Modeling of Policyholders Fund Switching Behavior within the Swedish Unit-linked Market

Joakim Mossberg

Master's thesis
2012:E6

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

TYP AV DOKUMENT		DOKUMENTBETECKNING
☒ Examensarbete	☐ Kompendium	
☐ Delrapport	☐ Rapport	LUTFMS-3188-2012

INSTITUTION

Matematikcentrum. Matematisk statistik, Lunds universitet, Box 118, 221 00 LUND

FÖRFATTARE

Joakim Mossberg

DOKUMENTTITEL OCH UNDERTITEL

Modeling of Policyholders Fund Switching Behavior within the Swedish Unit-linked Market

SAMMANFATTNING

This thesis is written at the actuarial department at SEB Trygg Liv. The project is a pre study on the impact of fund switching within the Swedish unit-linked market. This includes individual factors that are believed to explain why policyholders switch funds. Generalized linear models will be used to predict the probability of occurrences of fund switches. The effect the fund switches will have on future revenues, on the basis how the economic climate on the Swedish equity and fixed-income markets will also be analyzed.

NYCKELORD

DOKUMENTTITEL OCH UNDERTITEL - SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL

Modellering av fondbyten inom fondförsäkring på den svenska marknaden

UTGIVNINGSDATUM		ANTAL SID	SPRÅK		
år 2012	mån 04		☐ svenska	☒ engelska	☐ annat

ÖVRIGA BIBLIOGRAFISKA UPPGIFTER	ISSN
	ISBN
	2012:E6

Signature ______________________ Date ______________

Table of Contents

Table of Contents 3

1 Background 6

1.1 SEB Trygg Liv 6

1.2 Purpose and outline 6

2 Theory 7

2.1 Binary Regression Models 7

2.1.1 Linear Probability Model 7

2.1.2 Probit and Logit model 8

2.1.3 Model diagnostics 9

2.1.4 Variance inflation factor analysis 9

2.2 Panel data 10

2.2.1 Binary choice models 10

2.2.2 Repeated Cross sectional analysis 11

2.3 Time series 11

2.3.1 Testing for normality 11

3 Unit-linked contracts 12

4 Analysis – Individual factors 14

4.1 Fund switch categories 14

4.2 Explanatory variables & hypothesis formulation 15

4.2.1 Value change 15

4.2.2 Value 16

4.2.3 Pension insurance 16

4.2.4 Male 16

4.2.5 External insurance broker 16

4.2.6 Fund placement service 17

4.2.7 Age 17

4.3 Data 18
4.3.1 Dataset construction 18
4.3.2 Dataset 18
4.4 Correlation analysis 20
4.4.1 Variance inflation factor analysis 21
4.5 Repeated Cross sectional models for switching behavior 22
4.5.1 Logistic regression analysis 23
4.5.2 Model diagnostics 24
4.6 Average model for switching behavior 26
4.7 Hypothesis testing 27
4.7.1 Value change 28
4.7.2 Value 28
4.7.3 Pension insurance 29
4.7.4 Male 29
4.7.5 External insurance broker 29
4.7.6 Fund placement service 30
4.7.7 Age of the policyholder 30
4.8 Result 31
5 Analysis – Impact on future revenues 36
5.1 Data 36
5.2 Model for change in future revenue 36
5.2.1 Model diagnostics 38
5.2.2 Model specification validation 39
5.3 Result 41
6 Conclusion and closing discussion 42
7 References 43
Appendix 1 – GLM results 44

1 Background

Unit-linked contracts stand for about one third of the Swedish life-insurance market, when it comes to premium payments in 2011 (Svenskförsäkring 2012). Unit-linked contracts differ from traditional contracts since the policyholder chooses the investment portfolio (to some extent) instead of the life-insurance company. In the unit-linked case the life-insurance company would typically offer a range of mutual funds in which the policyholder can choose to invest. Therefore, the return under the policy is dependent on the investment choice, while in the traditional case, when the company chooses the portfolio the return to the policyholder is set to a guaranteed level.

The policyholders are usually free to switch between funds and dependent on the choice of funds the policyholder holds in his or hers portfolio the life-insurance companies future revenues are effected. Typically fixed income funds are cheaper for the policyholder to hold than funds that invest their capital in the stock market.

1.1 SEB Trygg Liv

Skandinaviska Enskilda Banken (SEB) is one of the leading corporate banks in Sweden, among their products they provide life-insurance contracts and SEB Trygg Liv is the division handling all unit-linked contracts. The project will be done at SEB Trygg Liv beginning in November 2011.

1.2 Purpose and outline

Modeling how the policyholders behave in respect to fund switching within their insurances is needed for future forecasts and solvency capital requirements calculations.

The main goal for the project is to find factors that affect policyholder's behavior in respect to switching funds. Is there a connection between fund switching and the performance of the unit-linked contract? Intuition says that there is a connection, for instance if the stock market goes down policyholders will switch to more fixed-income type funds and decrease the market risk in their portfolio. However is this statement true within the Swedish unit-linked market?

The main hypothesis to be tested is:

H1: If the performance of the insurance contract decrease policyholders will be more likely to switch to one or more funds with a lower risk profile.

Other individual factors that will impact policyholders switching behavior is also sought, possible candidates and hypothesis are presented in section 4.2.

Second to this, non-individual factors such that market indices is sought to explain how fund switches effect future revenues. This is done in section 5. For the same reasons as above, it is expected that policyholder switches to more fixed-income type of funds if the stock market goes down, a model that explains the effect that these changes will have on future revenues is sought. If a model of this can be found, stress test on future revenues can be made, depending on once assumptions of the stock and fixed-income market.

2 Theory

Switching behavior of customers has been studied a lot in several fields. Numerous studies on customer behavior and customer loyalty have been published, see for instance Guadagni and Little 2008; Clemes, Gan and Zhang 2010; Wieringa and Verhoef 2007. Most of the studies made have focused on how a customer might switch to a competitor (i.e. Bank or Grocery store) or how customer change their consumption behavior (Which brand to buy etc.). Since the econometric nature of the data modeled most methods used to explain switches of some sort is regression analysis.

In order to test the hypothesis our method of choice is to modeling fund switches with a binary regression model, since either we observe a certain switch or we do not observe a switch. The most common ones are the Probit and Logit (Logistic regression) Models.

2.1 Binary Regression Models

Binary regression models are a technique in which the dependent variable in the regression takes values which are either 0 or 1.

2.1.1 Linear Probability Model

The linear model for a binary variable y_i were i represent the sample size (i=1...N) and x_i represent the associated vector of explanatory variables is given by

$$y_i = x'_i\beta + e_i \,. \qquad (2.0)$$

The standard assumption is that the residual or error term has a zero expected value conditioned on its vector of explanatory variables, thus:

$$P(y_i = 1 \,|x_i) = x_i'\beta$$
$$E[y_i|x_i] = P(y_i = 1|x_i) = x_i'\beta \qquad (2.1)$$

The parameter vector β can be estimated using ordinary least squares (OLS). The major drawback of this model is that unless restrictions are placed on β one can observe probabilities outside the closed interval [0,1], for this reason there exists a class of Generalized linear model (GLM). In general we have

$$P(y_i = 1 \,|x_i) = G(x_i, \beta). \qquad (2.2)$$

So the probability of $y_i = 1$ depends on the explanatory variables vector x_i. In our case the probability that an individual does a fund switch depends on a vector of individual explanatory variables. The function G(.) should only take values in the interval [0,1]. Denoting $G(x_i, \beta) = F(x_i'\beta)$ the function F(.) also has to take values in the interval [0,1]. The natural choice for binary choice models is to choose F to be some distribution function (Verbeek 2008, Ch. 7). Note that if F is chosen as the uniformed distribution over the interval [0,1], we will get a similar model as in 2.1, but without any possibility of observing a probability outside the interval [0,1] because of the boundaries from the uniformed distribution function.

2.1.2 Probit and Logit model

The Probit model uses the standard normal distribution function

$$F(w) = \Phi(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\} dt. \tag{2.3}$$

While the Logit model similarly uses the standard logistic distribution function

$$F(w) = L(w) = \frac{e^w}{1+e^w}. \tag{2.4}$$

Note that the function F(.) only is being used to model p_i and does not denote the cumulative distribution of the dependent binary variable itself. The logit function is defined as

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right), where\ p_i = P(y_i = 1|x_i). \tag{2.5}$$

This is sometimes called the log-odds ratio, a ratio of ρ means that the odds of y_i = 1 are ρ-times those of y_i = 0.

The model in 2.2 can thus be written as

$$\ln\left(\frac{p_i}{1-p_i}\right) = x_i'\beta. \tag{2.6}$$

Equation 2.6 displays the standard model for the simple linear logistic regression. In ordinary regression the interpretation of a single coefficient β_k displays the marginal effect, so if the mean of x increases by one the response increases with β_k. A positive β_k in a logistic regression indicates an increase in the odds of the response.

Transforming probabilities to odds is a monotonic transformation. The odds are defined as the probability of success divided by the probability of failure. Probabilities ranges from 0 to 1 and the corresponding odds range from 0 to positive infinity. If the probability is 0.5 the odds of success is then 1, usually we say that the odds are 1 to 1, i.e. a 50-50 chance.

To estimate these binary choice models one typically uses maximum likelihood (Christensen 1997, Ch. 2). The likelihood contributions from the two discrete outcomes are

$$P(y_i = 1|x_i)\ for\ y_i = 1 \tag{2.7}$$
$$P(y_i = 0|x_i) for\ y_i = 0 \tag{2.8}$$

So the likelihood function for the entire sample is

$$L(\beta) = \prod_{i=1}^{N} P(y_i = 1|x_i; \beta)^{y_i} P(y_i = 0|x_i; \beta)^{1-y_i}. \tag{2.9}$$

And the corresponding log-likelihood function is thus given by

$$l(\beta) = \sum_{i=1}^{N} (y_i \ln F(x_i'\beta) + (1-y_i)\ln(1 - F(x_i'\beta))). \tag{2.10}$$

Maximizing the log-likelihood function is done in either closed form or with numerical techniques depending on the model choice, F function.

Interpreting the regression coefficients whether in a binary or in an ordinary regression can be very tricky. The main problems consist of the possible correlation between the explanatory variables, omitted variables or unobserved heterogeneity. Various diagnostics for the model and methods for finding and obtaining valid statistical inferences under these conditions exist and the relevant methods/diagnostics for this thesis are described in the sections that follow.

2.1.3 Model diagnostics

In OLS regression the most widely used measure for model validation is the coefficient of determination (R2). Even though many econometricians criticisms the usefulness of this measure, for instance Achen 1990. For the GLM models it does not exist an equivalent to this measure. Several pseudo R-square measures has been introduced, the two mostly used are Efron and Mcfadden R2. These measures have several different names but these ones are the two most notoriously names for the measures.

$$R^2_{Efron} = 1 - \frac{N}{N_1 N_0} \Sigma(y_i - \widehat{p_\iota})^2 \tag{2.11}$$

$$R^2_{Mcfadden} = 1 - \frac{l_u(\widehat{\beta})}{l_r(\widehat{\beta}} \tag{2.12}$$

Where N_0, N_1 represents the number of zero and non-zero elements in the dependent vector y respectively, and l_u and l_r represents the log-likelihood function for the unrestricted and the restricted model respectively. Where the restricted model, only has the intercept term as the regressor.

