Unobserved heterogeneity. Black line: Kernel estimate of empirical density. Grey line: Normal density. "p(SW)": p-value of Shapiro and Wilk's test for normality.

7 Bibliography

Allenby, G. M. and P. E. Rossi (1999), "Marketing models of consumer heterogeneity," Journal of Econometrics, 89 (1-2), 57-78.

Anderson, B., A. McWilliam, H. Lacohee, E. Clucas, and J. Gershuny (1999), "Family life in the digital home - domestic telecommunications at the end of the 20th century," Bt Technology Journal, 17 (1), 85-97.

Blattberg, R. C. and J. Deighton (1996), "Manage marketing by the customer equity test," Harvard Business Review, 74 (4), 136.

Chib, S., P. B. Seetharaman, and A. Strijnev (2002), "Analysis of multi-category purchase incidence decisions using IRI market basket data," in Econometric Models in Marketing, Vol. 16.

Donkers, Bas, Peter C. Verhoef, and Martijn de Jong (2003), "Predicting Customer Lifetime Value in Multi-Service Industries," in ERIM Report Series. Rotterdam.

Edwards, Y. D. and G. M. Allenby (2003), "Multivariate analysis of multiple response data," Journal of Marketing Research, 40 (3), 321-34.

Heckman, J. J. (1981), "Heterogeneity and State Dependence," in Studies in Labor Markets, S. Rosen, Ed.: University of Chicago Press.

Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric-Models for Duration Data," Econometrica, 52 (2), 271-320.

Hoel, Paul G., Sidney C. Port, and Charles J. Stone (1987), Introduction to Stochastic Processes. Boston: Waveland Press, Inc.

Hsiao, Cheng (2003), Analysis of Panel Data: Cambridge University Press.

Lancaster, Tony (1990), The Econometric Analysis of Transition Data: Cambridge University Press.

Luenberger, David G. (1997), Investment Science: Oxford Press.

Manchanda, P., A. Ansari, and S. Gupta (1999), "The "shopping basket": A model for multicategory purchase incidence decisions," Marketing Science, 18 (2), 95-114.

Miravete, Eugenio J. and Ignacio Palacios-Huerta (2002), "Learning Temporal Preferences," CEPR Discussion Papers.

Morrison, Donald G., Richard D. H. Chen, Sandra L. Karpis, and Kathryne E. A. Britney (1982), "Modeling Retail Customer Behavior At Merrill Lynch," Marketing Science, 1 (2).

Reinartz, W. J. and V. Kumar (2003), "The impact of customer relationship characteristics on profitable lifetime duration," Journal of Marketing, 67 (1), 77-99.

Robert, Christian P. and George Casella (1999), Monte Carlo Statistical Methods (Third ed.): Springer.

Rossi, P. E. and G. M. Allenby (2000), "Statistics and marketing," Journal of the American Statistical Association, 95 (450), 635-38.

Schmittlein, D. C. and R. A. Peterson (1994), "Customer Base Analysis - an Industrial Purchase Process Application," Marketing Science, 13 (1), 41-67.

Working Papers published by the Department of Management

2005-1: Esben Koling Laustrup and Johannes Raaballe: Udbytteannonceringseffekten i Danmark.

2005-2: Carina Sponholtz: Separating the Stock Market's Reaction to Simultaneous Dividend and Earnings Announcements.

2005-3: Kyle Bagwell and Per Baltzer Overgaard: Look How Little I'm Advertising!

2005-4: Jerker Nilsson and Ole Øhlenschlæger Madsen: Cross-border Mergers Between Agricultural Co-operatives - a Governance Perspective.

2005-5: Bent Jesper Christensen and Morten Ørregaard Nielsen: The Effect of Long Memory in Volatility on Stock Market Fluctuations.

2005-6: Bent J. Christensen and Nicholas M. Kiefer: Investment in Advertising Campaigns and Search: Identification and Inference in Marketing and Dynamic Programming Models.

2005-7: Flemming Witt and Jørn Flohr Nielsen: The Difficult Empowerment in Danish Hospitals: Power to the Nurses!?

2006-1: Thomas Busch, Bent Jesper Christensen and Morten Ørregaard Nielsen: The Information Content of Treasury Bond Options Concerning Future Volatility and Price Jumps.

2006-2: Thomas Busch, Bent Jesper Christensen and Morten Ørregaard Nielsen: Forecasting Exchange Rate Volatility in the Presence of Jumps.

2006-3: Thomas Busch, Bent Jesper Christensen and Morten Ørregaard Nielsen: The Role of Implied Volatility in Forecasting Future Realized Volatility and Jumps in Foreign Exchange, Stock, and Bond Markets.

2006-4: Niels Stender: Expected Present Value of a Customer.

2006-5: Niels Stender: A Quantitative Model of Dynamic Customer Relationships.