

Understanding the Differences in Customer Portfolio Characteristics and Insurance Consumption across Distribution Channels

Summary

Note: Work in progress. A current version of the paper can be obtained from the authors.

Key words: non-life insurance · multichannel behavior · cross-channel customer · insurance consumption

1 Motivation

Today's multichannel presence of insurance companies allows customers to research and purchase insurance products over multiple channels with an increasing preference for online channels. As an outcome, a recent study showed that in 2012, 57% of insurance customers worldwide chose the insurer's own website for product and premium research, a 6% increase since 2010. A similar trend is also visible for policy purchases with 20% of customers using the insurer's website in 2012, a 3% growth since 2010.

In order to extend the understanding of multichannel research and purchase behavior for insurance products we conduct a data-driven study with a special focus on cross-channel customers, addressing the following questions: How customer characteristics (1) impact the channel choice, (2) lead to channel related customer typologies/clusters, and (3) lead to channel specific consumption behavior. Prior studies for various industries in the field of customer channel preferences focused on the difference between pure online and pure offline customers, but lacked the segment of cross-channel customers also referred to as research shoppers, who research insurance products online but purchase the policy offline. These customers have become a research topic of interest in the recent years. Furthermore previous work in this domain included only few general customer characteristics, such as age or gender, and was limited to minor samples created through (online) surveys.

In our study we are able to expand existing work by adding more detailed characteristics like nationality, level of urbanicity of the customer residence, customer status (existing or new customer), and annual premium. Moreover we are taking a data-driven approach where we derive our conclusions from a major insurance customer sample of over 370'000 observations. We hypothesize that all the observed channel segments vary in their characteristics and form distinct customer typologies/clusters, however we expect cross-channel and pure online customers to be alike, due to the fact that both segments show an affinity for the usage of online media during the phase of product research. We further assume pure offline customers to have the highest insurance consumption, because only in this channel the full range of coverage is sold by the insurer. Our findings are relevant to academics and practitioners alike and are important for strategic customer management and detailed understanding of multichannel customer behavior.

2 Methodology

For this work we apply a data-driven approach to verify our hypotheses and build our study on a recent sample of 370'000 customer and policy sets from the data warehouse of a large Swiss insurer. The dataset captures the period July 2011 to June 2014. This company has a total non-life market share of 26% in Switzerland, therefore the results of the analysis are strongly relevant for the Swiss market.

The sample includes the private-line products *Household/Liability*, *Motor*, and *Travel* and is restricted to new policies, where all characteristics are measured at the inception date of the policy. Policyholders are classified into three channel segments: *Pure Online*, *Cross-Channel*, and *Pure Offline*. The channel segments pure online and pure offline are identified via a channel flag in the policy data, while the cross-channel cases are detected via a matching algorithm. The following table provides a summary of the dataset quantities per product and channel segment.

Insurance Products	Channel Segments			Total
	Pure Online	Cross-Channel	Pure Offline	
Household/Liability	1'958	2'384	123'459	127'801
Motor	2'211	10'833	205'570	218'614
Travel	1'131	1'625	24'920	27'676
Total	5'300	14'842	353'949	374'091

To test our different hypotheses, we include the following independent variables into our study: age, gender, nationality, civil status, urbanicity, local region, income, customer status, length of relationship, annual premium, bonus ratio (only for motor insurance). For the analysis of research question (1) – impact of customer characteristics on the channel choice – and in line with our expectations we compare channel segments in two constellations: (a) pure online & cross-channel vs. pure offline, and (b) pure online vs. cross-channel. For both constellations we first provide empirical evidence of the impact on channel choice through customer characteristics by applying statistical analysis, such as Student's t-test and Pearson's chi-square test, and second quantify the impact of each characteristic on the channel choice by conducting multiple logistic regression for dual choices. Our logistic regression model to quantify the channel impact is defined as:

$$\mathbb{P}(\vec{X}) = \frac{e^{\beta_0 + \vec{\beta}\vec{X} + \varepsilon}}{1 + e^{\beta_0 + \vec{\beta}\vec{X} + \varepsilon}} \quad (1)$$

with $\beta_0, \vec{\beta} = (\beta_1, \dots, \beta_n)$ being the estimated coefficients and $\vec{X} = (X_1, \dots, X_n)$ being the vector of the above mentioned independent variables (characteristics). We perform stepwise regression and test the accuracy of the model in each step using k -fold cross-validation.

To address research question (2) – impact of customer characteristics on the existence of channel related customer typologies/clusters – we apply clustering methods, such as hierarchical clustering and k-means clustering, and derive channel specific customer typologies, with a special interest in highly online affine customers. To verify our hypotheses in research question (3) – channel impact on insurance consumption – we conduct an empirical analysis of the annual premium and the bonus ratio using similar methods as for research question (1).

3 Preliminary Results

The results of this study are discussed separately for each product. For the product *Household/Liability* insurance our results indicate a significant impact of customer characteristics on channel choice and leading to channel specific insurance consumption.

Looking at research question (1), the results of the statistical analysis show empirical evidence for a significant impact on channel choice for all measured characteristics comparing pure online & cross-channel vs. pure offline customers. We found the first-mentioned segments more likely (a) to be younger; (b) to be female; (c) to have a nationality from a neighboring State of Switzerland; (d) to live in a single household; (e) to be residents of urban areas especially Switzerland’s metropolitan regions Zurich, Geneva, and Basel in a commune with a higher income level; (f) to be a new customer; and (g) to have a short relationship with the insurance, when purchasing this products as an existing customer. When performing the logistic regression, we could quantify the strongest impact factors having a positive relation with being pure online & cross-channel customer: (a) age (25-34 years); (b) nationality (neighboring State of Switzerland); (c) urbanicity (live in an urban area); (d) income (75-85 kCHF); (e) customer status (being a new customer).

When comparing pure online vs. cross-channel customers an impact on channel choice through the characteristics is still visible, but the effect is more moderate. We found the pure onliners more likely (a) to be even younger; (b) to be residents of urban areas especially Switzerland’s metropolitan regions Zurich and Geneva in a commune with a higher income level; and (c) to be new customer. In the regression model only the characteristics urbanicity (live in an urban area) and customer status (being a new customer) have a significant positive relation with being a pure online customer.

In summary the described results are as expected, however we found one unanticipated observation. Surprisingly the youngest customers (< 25 years) do not have the highest online affinity for insurance products, but the more mature customers (25-44 years). We assume that the youngest customers are less confident in an online-only purchase, as they have no experience with insurance products, whereas more mature customers have an advantage of being experienced insurance customers and online affine.

For research question (2) we find specific online affine customer clusters and even a distinction of pure online and cross-channel customers. The results regarding research question (3) were not as expected. We find the cross-channel customers to have the highest average annual premium with 336.10 CHF of all three channel segments, compared to pure onliners (300.24 CHF) and pure offliners (313.78 CHF). The effect between pure online and cross-channel customers might be explained best by up-selling through the personal contact of the cross-channel customers during the sales process, as both customer groups see an identical product offering (coverages and franchises) during the online product research at the insurer’s website.

For the insurance products *Motor* and *Travel* we expect to observe similar results.

This research work is part of the ongoing PhD thesis project of one of the authors. The working paper is expected to be fully finalized soon.