2.1.4 Variance inflation factor analysis

If there exist a correlation between the regressors there is a problem with collinearity. If the two or more regressors are perfectly collinear the maximum likelihood estimates are no longer efficient. In the presence of that, the best way is to simply drop one of the regressors for the analysis since all the variation in that regressor can be explained by its collinear counterpart. If two or more regressors are correlated but not perfect correlated the same problem may exists. In accordance with this the parameter estimates are hard to interpret since it is not possible to distinguish between the variables and the estimates are no longer unbiased. Signs of collinearity are a high R2, even though the model fit is not that good, large parameter estimates and standard deviation.

To find out whether regressors are collinear it is possible to perform a variance inflation factor analysis. The simplest way is to perform k linear regressions, where k represents the number of regressors. For each k regress the k:th variable on the rest of the regressors as

$$x_1 = \alpha_2 x_2 + \alpha_3 x_3 + \dots + a_k x_k$$
$$x_2 = \alpha_1 x_1 + a_3 x_3 + \dots + a_k x_k$$
$$\dots$$
$$x_k = a_1 x_1 + a_2 x_2 + \dots + a_{k-1} x_{k-1}$$

And calculate the ordinary R-square value for each of the k regressions. The variance inflation factor, VIF is then given by

$$VIF_k = \frac{1}{1-R_k{}^2}\,. \tag{2.13}$$

If VIF is larger than 5 we assume that the multicollinearity is high and one or more of the regressors should be omitted for the main regression analysis.

2.2 Panel data

Panel data consist of both cross sectional and time series dependent observations, e.g. observations of different policyholders fund switch, collected over a number of periods. The advantage with panel data in terms of regression is that estimators based on panel data are often more accurate (Verbeek 2008, Ch. 10), since the explanatory variables vary over two dimensions, rather from one, reducing the collinearity between the explanatory variables therefore improves the efficiency in the estimates (Hsiao 2003, page 3). Panel data makes it possible to analyses changes on an individual level, instead on an average level.

2.2.1 Binary choice models

Suppose we observe the discrete outcome y if an latent (unobserved) continuous variable y* crossing a certain threshold.

$$y_{it} = \begin{cases} 1 \ \ if\ y_{it}^* > 0 \\ 0 \ \ if\ y_{it}^* \le 0 \end{cases}$$

$$y_{it}^* = x_{it}\beta + v_{it} \tag{2.14}$$

Where t is the time index, t=1...T. To distinguish between average behavior and individual behavior the residual term v_{it} is decomposed as

$$v_{it} = \ \alpha_i + u_{it} \tag{2.15}$$

Where α_i displays the individual specific contribution to the error term. It is assumed to be fixed over all time periods. *The fixed effect logit model* is thus given by 2.16.

$$y_{it}^* = \alpha_i + x_{it}'\beta + u_{it},\ u_{it} \sim IID(0, \sigma_u^2) \tag{2.16}$$

The problem with the fixed effect logit model has to do with the estimation technique. To see this let us examine model 2.16.

$$P(y_{it} = 1) = P(y_{it}^* > 0) = P(u_{it} > \ -x_{it}'\beta - a_i) = F(a_i + x_{it}'\beta). \tag{2.17}$$

Expanding the log likelihood function given by eq 2.10 with result from 2.17, the log likelihood function looks like

$$l(\beta, a_1, \dots, a_N) = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}\ln F(x_{it}'\beta + a_i) + (1 - y_{it})\ln(1 - F(x_{it}'\beta + a_i))). \tag{2.18}$$

Maximizing 2.18 will lead to inconsistent estimates of the parameters $(\beta, a_1, \dots, a_N)$ for a fixed number of time points T, (see Hsiao 2003 Ch. 3-4 or Baltagi 2005 page 209.) The reason for this is that for a fixed T, the number of parameters grows with the sample size N, this is known as the incidental parameter problem. One way of estimating the parameters consistent is to use, as suggested by Chamberlain (1980) maximizing the conditional likelihood function.

Another approach for model panel data with a binary choice model is the random effects model.

2.2.2 Repeated Cross sectional analysis

One problem that often occurs with panel data is missing data points. Since the data on policyholders fund switch that is available does not consist of the same individuals over time. New individuals buys insurance contracts, new individuals transfer their insurance from one life-insurance company to another. This is in general not a problem since these individuals can be disregarded in the model. However there is a problem with annulments, transfers and deaths from the existing sample, because the same policyholders are then no longer observable over the different time periods. For this reasons, and for the estimation problems discussed in the previous section, a panel data analysis cannot be done to full extent. Instead a repeated cross section analysis will be performed to analyze some individual factors that might contribute to the customers fund switching behavior. This will be done repeatedly over some time steps to check wheatear that factors change over time. If the factors show to be stable over time an average model will be evaluated.

2.3 Time series

2.3.1 Testing for normality

To test whether a time series comes from a normal distribution with unknown mean and variance, one usually check for the third and fourth moment, i.e. the skewness and excess kurtosis. The normal distribution has expected zero third- and fourth moment. If the expected third moment is not zero, the distribution is skewed, i.e. it is not symmetric around zero. If the distribution has excess kurtosis it has fatter tails than the normal distribution. To test if the third- and fourth moment satisfies this condition one can use the Jarque-Bera test, the test statistics is given by

$$JB = \frac{N}{6}\left(s^2 + \frac{(k-3)^3}{4}\right) \tag{2.19}$$

Where s and k is the sample skewness and sample kurtosis respectively. Under the null hypothesis JB is Chi-square distributed with 2 degrees of freedom. (Verbeek 2008, page 195)

3 Unit-linked contracts

Several different life- insurance contracts exist on the Swedish market. In this section the various contracts will be briefly described, since a full description of every contract is outside the scope of this thesis, the reader need to have some basic knowledge about the Swedish market.

All unit-linked contracts are either considered to be an Endowment policy or pension insurance. Pension insurance can then be either an occupational pension or a privately own policy. The occupational pension plans, are either bound to a collective agreement or an individual occupational pension plan. Figure 1 displays the schematic off the different product types.

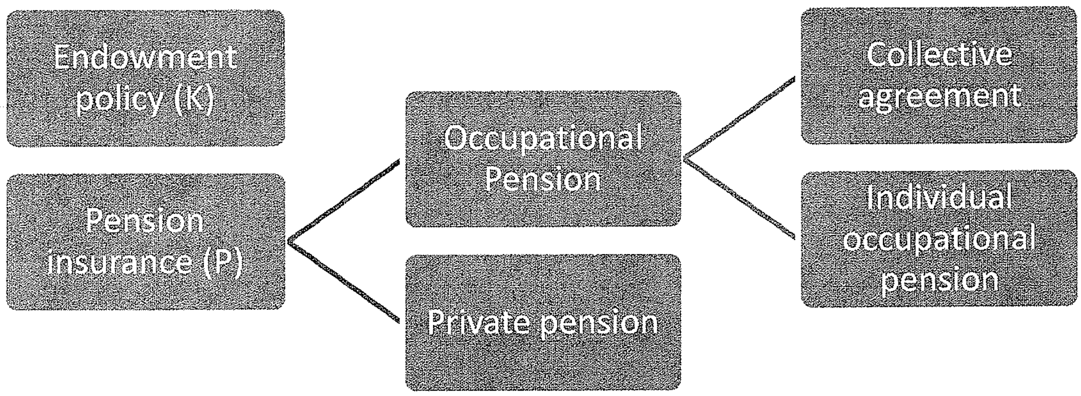


Figure 1 - Schematic presentation of products

In this analysis the contracts are divided as either an Endowment policy (K) or a Pension Insurance (P), the subcategories of P insurance will not be treated in a special way.

The term policyholder is used to describe the holder of a policy (contract). It is possible that a policyholder holds multiple contracts, for instance he can own a privately own endowment insurance and has an occupational pension plan from his employer. In this analysis such a person will be treated as two policyholders, because some attributes is bound to the policyholder (age, gender) are common over the two contracts he holds and some attributes are contract specific (contract type, value etc). To avoid confusion the word policyholder is bound to one policy, if he hold multiple there will be multiple policyholders that share some attributes.

Endowment insurance is a contract type that has special tax benefits. It can be viewed as a shell were you collect your investments and the tax paid is calculated from the value of the contract first of January each year times 27% of the government borrowing rate. It is a popular form of savings due to tax benefits, investors could decrease their tax on the capital gains by paying this standardized tax instead of capital gain tax of 30% if you were to hold the investments separately. As of 1 of January

2012 the tax rate was increased due to the release of a new savings form the Swedish government introduced, investment savings account (ISK).

On the Swedish market the investments that can be made within the insurances (K and P) is only mutual funds. So the policyholder chooses the investments in one or multiple funds that the company offer. In this thesis the categorization of different funds is not in focus and made strictly from their risk level and their management fee, as explained in section 4.1. Without any empirical prof, since this is outside the scope of this thesis, the following facts about the company fund range is stated.

- The funds with low management fee is typically fixed-income types
- The funds with low risk profile is typically fixed-income types

- The funds with high management fee is typically equity funds
- The funds with high risk profile are typically equity funds

There are some funds that this statements will not hold for, for instance an equity index fund has often a management fee that corresponds to the level for a fixed-income fund.

The revenues for the life-insurance company lies in a yearly fee on the contract and a kickback return from the mutual fund companies. The kickback return is typically a percentage on the management fee. It is a fixed percentage on the management fee that the life-insurance company negotiates with the mutual fund company. This kickback percentage is different for different funds and fund companies, it is however rather equal across the different funds. So typically if the customers hold equity funds instead of fixed-income funds the life-insurance company's future revenues are higher. Nevertheless if the customer holds more fixed-income type funds in case of a stock crash where the fixed-income type funds are not affected so much it withholds and maintains the value of the entire fund stock.

4 Analysis - Individual factors

In order to examine policyholders fund switching behavior, various dataset containing the possible variables were produced. In section 4.1 five different types of fund switches are explained. The explanatory variables that are believed to explain why policyholders switch funds, is described in section 4.2. In section 4.3 the available data for the analysis is explained. In section 4.4 the correlation between the fund switches and the explanatory variables is analyzed. Later the result from the VIF test and the results from the repeated cross sectional GLM are presented.

4.1 Fund switch categories

Hereafter the term fund switch will be used numerous of times throughout this thesis, and is used to explain when a policyholder switches all or part of her shares from one or multiple funds, into one or multiple other funds. Since the large amount of different funds, funds that share common attributes where grouped together in five different ways, this lead to the five different fund switch categories, which are explained in detail below.

Each of the fund switch categories examines a specific time interval. In the case of a "many to many" fund switch, only one of the ingoing funds have to match the ingoing criteria and one of the outgoing funds, has to match the outgoing criteria.

$FS1_i \equiv Policyholder\ i\ has\ switched\ to\ a\ fund\ with\ a\ lower\ management\ fee.$
$FS2_i \equiv Policyholder\ i\ has\ switched\ from\ a\ fund\ in\ category\ B, to\ a\ fund\ in\ category\ A.$

Where the categories, A and B are defined from their management fee as

$$Category = \begin{cases} A & if\ management\ fee \leq 1\% \\ B & if\ management\ fee > 1\% \end{cases}$$

$FS3_i \equiv Policyholder\ i\ has\ switched\ to\ a\ fund\ with\ a\ lower\ risk\ profile.$

All funds are given a risk level in the interval [1,7] where 1 represent the lowest possible risk and 7 represent the highest possible risk. This number is calculated continually from the lasts year historic volatility. Since the risk level is time variant but in the specific period regarded as time invariant this can cause trouble. However funds seldom changes risk level for the short periods investigated (1 -2 months).

$FS4_i \equiv Policyholder\ i\ has\ switched\ from\ a\ fund\ in\ category\ 2\ to\ a\ fund\ in\ category\ 1.$

Where the categories, 1 and 2 are defined from their risk level

$$Category = \begin{cases} 1 & if\ risk\ level\ is\ 1,2\ or\ 3 \\ 2 & if\ risk\ level\ is\ 4,5,6\ or\ 7 \end{cases}$$

$FS5_i \equiv Policholder\ i\ has\ switched\ \ from\ any\ kind\ of\ fund, to\ any\ kind\ of\ fund.$

The reason for this categorization lies in the nature of what is trying to be investigated. The first four categories describe more or less the same type of switches, while the last one is used to measure the activity of the policyholders.

A window of net fund flows where examined, the period examined was January to December in 2011. This period the Swedish stock market where rather volatile, mostly due to the national debt crisis that have grown in part of Europe in the aftermath of the financial crisis of 2008/2009. During that period the data at hand showed that a larger amount of inflow (both switches and new premiums) to fixed-income type funds and larger amount of outflow from more equity based funds. This is the behavior that is under investigation.

The behavior of the customer is believed to be triggered mostly by risk aversion, where to customer switches to a less risky fund, i.e. a fixed-income type fund. Since these kinds of funds have a lower risk level. However from the life-insurance company point of view the management fee categories lies more in their interests since that fee is directly correlated with the revenue stream from that contract. Since a great part of the revenues for the insurance companies lies in a kickback return percentage on the management fee. This is the reason for the categorization based on both risk level and management fee, and it is believed that the wider categories FS2 and FS4 will coincide. Since many of the funds in group A belongs to group 1 and vice versa.

The fifth category will be used only to look at the activity of the customer stock.

4.2 Explanatory variables & hypothesis formulation

Listed below are the explanatory variables chosen for the analysis, for each one of the explanatory variable a hypothesis about how that variable is believed to affect the switching behavior is formulated.

4.2.1 Value change

The change in value in percentage for the insurance contract is believed to affect the policyholders' behavior. It is assumed that if the value of the contract decreases the policyholder will be more likely to perform a switch of any kind, because as long as the insurance performance is positive or acceptable for the policyholder, he/she is assumed not to react in term of fund switching. However if the performance is not acceptable the policyholder is assumed to react on this fact, and switch funds.

Since we measure switches during a period the change in value prior to the period is believed to have an impact on the behavior. The source data contains fund information of every policyholder at the end of each month and weekly fund flows, payments, payouts switches etc. Observations of fund switches and payments that has occurred in a week that overlaps two consecutive months causes a problem, since it is not possible to distinguish in which month the action has occurred. Due to this problem it is hard to trace back the performance of the contract if payments are made in an overlapping week. For this reason the value change variable is calculated as follows: For each contract the fund-weights of each fund at the end of each month is calculated. If the policyholder has held the same portfolio (same weights in each fund) one and two months prior, even if this is not the case the value of that portfolio is calculated with the fund share prices. The value of the portfolio now is divided by the calculated value of the same weighted portfolio one and two months ago.

$ValueChange1_i \equiv Value\ change\ of\ all\ funds\ within\ contract\ i\ in\ percentage, from$
$one\ month\ prior\ the\ observation\ period.$

$ValueChange2_i \equiv Value\ change\ of\ all\ funds\ within\ contract\ i\ in\ percentage, from\ two\ months\ prior\ the\ observation\ period.$

Only one of the value change variables will be used in the final model, the one that fits the data best.

H1: The more the value of the contract decreases in percentage, policyholders will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders with a better performance in percentage.

4.2.2 Value

Policyholders with more amount of money within the insurance are believed to be more active in respect to switching between funds, than policyholders with less amount of money, because the more amount of money the policyholder holds within his or hers insurance it is assumed that the insurance contract is more important to the that individual. And if the contract is more important to the individual he or she will be more active in securing and growing the value of the contract, triggering more fund switches. This is mostly believed to be true for the extreme cases, e.g. a person with a contract value of 100 SEK is not believed to bother so much about the fund allocation as a person with a contract value of 1 000 000 SEK.

$Value_i \equiv Value\ of\ all\ funds\ within\ insurance\ i\ in\ monetary\ terms\ at\ the\ end\ of\ the\ period.$

H2: Policyholders with a higher contract value will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholder with a lower contract value.

4.2.3 Pension insurance

Policyholder with endowment insurance is assumed to be more active than other holders with pension insurance. Since the tax benefits, an endowment insurance suits investors which are more active and it is more of a shell that holds your investments rather than a typical insurance contract.

$$P_i \equiv \begin{cases} 1\ if\ policyholder's\ i\ contract\ type\ is\ P\ (Pension\ insurance). \\ 0\ if\ policyholder's\ i\ contract\ type\ is\ K\ (Endowment\ Insurance). \end{cases}$$

H3: Policyholders with an endowment insurance (K) will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders with a Pension insurance (P).

4.2.4 Male

The gender of the policyholder is believed not to affect the activity in terms of fund switching.

$$Male_i \equiv \begin{cases} 1\ if\ policyholder\ i\ is\ male. \\ 0\ if\ policyholder\ i\ is\ female. \end{cases}$$

H4: Policyholders switching behavior will be indifferent between genders.

4.2.5 External insurance broker

An insurance broker is an accommodator that handles the insurance contract for the customer. A customer is believed to assign a broker if he or she lack the will, information, knowledge or time to handle his or hers savings for the future. The policyholder can self-initiate a fund switch but it may also be initiated by the broker.

Policyholders that have connected an external insurance broker, they will be more "active" in terms of fund switching. Since it is believe that the policyholder seeks some kind of added value by connecting an insurance broker. And that this added value might be that the broker has to be more active with its client and that this may trigger more fund switches than for customers who lack a broker.

$$hasEB_i \equiv \begin{cases} 1 \text{ } if \text{ } policyholder \text{ } i \text{ } has \text{ } an \text{ } external \text{ } insurance \text{ } broker. \\ 0 \text{ } if \text{ } policyholder \text{ } i \text{ } does \text{ } not \text{ } have \text{ } an \text{ } external \text{ } insurance \text{ } broker. \end{cases}$$

H5: Policyholders with an external insurance broker will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders who does not have one.

4.2.6 Fund placement service

A fund placement service is a service available for the policyholders, if you chooses to have a fund placement service you choose to leave all fund choices in the hands of the service provider. All contracts that are connected to the same fund placement service share the same fund allocation. There are many different fund placement services available for the customers, first the insurance company typically offer a fund placement service where you might be able to choose from different model portfolios according to the level of your risk aversion. Then there exist numerous other of fund placement services that one can have, typically offered from a broker house.

Even if the fund switching behavior is left in the hands of the service provider, it is still an important behavior to choose to have such a service, and since we are trying to capture the behavior of the entire universe of contracts they are included in the analysis.

$$hasFPS_i \equiv \begin{cases} 1 \text{ } if \text{ } policyholder \text{ } i \text{ } has \text{ } a \text{ } fund \text{ } placement \text{ } service, i.e \text{ } the \text{ } policyholder \text{ } does \text{ } not \text{ } controll \text{ } the \text{ } portfolio. \\ 0 \text{ } if \text{ } policyholder \text{ } i \text{ } does \text{ } not \text{ } have \text{ } a \text{ } fund \text{ } placement \text{ } service. \end{cases}$$

H6: The probability that one observe a fund switch of type FS1, FS2, FS3, FS4 or FS5 is greater for the policyholders that is connected to a fund placement service.

4.2.7 Age

The age of the policyholder is believed to affect the fund switching behavior. It is believed that the older the policyholder becomes the more risk averse will the policyholder become, since the maturity date is either getting closer or if the contract is already under payment, the policyholder will not be likely to risk the value of the contract.

$$Age_i \equiv Age \text{ } of \text{ } policyholder \text{ } i.$$

H7: Older policyholders, will be more risk averse than their younger counterpart. The probability of an older policyholder will perform a switch of type FS1, FS2, FS3 or FS4 is greater than for their younger counterpart.

4.3 Data

Two dataset were constructed for the analysis. The data construction phase of the project was the most time consuming due to the large amount of data. First a brief presentation how the dataset were constructed, then an explanation about the data contained in the dataset is presented.

4.3.1 Dataset construction

The raw data at hand consisted of weekly fund flows, payment in and out of a specific fund and fund switches, stored in a SQL server. From the SQL server the data were extracted and inserted into a local SQL server for data manipulation. Most of the time where spent on building various SQL scripts that set up temporary tables in the database for the different variables, for instance the Fund switch categories and the value change variable had to be calculated backward for each policy in each period. These were then collected with the other more static variables as age etc. and inserted into a sample table. From this table the data were extracted into MATLAB for the statistical analysis.

The company mainly uses Microsoft Access to link in the data from the SQL server for data manipulation. Due to the file size limit of 2GB for an access database this software could not be used. The choice of software fell on the free version of Microsoft SQL Server 2008, the express edition, which allows for 10 GB of space in each database. However the entire dataset were larger than this limit, the data had to be manipulated chunk by chunk and then stored to local CSV raw data files and removed from the local database to free up space.

MATLAB also encountered problems due to the large sample size. The estimation routine encountered out of memory error, because the physical memory limit on the 32 bits operating system used by the company. This was overcome with a 64 bit version of MATLAB on another machine running a 64 bit operating system.

4.3.2 Dataset

Two dataset were constructed one with 12 periods (each month) and one with 6 periods (2 months group together), two value change variables were constructed for each contract, later which one is better to use will be evaluated. The Value change variables for each contract in each period are calculated for both one and two months backwards in time. Se figure 2.

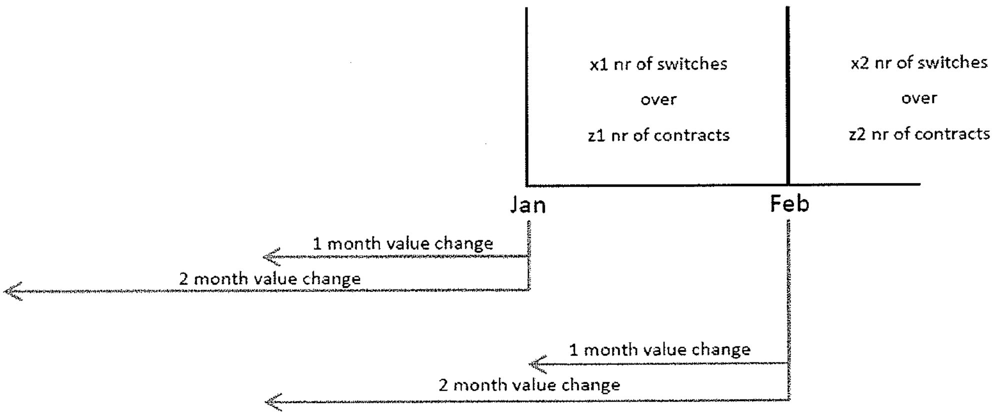


Figure 2 - Graphical representation of the first dataset

For each of the twelve time points, information regarding what kinds of fund switch that has occurred, if any, for each of the contract is stored. In addition to this, the explanatory variables for each contract at every time point are stored.

The second dataset consist of fund switching data over 2 consecutive months starting in Jan/Feb 2011 to Nov/Dec 2011 for a total of 6 time points.

The number of contracts available for each time point varies, since there are several actions that may occur, for instance death, annulment of the contract, surrender of the contract and transfer of the contract. There are also contracts that suffer from missing data in the value change calculations, since some historic fund share prices are missing. These are omitted from the sample in that period. Statistics on the first dataset are given in table 1 and 2. An "active" policy means that a switch of any kind (FS5) has occurred during the period.

	Jan	Feb	Mar	Apr	May	Jun
Number of policies	792 994	797 928	971 008	972 373	971 013	977 584
Number of active policies	117 084	99 983	137 338	117 520	111 905	129 785
Number of policies with a fund placement service	86 090	87 451	110 113	110 968	109 984	110 964
Proportion of active policies	14,76%	12,53%	14,14%	12,09%	11,52%	13,28%

Table 1 - Statistics on data (Jan-Jun)

	Jul	Aug	Sep	Okt	Nov	Dec
Number of policies	979 206	979 783	980 712	929 310	890 202	907 666
Number of active policies	93 385	205 471	158 354	134 183	86 265	55 925
Number of policies with a fund placement service	111 123	110 856	111 051	85 942	47 145	49 832
Proportion of active policies	9,54%	20,97%	16,15%	14,44%	9,69%	6,16%

Table 2 - Statistics on data (Jul-Dec)

4.4 Correlation analysis

The correlation between the fund switching variables and the explanatory variables are of interest, if the correlation between fund switching and the explanatory variables are stable over time there might be a connection. Later in this section the correlation between the explanatory variables themselves are analyzed.

We start by examine the correlation between the explanatory variables and the fifth fund switching category, which is any kind of fund switches, i.e. we examine the activity of the policyholders.

Variable / Month	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Okt**	**Nov**	**Dec**
ValueChange1	0,13	-0,14	0,15	-0,13	-0,14	0,16	0,12	-0,01	0,04	0,13	0,05	-0,10
ValueChange2	0,11	0,14	0,07	-0,06	-0,09	0,04	0,18	0,07	0,07	0,12	-0,10	0,01
Value	0,10	0,11	0,09	0,06	0,09	0,08	0,05	0,03	0,08	0,07	0,06	0,06
P	-0,11	-0,08	-0,06	-0,07	-0,10	-0,04	-0,07	0,02	-0,04	-0,03	-0,07	-0,11
Male	0,04	0,05	0,05	0,03	0,03	0,05	0,03	0,06	0,05	0,05	0,04	0,03
hasEB	0,24	0,32	0,25	0,14	0,18	0,26	0,20	0,18	0,22	0,27	0,18	0,20
hasFPS	0,61	0,81	0,66	0,60	0,64	0,67	0,53	0,52	0,73	0,74	0,69	0,70
Age	0,02	0,02	0,00	0,01	0,04	-0,01	0,05	-0,02	0,02	-0,01	0,02	0,03

Table 3 - Correlation between FS5 and Explanatory variables

Without analyzing the significance of the correlation, as this will be done later on, some weak correlation analysis follows. The variables that show signs of a positive correlation with FS5 over time is

- Value
- Male
- hasEB
- hasFPS

Policyholder that has a fund placement service, has an external insurance broker seems to be more "active" which make sense and was also expected. Policyholder with greater insurance value will also be more active and the gender dummy variable for male is slightly positive correlated over time with fund switches, however this correlation is very small and may also be insignificant.

And the variables that show signs of negative correlation over time

- P (11 out of 12 months)

Consequently policyholders with an endowment insurance (P=0) contract are more active in 11 months of 12 than policyholder of a pension insurance (P=1) contract.

If this result is compared with the correlation for the fund switch category FS4 and the explanatory variables, the age variable is positively correlated with FS4 over the twelve time points. This make sense since we believe that the older the policyholder the more risk averse he/she will become. See table 4.

Variable/Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Okt	Nov	Dec
ValueChange1	0,01	-0,15	0,14	-0,11	-0,16	0,15	0,11	0,16	0,20	0,13	0,01	-0,09
ValueChange2	0,03	0,08	0,09	-0,05	-0,10	0,04	0,17	0,16	0,22	0,20	-0,16	0,02
Value	0,08	0,09	0,09	0,06	0,07	0,09	0,04	0,11	0,10	0,07	0,07	0,04
P	-0,03	-0,08	-0,09	-0,11	-0,12	-0,07	-0,08	-0,09	-0,09	-0,10	-0,12	-0,12
Male	0,01	0,03	0,02	0,01	0,01	0,04	0,02	0,04	0,03	0,02	0,03	0,02
hasEB	0,10	0,30	0,25	0,17	0,20	0,32	0,19	0,30	0,26	0,23	0,24	0,22
hasFPS	0,23	0,73	0,64	0,65	0,59	0,77	0,51	0,74	0,82	0,60	0,85	0,77
Age	0,03	0,04	0,04	0,07	0,06	0,02	0,06	0,03	0,04	0,04	0,06	0,03

Table 4 - Correlation between FS4 and Explanatory variables

The correlation between the explanatory variables themself was investigated. As described in section 2.1.4, there is a problem if the variables are collinear. Most of the seven variables show a very low correlation to the other variables. The correlation between two of the variables might raise a concern:

- Corr(Value,Age)
- Corr(hasFPS,hasEB)

There exists a positive correlation between the value of the contract and the age of the policyholder. There is also a positive correlation between the contracts that are connected to a fund placement service and the contracts that has an external insurance broker. This is expected since most of the broker houses offer their own fund placement service to their customers. And an older policyholder has on average more money within their insurance than a younger policyholder. The correlation coefficient over the twelve time points lies in the interval of:

$$0 < Corr(Value, Age)_i < 0.2 \; \forall i, i = 1 \ldots 12$$
$$0 < Corr(hasFPS, hasEB)_i < 0.4 \; \forall i, i = 1 \ldots 12$$

In order to determine if the correlation between two explanatory variables is too strong for a regression analysis, a variance inflation factor test was performed.

4.4.1 Variance inflation factor analysis

The VIF test is describes in section 2.1.4 and is used to determine if one or two variables is collinear. Typically one uses a VIF of 5 to say that the variables are collinear. This means:

$$VIF = \frac{1}{1 - R^2} = 5 => R^2 = \frac{4}{5} = 0.8$$

If the R-square value of one the k regressions is above 0.8 it can be concluded that there is a too strong relationship between two or more of the regressors so that one or more of them has to be omitted.

In this test the periods that showed the highest correlation coefficients between the variables explained in the previous section was sought out and the VIF-test were performed on the sample from that period. The results from the test conclude that all the seven variables should be included in the main regression since the largest VIF observed was around 1.2.

4.5 Repeated Cross sectional models for switching behavior

The cross sectional fund switching was modeled as described in the theory section, as a generalized linear model. Three different link functions were used, Logistic-, Normal- and Uniformed distribution.

The logistic regression model is given as

$$P(switch)_i = \frac{e^{FS_i}}{1+e^{FS_i}} \tag{4.1}$$

While the Probit model is given as

$$P(switch)_i = \int_{-\infty}^{FS_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\} dt \tag{4.2}$$

And the linear probability model

$$P(switch)_i = FS_i \tag{4.3}$$

Where FS_i is given as a linear combination of the explanatory variables and a residual term.

$$FS_i = \beta_0 + \beta_1 ValueChange_i + \beta_2 Value_i + \beta_3 P_i + \beta_4 Male_i + \beta_5 hasEB_i + \beta_6 hasFPS_i + \beta_7 Age_i + \varepsilon_i$$

Which model that is preferred depends of the data generation process (dgp), we choose the model which we believes approximate the dpg sufficiently. The linear probability model has some drawbacks presented in section 2.1, however if the regressors show to be endogenous this method for parameter inference can still be useful (Cameron and Trivedi 2005, page 473)

A test to determine whether the explanatory variables could be considered exogenous or not (endogenous) was performed. If the variables show signs of endogeneity there are methods to overcome this in the linear probability model. The main reason for endogeneity, is the problem with omitted and/or unobserved variables, i.e. one or more significant variables have been left out from the model. However it was shown that we could consider the following to hold.

$$cov(X, \varepsilon) = 0 \; for \; all \; time \; periods$$

Where X is a Nx7 matrix containing the 7 explanatory variables and ε is a Nx1 vector of the residuals. This was shown to be true for all time periods. Since there were no obvious signs of endogeneity it was concluded that the Probit or Logit model should be used.

Therefore, which model that fits the data the best was investigated based on the maximum likelihood estimate, but as discussed by Cameron and Trivedi (2005), they conclude that the choice of link function is not that crucial if certain conditions upon the regressors hold.

"Misspecification in this matter is not as great if the regressors are distributed such that the mean of each regressor, conditional on the linear combination $x_i'\beta$ is linear in $x_i'\beta$, then choosing the wrong function F, can be shown to affect all slope parameters equally so that the ration of slope parameters is constant across all models" (Cameron and Trivedi 2005, page 472)

Even though the fact that the Probit and Logit model typically yields the same result (Verbeek 2008, page 201) both models were estimated for all possible combinations of the data (different value change variables and the length of the switching period).

The two models 4.1-4.2 was estimated for each of the time period for the two datasets and for both the FS5 and FS4 kind of fund switches as the dependent variable, since the other fund switch categories was shown to coincide with the FS4. There is little difference between the results (sign and significance of the estimates) from the estimation for the two models, see appendix 1. Looking at the 12 month dataset and denoting $maxl_t^{logit}$ and $maxl_t^{probit}$ as the maximum likelihood estimate for the Logit and Probit model respectively it was shown that

$$maxl_t^{logit} > maxl_t^{probit}, \forall t, t = 1 \dots 12.$$

Based on this it was concluded that the Logit model fits the data better, the pseudo R2 values aggress with this. Further diagnostics and hypothesis testing will thus be based on the logistic regression model.

4.5.1 Logistic regression analysis

Let us start by analyzing the estimation result for the FS5 variable, the parameter estimates are given in appendix 1. Significance test were performed for each of the variable in each period. Table 5 shows which variables that is statistical significant at the 99% level, over all periods for the FS5 category.

DV: FS5 Variable	Number of significant periods	Number of periods that the parameter is <u>positive</u> and significant	Number of periods that the parameter is <u>negative</u> and significant
Constant	12 (12)	6	6
ValueChange1	12 (12)	5	7
Value	11 (12)	11	0
P	12 (12)	7	5
Male	12 (12)	12	0
hasEB	11 (12)	7	4
hasFPS	12 (12)	12	0
Age	10 (12)	5	5

Table 5 - Parameter stability, FS5, Logit

When looking at the activity of the policyholders, no conclusions about the value change or age variable can be drawn. The value change variable is significant in all the periods but due to the change in sign in each period, no conclusion about the performance of the individual contract and activity in fund switching can be drawn. Table 5 shows that the parameters that correspond to the Value, Male and hasFPS variables are positive and rather stable, se appendix 1, over time. The parameter that corresponds to P and hasEB are positive in more periods than they are negative.

DV: FS4 Variable	Number of significant periods	Number of periods that the parameter is <u>positive</u> and significant	Number of periods that the parameter is <u>negative</u> and significant
Constant	12 (12)	5	7
ValueChange1	12 (12)	7	5
Value	12 (12)	12	0
P	12 (12)	0	12
Male	11 (12)	7	4
hasEB	11 (12)	10	1
hasFPS	12 (12)	12	0
Age	10 (12)	9	1

Table 6 - Parameter stability, FS4, Logit

The difference when looking at the FS4 category, see table 6, is that the P parameter now is negative and significant in all twelve periods. There is also a difference corresponding to the Age and hasEB variable. Older and brokered policyholder seems to be more risk averse than their younger and un-brokered counterpart. The likeliness of switch for policyholder that holds an endowment policy (P=0) seem to be greater than policyholder that holds a pension insurance (P=1).

4.5.2 Model diagnostics

To evaluate the fit of the models, the periods that showed the highest and the lowest pseudo r2 value were examined and a cross tabulation of actual and predicted outcomes were made. The difference in fit for the two value change variables was also under investigation.

Let start by looking at the model with FS5 as the dependent variable. The periods under investigation is chosen to be Aug and Feb. The numbers within parenthesis is from the regression with the valueChange2 variable instead of valueChange1.

DV: FS5 Month: Aug McfaddenR2: 0.22 (0.216)		**Predicted**		**Percentage of right prediction**
		0	1	
Switch	0	*751 832 (751 841)*	*22 480 (22 471)*	*97.10% (97.10%)*
	1	*117 099 (117 109)*	*88 372 (88 362)*	*43.01% (43.00%)*
			Total prediction success	**85.75% (85.75%)**

Table 7 - Cross tabulation, FS5, Lowest R2

In the period with the lowest Mcfadden R2, see table 7, the model is good to predict non-switches but seems to miss a lot of switches, only 43% of the them are captured. Overall the estimated logistic model in this period predicts rights in around 86% of the cases.

DV: FS5 Month: Feb McfaddenR2: 0.585		Predicted		Percentage of right prediction
		0	1	
Switch	0	*687 987* *(688 021)*	*9 958* *(9 924)*	*98.57%* *(98.58%)*
	1	*22 209* *(22 214)*	*77 774* *(77 769)*	*77.79%* *(77.78%)*
			Total prediction success	**95.97%** **(95.97%)**

Table 8 - Cross tabulation, FS5, Highest R2

In the period with the highest Mcfadden R2, see table 8, the model predicts the non-switches good and has a prediction success of switches at 77% which is acceptable. The overall prediction success is 95%.

Looking at the FS4 category, i.e. the switch from a riskier fund group to a less risky fund group, the period with the lowest Mcfadden R2, see table 9, is Jan and the period with the highest R2 value is Nov, see table 10.

DV: FS4 Month: Jan McfaddenR2: 0.20 (0.216)		Predicted		Percentage of right prediction
		0	1	
Switch	0	*782 629* *(782 634)*	*108* *(103)*	*99.99%* *(99.99%)*
	1	*13 263* *(13 282)*	*45* *(26)*	*0.003%* *(0.002%)*
			Total prediction success	**98.32%** **(98.32%)**

Table 9 - Cross tabulation, FS4, Lowest R2

The model is in the period with the lowest R2 value is insufficient at predicting switches, some explanation lies in the fact that the months Jan and Feb has lots of missing data so approximately 200 000 policies are omitted in these two months. This will result in only about 13 000 fund switches of FS4 has occurred, which corresponds to about 1.5% of the policyholders perform a switch. There is a problem with the estimation if the dependent variable seldom changes value, since the estimates are the estimates that fit the entire data best. In this case the parameter estimates fits the non-switches best because of their domination in this month's sample. This is however a problem in all periods, since the number of non-switches are greater than the number of switches, the best case scenario would have been if the number of switches and number of non-switches would be rather equal for the estimation accuracy point of view. This is one of the reasons why there exists a second data set that stores switches made over 2 months. However, looking at the prediction success for that dataset no improvement in the hit success was shown.

In the month with the second lowest R2 value the success prediction rate for switches is about 35% instead of 0.002%, in this period there are not so many omitted polices.

DV: FS4 Month: Nov McfaddenR2: 0.67 (0.67)		Predicted		Percentage of right prediction
		0	1	
Switch	0	*832 028 (832 028)*	*3 361 (3 361)*	*99.60% (99.60%)*
	1	*11 012 (11 011)*	*43 801 (43 802)*	*79.91% (79.91%)*
			Total prediction success	**98.39% (98.39%)**

Table 10 - Cross tabulation, FS4, Highest R2

In the period with the highest Mcfadden R2 value, see table 10, the model is good to predict non-switches and has a prediction success of switches at almost 80% which is acceptable, and the overall prediction success is also very high.

From this we can conclude that the logistic regression model fits the overall data well, however just at predicting fund switches it is a rather insufficient model in some periods. There is a slight increase in prediction success of switches using the valuechange1 variable instead of valuechange2, hereafter the valuechange1 variable will thus be used.

As stated in section 1.2 a model of switching behavior is sought, and even though the result is not as good as hoped, we continue with caution and conclude that the parameters estimate are stable enough to find an average model of all periods.

4.6 Average model for switching behavior

In this section the 12 different sets of data are grouped together in one large dataset in order to find a single estimate for each of the explanatory variables. The previous estimations did not show the stability of the parameters as hoped. The parameter estimate for the value change parameter did change sign from period to period which was concerning, since we do believe in the fact that policyholders will change to less riskier fund when the value of the contract goes down. In this section the entire 12 months are evaluated. The hope is to find something that explains the average behavior of the policyholder with respect to the performance of their contract.

One of the fundamental rules of regression analysis is that the observations is assumed to be independent, here this is stretched as we are now grouping the approximately 950 000 polices for each month in one matrix over all periods, in order to do a cross sectional estimation over approximately 12*950 000 policies. Since the same policyholder is observed (in most cases) 12 times in the set, the observations cannot be assumed to be independent since it is the same person observed 12 times.

However the value change and value variable has most certainly different values. There are also possible that the policyholder has changed his status in respect to a fund placement service and broker. This combined with the fact the very large amount of different policyholders it is decided that this is acceptable to do in order to find a model that displays the average behavior.

A new fund switch category is introduced in order to examine the difference between different fund switches. The categories FS1-FS4 are more or less the same switch as concluded earlier. Since FS4 is chosen the opposite switch is introduced, it is the complement to the FS4; namely FS4C. FS4C is the switch from a fund within the lower risk class to a fund within the higher risk class, for instance a fixed-income fund to a global equity fund.

Once again the following logistic model was estimated

$$P(FSx) = \frac{e^{(\widehat{\beta_0}+\widehat{\beta_1}ValueChange1+\widehat{\beta_2}Value+\widehat{\beta_3}P+\widehat{\beta_4}Male+\widehat{\beta_5}hasEB+\widehat{\beta_6}hasFPS+\widehat{\beta_7}Age)}}{1+e^{(\widehat{\beta_0}+\widehat{\beta_1}ValueChange1+\widehat{\beta_2}Value+\widehat{\beta_3}P+\widehat{\beta_4}Male+\widehat{\beta_5}hasEB+\widehat{\beta_6}hasFPS+\widehat{\beta_7}Age)}} \quad (4.0)$$

The estimation result can be found in table 11.

Fund switch	**FS4**	**FS4C**	**FS5**
$\widehat{\beta_0}$ (Constant)	3.1987	-9.3345	4.0527
$\widehat{\beta_1}$ (ValueChange1)	-8.1133	4.3275	-6.9856
$\widehat{\beta_2}$ (Value)	1.8455e-7	1.0275e-7	1.9708e-7
$\widehat{\beta_3}$ (P)	-0.5217	-0.7330	-0.0581
$\widehat{\beta_4}$ (Male)	0.0357	0.0136	0.1871
$\widehat{\beta_5}$ (hasEB)	0.5886	0.5375	0.1361
$\widehat{\beta_6}$ (hasFPS)	4.3809	4.5674	4.0217
$\widehat{\beta_7}$ (Age)	0.0137	0.0121	-0.0027

Table 11 - Parameter estimates from the average model

A t-test concluded that each variable was significant at 99% confidence level.

The same diagnostics as in the multiple cross sectional analysis were made, and the following prediction success ratios were obtained.

Fund switch	**FS4**	**FS4C**	**FS5**
Prediction success non-switch (FSx=0)	96.62%	97.72%	97.29%
Prediction success switch (FSx=1)	73.53%	53.58%	61.65%
Total prediction success	94.98%	95.04%	92.79%

Table 12 - Average model prediction success ratios

From table 12 it can be concluded that the model prediction success is acceptable for FS4 and FS5. For the FS4C switch only around half of the switches are captured. The overall prediction success is however at a high level and the model is acceptable.

4.7 Hypothesis testing

The different explanatory variables are analyzed separately and the corresponding hypothesis is either rejected or accepted. Due to the large amount of models (various dataset and different value change variables) the hypothesis testing will be performed on the basis of the average model at first with support from the 12 cross sectional regressions, and the correlation analysis. As mentioned earlier switches of type FS1, FS2 and FS3 was shown to coincide with the wider FS4 category. Each of the explanatory variables will be investigated and the corresponding hypothesis formulated in section 4.2 will be tested.

4.7.1 Value change

The hypothesis formulated in section 4.2.1 was:

H1: The more the value of the contract decreases in percentage, policyholders will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders with a better performance in percentage.

As discussed in the cross section analysis the parameter estimate changes sign from period to period. The average model shows a negative significant estimate for the FS4 and FS5 switch which is expected and a positive significant estimate for switches of type FS4C. No conclusion about the variable can be drawn on the monthly basis. One month might be too short of an interval and another explanation lies in the fact that most of the policyholders that perform a switch also have a fund placement service, so the value performance of the different fund placement services versus the average performance of the other policies will contribute to the sign of the estimate. For instance if the large fund placement services has a greater value performance than the average performance of the rest of the policies the parameter estimate would become positive and if the value performance of the large fund placement services are less than the average performance of the rest, the parameter estimate would typically become negative.

On this basis we conclude that the more data from more periods the better to say something about the average behavior of the policyholders, and the parameter estimate is significant and negative for FS4 and FS5 from the average model, so it is concluded that the hypothesis holds on average. For the other switch FS4C even if it is not investigated in the hypothesis the opposite relation holds i.e. the parameter estimate is positive for the FS4C switch which is also expected. However the model works poorly predicting switches of kind FS4C. Table 13 shows the relation between the change in value change and probability of switch for the different switching categories

Switches	FS4	FS4C	FS5
If Value change *increase*	The probability of this switch *decrease*	The probability of this switch *increase*	The probability of this switch *decrease*

Table 13 - Value change, average behavior

The hypothesis is accepted on average.

4.7.2 Value

The hypothesis formulated in section 4.2.2 was:

H2: Policyholders with a higher contract value will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholder with a lower contract value.

The value parameter for the FS4 and FS5 regressions is significant and positive in most periods (except for one) on the monthly basis. The parameter estimates are also positive and significant in the average model. The estimate is very small in scale due to the fact that the large variation in value within the insurance between policyholders. So the hypothesis holds both on monthly basis and on average. Table 14 shows the relation between change in value and probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If Value *increase*	The probability of this switch *increase*	The probability of this switch *increase*	The probability of this switch *increase*

Table 14 – Value, average behavior

The hypothesis is accepted.

4.7.3 Pension insurance

The hypothesis formulated in section 4.2.3 was:

H3: Policyholders with an endowment insurance (K) will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders with an Pension insurance (P)

The parameter estimate changes sign from period to period for switches of any kind, FS5 on a monthly basis. However it seems to hold for FS4 types of switch on a monthly basis. From the average model we conclude that the hypothesis holds on average. Table 15 shows the relation between the P variable and probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If P *is true (1)*	The probability of this switch *decrease*	The probability of this switch *decrease*	The probability of this switch *decrease*

Table 15 - P, average behavior

The hypothesis is accepted on average.

4.7.4 Male

The hypothesis formulated in section 4.2.4 was:

H4: Policyholders switching behavior will be indifferent between genders.

The parameter estimate is positive and significant for all periods on a monthly basis for FS5 types of switches. The average model estimate agrees with this fact. It can be concluded that males are more likely to switch funds than females. Table 16 shows the relation between the Male variable and probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If Male *is true (1)*	The probability of this switch *increase*	The probability of this switch *increase*	The probability of this switch *increase*

Table 16 - Male, Average behavior

The hypothesis is rejected.

4.7.5 External insurance broker

The hypothesis formulated in section 4.2.5 was:

H5: Policyholders with an external insurance broker will be more likely to perform a switch of type FS1, FS2, FS3, FS4 or FS5 than policyholders who does not have one.

The parameter estimate changes sign from period to period on a monthly basis, it is however more often positive than negative, i.e. the likeliness of switch if you have an insurance broker is greater in more periods for FS4 and FS5 types of switch. On an average level this holds, we conclude that the hypothesis holds on average. Table 17 shows the relation between the hasEB variable and probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If hasEB *is true (1)*	The probability of this switch *increase*	The probability of this switch *increase*	The probability of this switch *increase*

Table 17 - hasEB, Average behavior

The hypothesis is accepted on average.

4.7.6 Fund placement service

The hypothesis formulated in section 4.2.6 was:

H6: The probability that one observe a fund switch of type FS1, FS2, FS3, FS4 or FS5 is greater for the policyholders that is connected to a fund placement service.

The parameter estimate is positive and significant in all periods on a monthly basis, the average model aggress with this, for switches of type FS5. The same is true for FS4 type of switches. It is concluded that the hypothesis holds. Table 18 shows the relation between the hasFPS variable the probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If hasFPS *is true (1)*	The probability of this switch *increase*	The probability of this switch *increase*	The probability of this switch *increase*

Table 18 - hasFPS, Average behavior

The hypothesis is accepted.

4.7.7 Age of the policyholder

The hypothesis formulated in section 4.2.7 was:

H7: Older policyholders, will be more risk averse than their younger counterpart. The probability of an older policyholder will perform a switch of type FS1, FS2, FS3 or FS4 is greater than for their younger counterpart.

The parameter estimate changes sign from period to period on a monthly basis for FS5 switch, however this is not the switches under investigation here. On a monthly basis the age variable is significant in 10 out of 12 periods and positive in 9 out of those 10 periods. The average model agrees with the fact that older policyholders are more likely to perform a FS4 switch. The hypothesis is concluded to hold on average. Older policyholders are more likely to perform a switch of FS4 and FS4C, but on average they are less active in terms of fund switching (FS5) than their younger counterparts. Table 19 shows the relation between the age variable and the probability of switch for the different categories.

Switches	FS4	FS4C	FS5
If Age *increase*	The probability of this switch *increase*	The probability of this switch *increase*	The probability of this switch *decrease*

Table 19 - Age, Average behavior

The hypothesis is accepted on average.

4.8 Result

The result from the hypothesis testing showed that six out of seven hypotheses holds, at least on average. How much the probability of the different switches changes for the explanatory variables is presented in this section. Here the average model for the entire set is in the only one in focus.

Due to the fact that the parameter estimates from a logistic regression are little more complex to interpret than in a normal regression the following example shows the impact on switches by having a fund placement service.

The parameter estimates for the average model are given in table 11. Recall the theory section that the parameter estimate in a logistic regression can be viewed as the log odds. For instance if we examine the estimate of the coefficient for hasFPS for any kind of switch, FS5.

$$\widehat{\beta_6} = 4.0217$$

A one unit increase in hasFPS results in a 4.0217 change in the log of the odds. The odds ratio can then be calculated by talking e to the power of the coefficient estimate.

$$OR = e^{\widehat{\beta_6}} = e^{4.0217} = 55.7959$$

By having a fund placement service the *odds* of making a FS5 switch increase by 5479.59% compared with not having one, if all other variables were the same (cateris paribus). I.e. the odds for performing a switch without having a fund placement service active, multiplied by the odds ratio (55.7959) gives the odds for performing a switch with a fund placement service. Table 20 shows the odds ratio of all explanatory variables

Explanatory variable	Odds ratio FS4	Odds ratio FS4C	Odds ratio FS5
Constant	24.5005	0.0001	57.5508
Value	1.0000	1.0000	1.0000
P	0.5935	0.4805	0.9435
Male	1.0363	1.0137	1.2057
hasEB	1.8016	1.7117	1.1457
hasFPS	79.9112	96.2907	55.7946
Age	1.0138	1.0122	0.9973
Valuechange	0.0003	75.7579	0.0009

Table 20 - Odds ratios

For simplicity it is more convenient to present the result with probabilities instead of odds ratios. Therefore the "average policyholder" is used to say how the probability of switches changes while one of the explanatory variables changes. The average policyholder has the attributes according to table 21.

Explanatory variable	Value
Value	X SEK
P	1
Male	1
hasEB	0
hasFPS	0
Age	49.5
Valuechange	1.0

Table 21 - Average policyholder attributes

Hence the average person is a male with X amount of money within his pension insurance. He is 49.5 years old and do not have a fund placement service nor an external insurance broker. The value change during the last month is assumed to be 1, i.e. no change at all.

First the effect on the binary attributes are investigated then the effect on the continuous variables value, age and value change is investigated. The marginal effect on every variable is examined (cateris paribus).

Binary variable		Is false (=0)	Is true (=1)	Change (%) in probability from false to true
P	Probability Switch FS4	0.015	0.009	-40%
	Probability Switch FS4C	0.012	0.006	-50%
	Probability Switch FS5	0.054	0.0514	-4.81%
Male	Probability Switch FS4	0.0088	0.0090	2.27%
	Probability Switch FS4C	0.0059	0.0060	1.69%
	Probability Switch FS5	0.0449	0.0514	14.48%
hasEB	Probability Switch FS4	0.0090	0.0161	78.89%
	Probability Switch FS4C	0.0060	0.0102	70%
	Probability Switch FS5	0.0514	0.0584	13.62%
hasFPS	Probability Switch FS4	0.0090	0.4205	4572.22%
	Probability Switch FS4C	0.0060	0.3667	6011.67%
	Probability Switch FS5	0.0542	0.7514	1286.35%

Table 22 - Probabilities of different switches

From table 22 one can conclude that there is a massive impact on the probability of fund switches if the policyholder is bound to a fund placement service or not, for instance the probability of any switch for the average policyholder is 0.0542 (5.42%) while a policyholder with the same attributes, expect that he has a fund placement service active, is 0.7514 (75.14%). I.e. an increase in the probability of FS5 switches with over 1000%.

For the continuous variables table 23 displays the change in each of the variable (cateris paribus), the increase is not set to one unit, due to minimal effect a one unit increase in for instance value and age will have.

Continuous variable		Average (x SEK)	Increase of 10 000 SEK	Change (%) in probability from Average to increase
Value	Probability Switch FS4	0.0090	0.0090	0%
	Probability Switch FS4C	0.0060	0.0060	0%
	Probability Switch FS5	0.0514	0.0515	0.19%
		Average (49.5 years old)	**Increase of 10 years**	**Change in probability %**
Age	Probability Switch FS4	0.0090	0.0103	14.44%
	Probability Switch FS4C	0.0060	0.0067	11.67%
	Probability Switch FS5	0.0514	0.0501	-2.53%
		Average (No change past month)	**Increase of 1% in value past month**	**Change in probability %**
Valuechange	Probability Switch FS4	0.0090	0.0083	-7.78%
	Probability Switch FS4C	0.0060	0.0062	3.33%
	Probability Switch FS5	0.0514	0.0481	-6.42%

Table 23 - Probabilities of different switches

The change in probability on the marginal for the average policyholder is thus given. The fund placement service is by far the most significant variable.

By plotting the value change variable against the probability of switching for the average policyholder with and without a fund placement service figure 3-5 is given. The red line in figure 5 is thus the probability of an average policyholder perform a switch of type FS5 (any kind of switch), the probability line is downward sloping, but however not very steep. The value change of a contract during one month is seldom outside the interval +/- 10%. By widening the value change to range from 0.1 to 1.9, i.e. -90% to +90% change in the value of the contract in one month, one can see that the value change has to be approx. 0.6 i.e. a decrease of forty percent in the value of the contract in one month to reach a 50-50 chance of a fund switch. This is however very unlikely and no such drastic value change during one month is observed out of the approx. 11 million observations.

The probability of an average policyholder of performing a switch of FS4 is very small. The slope of the red curve in figure 3 is downward sloping as expected but even if the contract value decreases with 20% the probability of switching is still very low.

The probability of an average policyholder of performing a switch of FS4C is very small. The slope of the red curve in figure 4 is barley upward sloping, the value change of the contract does only affect the probability of switching at a very low level.

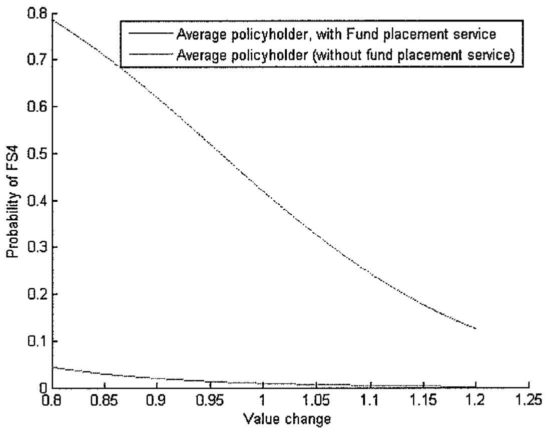


Figure 3 – Probability of FS4

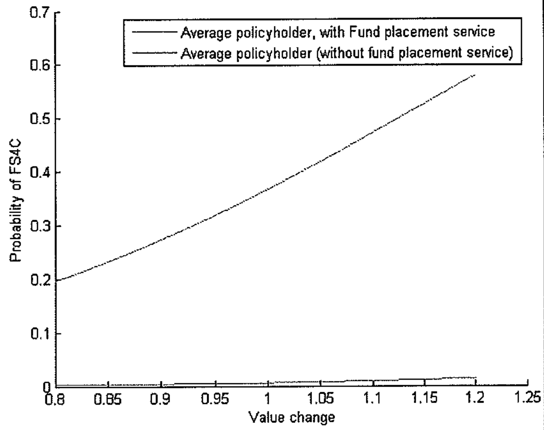


Figure 4 – Probability of FS4C

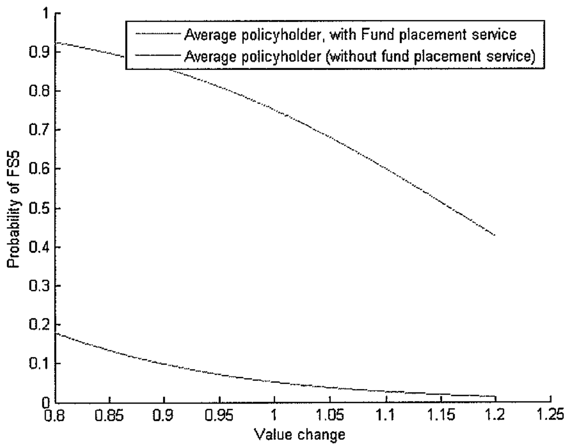


Figure 5 – Probability of FS5

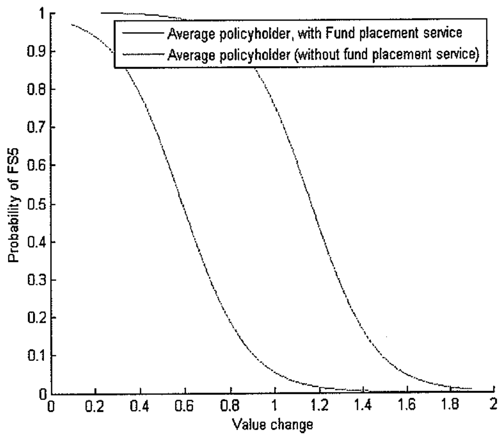


Figure 6 – Probability of FS5 (wider x-axsis)

From the result above the conclusion about the probability of switches has do to most with whether the policy is bound to a fund placement service or not. The impact of the value change is less than expected, especially for the risk averse switch FS4. Even by looking at the longer value change variable no difference where found.

The goal was to find a model that did work well and then stress the value change variable in different directions to see how the probability of different switches were effected, from this estimations of future revenues could in best case be done.

On average we see the behavior which is expected - if the value decreases the policyholders will react and perform a switch. However due to the change in sign in different periods on a monthly basis and that the model fails to predicts a part of the switches and with the fact that there seems to be an increase of switches when the stock market goes down, see table 2. In august 2011 there were a spike in the number of switches of FS5, during that period the return from the OMX Stockholm benchmark Index (Tr) was around -12%, which is large fall in one month.

This raises the question that it might not be as much the *individual* performance of the contracts that trigger switches, instead some sort of unobserved variable that explains the current macro-economic status. In the year investigated there were a lot of stress and uncertainty in the market due to the rising debt crisis in some European countries. It is possible that even if the policyholder's funds have performed well, the macro-economic variable might have gone down and this is what triggered switches of some sort. This in accordance with the goal stated in the section 1, that a model predicting how the revenue stream in the future is affected by fund switching, a model that only takes in non-individual factors is introduced in section 5.

From this section we conclude that the individual factors affect the average switching behavior as in table 13 – 19 and the corresponding odds ratios from the average model, see table 20. However as shown above all variables are significant at explaining the probability of switches, but when a fund placement service is not active the absolute change in probability of switches by being a couple of years older, having a broker etc. is small.

5 Analysis – Impact on future revenues

In this section the data and the model for change in revenue due to fund switches is explained. Later, as diagnostics for the model, information about the fund allocation of the entire fund stock will be used. There are some clear distinctions made in this analysis from the individual factor analysis. In the previous section there was no distinction between policyholders that did one or multiple fund switches. Here all fund switches made are under investigation. There is no longer need for various fund switch categories. Here the effect of the switch is taken from the funds that are involved in the switch, therefore no need for a categorization of funds. As described in section 3, the revenue for the life-insurance company partly lies in a kickback return from the mutual fund companies. In this section this kickback is set to an average level.

5.1 Data

The data available is weekly fund flows from approximately 100 weeks. The dataset constructed contains the net effect on future revenues for the life insurance company the switches in every week contributed to.

$$\Delta fee_t = \sum_{policies}\{(-Amount * fee_{outgoing}) + (Amount * fee_{ingoing})\} * kickback \quad (5.1)$$

Where Δfee_t and $Amount$ are given in monetary terms, $Amount$ is the amount switched, while $fee_{outgoing}$ and $fee_{ingoing}$ is the management fee in percentage for the outgoing respectively the ingoing fund. The kickback is the average kickback return level.

Every fund switch in every week is captured and the net effect on the management fee in monetary terms is calculated. For instance, in a week where many policyholders changed from equity funds to fixed-income types of funds, the delta fee value is negative. Due to the fee for equity funds is typically higher than the fee for the fixed-income type funds. By taking the cumulative sum over the vector Δfee the y series is given, y thus displays the cumulative change in the fees.

$$y_t = cumsum(\Delta fee_t) \quad (5.2)$$

In addition, four different index series for the same weeks were gathered:

- OMX Stockholm benchmark Index (TR) (Swedish stock index)
- MSCI World SEK (Global stock index)
- OMRX TBill (Short-term government debt index)
- OMRX Tbond (Long-term government debt index)

5.2 Model for change in future revenue

The goal is to find a model that explains y in a sufficient way. Since y displays the cumulative change in fees so by looking at the last period one can make forecasts in the future how the revenue stream is changed from the last observed period to a period in the future based on how the market goes. By looking at the cumulative sum it does not matter how delta fee changes in every period between the last observed period and the forecast period.

Denoting indices value from week 1, until last week T

$$Index = \begin{pmatrix} OMX\ Stockholm_1 & MSCI\ World\ SEK_1 & OMRX\ TBill_1 & OMRX\ TBond_1 \\ \vdots & \vdots & \vdots & \vdots \\ OMX\ Stockholm_T & MSCI\ World\ SEK_T & OMRX\ TBill_T & OMRX\ TBond_T \end{pmatrix}$$

$$w = (w_1\ w_2\ w_3\ w_4)$$

The goal is to find the weights w, for a weighted index consisting of these four indices such that the correlation between the weighted index and the y series is maximized.

$$weightedIndex = w * Index' \tag{5.3}$$

$$\max_{\{w \in R\}} CORR(weightedIndex, y) \tag{5.4}$$

No constraints are set upon the weights for each index series, the weighted index will be used just as a variable in a regression model. If the correlation is high, the following simple linear regression model would prove as a good model

$$y_t = \alpha_0 + \alpha_1 weightedIndex_t + \varepsilon_t \tag{5.5}$$

Multiple versions of model 5.5 were evaluated, different lagged variables of the weighted index series were tested but it was shown that model 5.5 with the four indices included did achieve the highest correlation. There were only a slight difference between using all four indices and just a combination of two of them. The latter is chosen for this analysis since the simplicity and the result is better presented if there are only 3 dimensions instead of 5. The indices finally used were

- OMX Stockholm benchmark index (TR)
- OMRX Tbill

The optimization problem 5.4 were solved numerically in MATLAB, the estimated weights are given in table 24.

Weight	Index	Estimate
w_1	OMX Stockholm benchmark index (TR)	0.8969
w_2	OMRX Tbill	-2.6090
Corr(weightedIndex,y)	**0.92**	

Table 24 – weights etstimate for the weightedIndex

Together with the weighted index in Table 24, model 5.5 were estimated with ordinary least squares, see table 25.

Parameter (Variable)	Estimate
$\widehat{\alpha_0}$ (Constant)	$3.37 * 10^9$
$\widehat{\alpha_1}$ (Weighted Index)	$4.38 * 10^5$

Table 25 - Parameter estimate for model 5.5

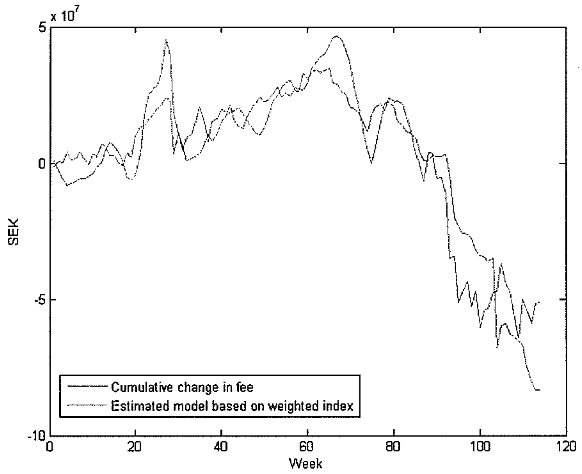


Figure 7 - Blue: y, Red: estimated model with weighted index

Figure 7, displays the actual cumulative fee change (blue) and the estimated model 5.5 (red). The correlation between the weighted index and the cumulative change is fees is very high. Figure 7 also displays that the model seems to fit the data well, which is backed by the coefficient of determination R2 is 0.8363.

5.2.1 Model diagnostics

To evaluate the fit of the model, the residual term is under investigation. If the residual sequence can be considered as white noise the model has an acceptable fit. A QQ plot of the quantiles from the residual sequence versus the quantiles from the normal distribution is given in figure 8, the quantiles of the residual sequence seems to fit those of a normal distribution. A few outliers but it seems that the sequence can be considered to be white noise.

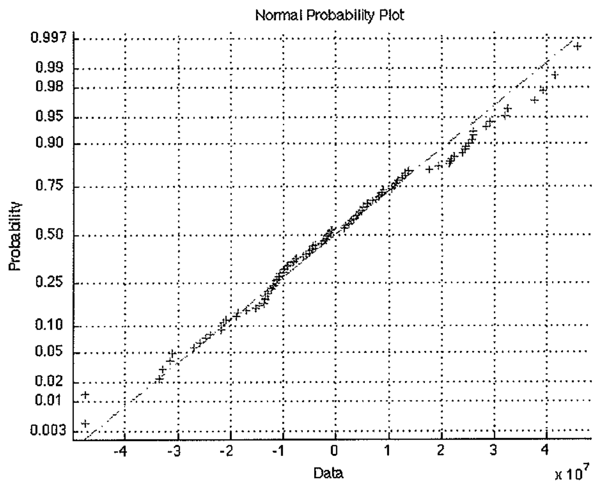


Figure 8 – QQ plot of the residuals against standrard normal distribution

A Jaque-Bera (J-B) test was also performed, see theory section. The J-B test statistic is evaluated to

$$JB = 0{,}2273 < CriticalValue = 5{,}49 \tag{5.6}$$

This indicates not to reject the J-B test null hypothesis, and conclude that the sample (residual sequence) comes from the normal distribution.

There seems to be some outliers but we conclude that the residual sequence can be considered as white noise, and that the model fit is good, the R2 value from the regression is 0.8363 which also indicates a good fit.

5.2.2 Model specification validation

The model fit is well, i.e. the linear regression model based on the weighted index fits y well, however if y is a good proxy for the real effect the fund switching will have on future revenues is in question. The raw data for the entire analysis is based on weekly fund flows for approximately 100 weeks backwards in time. Monthly information about the entire customer stock (which funds and the allocation they hold) is available for approximately 13 months backwards in time. The ideal case would have been if this data matched, for instance that the customer stock was given weekly and for the same period.

The limitations in this model come from the fact that the information of the customer stock is not synced and available with the fond flow data. There is also no reasonable easy way to backtrack the fund switches for a policyholder. If the policyholder changes from and back to a fund multiple times during a year this causes trouble, since the fee is only paid for a fund for as long as the fund is held.

Figure 9 shows the estimated cumulative change in fees plotted against the two indices. From the surface the change in revenue due the performance of the two indices can be seen.

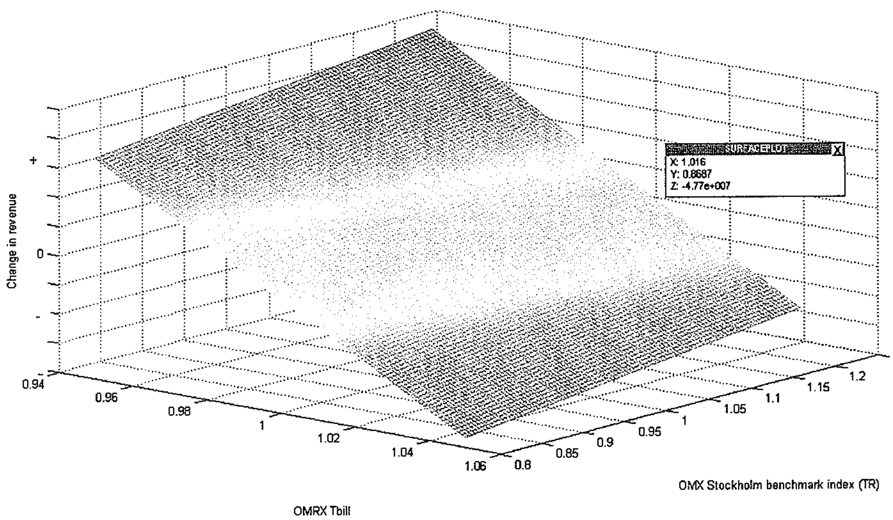


Figure 9 - Change in revenue versus OMRX Tbill and OMX Stockholm Benchmark index

The highlighted surface coordinates represent the performances of the two indices from 2011. I.e. if the OMX Stockholm benchmark index and OMRX Tbill will perform as they did in 2011, the change in revenue due to fund switches will be -47,7 Million SEK.

The main concern is how to validate the specification of this model, due to the limit in the data available. The choice fell on a visual inspection of the entire fund stock for 2011. Figure 5 shows the value of the entire stock (blue line) over 2011. The dotted red line shows the value of the fund stock that is placed in fixed-income funds while the dotted green line shows the value of the fund stock that is placed in equity funds. The red solid line is the cumulative sum of the amount in monetary terms that is switched to fixed-income funds. One can see that there is an increase in the amount switched to fixed-income funds over the year, which follows the value of the fixed-income stock almost precisely.

The solid green line shows the premium inflow minus premium outflow from equity fund. The black line shows premium in minus premium out from fixed-income funds. It can be concluded that the increase in fixed-income funds allocation of the entire portfolio is mostly based on fund switches to more funds of that type, rather than new premiums in and out.

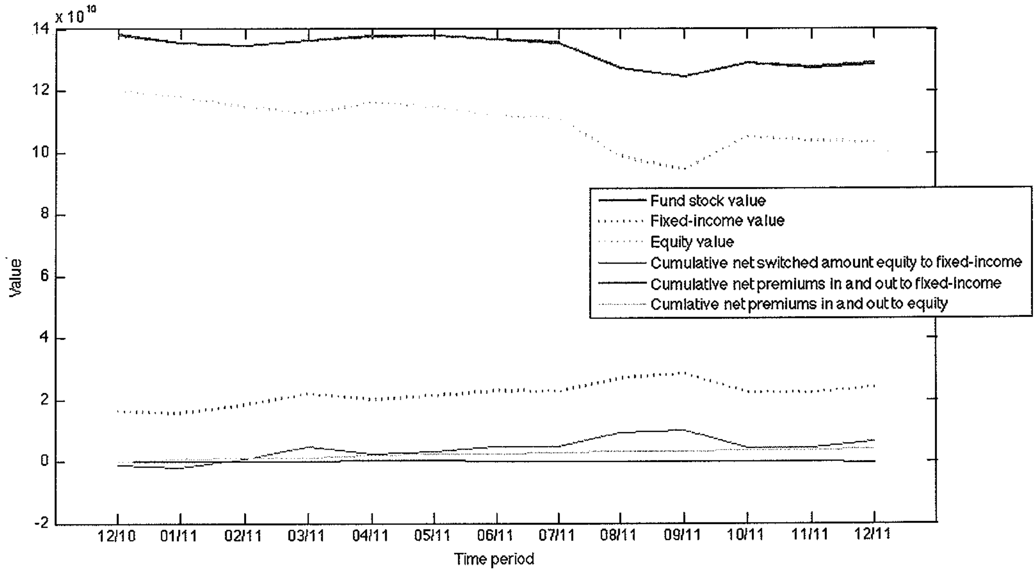


Figure 10 – Entire fund-stock

The re allocation from equity funds to fixed income type funds has a large impact on the insurance companies future revenues. The following data is gathered from 2011.

	Year start	Year end	Difference %
Total fund stock value	137, 99 B	128,83 B	0.9336
Equity value	119,98 B	103,30 B	0,8610
Fixed-income value	16,32 B	23,91 B	1,4651
Switch amount to fixed-income	-	8.10 B	
Premium in/out to fixed-income	-	-0.31 B	
OMX Stockholm Benchmark Index	519,33	451,07	0,8686
OMRX Tbill	3100,58	3151,09	1,0163

Table 26 - Statistics from 2011

Over the year of 2011 8.1 B SEK switched from equity to fixed-income while the value of the entire fixed-income stock increased with 7.59 B SEK. Premium in/out from fixed-income decreased with 0.31 B.

The following three average fees are calculated:

$$\overline{fee} = \frac{1}{N}\sum_{funds} fee_{i,} = 1{,}417\%$$

$$\overline{fee_{stock}} = \frac{1}{N_{stocks}}\sum_{stock\ funds} fee_i = 1{,}454\%$$

$$\overline{fee_{fixedIncome}} = \frac{1}{N_{fixedIncome}}\sum_{fixed\ income\ funds} fee_i = 0{,}818\%$$

If the model specification is correct the -47,7 M in change in revenue corresponds to the amount switched from equity to fixed-income funds.

$$Total\ amount\ switched = \frac{y_{T+1} - y_T}{fee_{stock} - fee_{fixedIncome}} = \frac{47{,}7}{0.01454 - 0.00818} = 7500\ M \qquad (5.7)$$

Thus if 7500 million in value is changed from equity funds to fixed-income funds the change in revenues to the life-insurance company based on these switches is approximately -47,7 M, comparing 5.7 with table 26 (8.1 B vs 7.5 B) it can be concluded that the model specification is acceptable. Note that the model takes in switches of all kind, for instance switches from equity to equity and the impact of these switches is not accounted for in this reasoning.

5.3 Result

The weighted index series can be used fine to model the cumulative change in fees. Depending on once assumptions on the stock and fixed-income market, conclusion about the future effect on the revenues that can be bound to fund switches can be drawn.

Inspecting the surface plot in figure 9, one can conclude that if short term government debt index decreases while the Swedish equity index increases, this will trigger many switches were the customer switches to a fund with a higher management fee, and the effect these switches will have on the revenue is positive.

Solvency capital requirements calculations stresses the current fund stock to forecast future revenues, in those calculations the fund stock is assumed to lay still, i.e. no fund switches are assumed to happen. This model can be used with current calculations in order to find more realistic best estimate calculation.

6 Conclusion and closing discussion

Factors that contributed to probabilities of certain types of fund switch were found. How much factors like age, type of product, having an insurance broker, value and the performance of the contract contributed were surprisingly small for some type of switches, even if they had the correct impact as formulated in the corresponding hypothesis.

The active customers is somewhat semi-active, they have chosen to have a fund placement service which handles all fund placement decisions for them. The value performance of the contract seems to affect any kind of fund switch in a way expected, however the probability of a risk averting switch when the value decreases is surprisingly low. Policyholder seems to switch to funds within the same type of risk category if the value of the contract decreases.

The average model for the probability of certain switches predicts correct in over 90% of the switch and non-switches investigated. The model explains risk averting switches better than the opposite type of switch. It can be concluded that even though the model is based on a lot of data, observation from a longer time perspective is needed, it would have been interesting to study other market cycles and data over several market cycles, so a more accurate model could be estimated.

The change in revenue model tries to explain the effect on the future revenue stream caused by fund switches. The model is satisfying, and the effect on the revenues from fund switches can be consider to be true for the short period investigated. Stress tests on the market indices can be done to see how the change in revenue estimates varies, this can be used with current solvency requirement calculations in order to find a more accurate estimate of total future revenues. Even if future revenues will fall if policyholders switches over to more fixed-income funds, it is important to understand that if the stock market goes down these switches can be considered profitable as they might help the entire fund stock withstand the decrease in the stock market. In the selected period just by visual inspecting the entire fund-stock figure (figure 10) one can believe that this is the case in 2011. This however has not been investigated further since the short period of time inspected. It would have been interesting to say something about this for a longer period of time, not just that in 2011 the switches helped to withstand the decrease of the entire fund stock and the future revenues will in fact be higher due to these switches, if this now is the case.

7 References

Achen C (1990) "What does "Explained Variance" Explain?: Reply" In: *Political Analysis* 1990, Vol 2, Issue 1, pages 173-184.

Baltagi B (2005), *Econometric Analysis of Panel Data*, 3rd edition, John Wiley & Sons Ltd.

Cameron A Colin, Trivedi Parvin K (2005), *Microeconometrics: Methods and Applications,* illustrated edition, Cambridge University Press.

Chamberlain G (1980) "Analysis of Covariance with Qualitative Data" In: *The Review of Economic Studies*, Vol 47, No. 1, Econometrics Issue (Jan 1980), pages 225-238, Oxford University Press.

Christensen R (1997), *Log-Linear Models and Logistic Regression*, 2nd edition, Springer.

Clemes M, Gan C, Zhang D (2010) "Customer switching behaviour in the Chinese retail banking industry" In: *International Journal of Bank Marketing* Vol 28, Issue 7, pages 519-546

Guadagni P, Little J (2008) "A Logit Model of Brand Choice Calibrated on Scanner Data." In: *Marketing Science* Vol 27 No 1. January – Februrary 2008, pages 29-48

Hsiao C (2003), *Analysis of Panel Data*, 2nd edition ,Cambridge University Press.

Madsen H (2008), *Time Series Analysis*, Chapman & Hall/CRC.

Svenskförsäkring (2012) ,*"Svensk Försäkrings branschstatistik Nr 1/2012 – 4e kvartalet 2011" (www.svenskforsakring.se/Statistics/StatBranch/Branschstatistik/2011/Kvartal%204%202011%20Rev%2023%20feb%202012.pdf) Accessed at 2012-02-27*

Verbeek M (2008), *A Guide to Modern Econometrics*, 3rd edition, John Wiley & Sons Ltd.

Wieringa J, Verhoef P (2007) "Understanding Customer Switching Behavior in a Liberalizing Service Market: An Explanatory Study" In: *Journal of Service Research* 2007 10, page 174-

Appendix 1 – GLM results

Parameter estimate for Logit and Probit model for both dataset, for both FS4 and FS5 as the dependent variable and for the value change variable calculated over 1 month backwards (ValueChange1) and calculated over 2 months backwards (ValueChange2)

Red cell indicates that the variable is insignificant in that period at 99% level.

DV: FS5, 12 month dataset for ValueChange1 and ValueChange2

FS5	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Variables						**Parameter estimate**						
Logit												
Constant	-10,69	13,70	-39,13	50,44	27,37	-53,90	-51,20	13,61	17,84	-1,73	-12,34	8,07
ValueChange1	7,17	-18,02	36,60	-52,66	-30,14	51,73	47,90	-15,84	-22,73	-1,26	8,49	-12,08
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,85	-0,59	-0,11	-0,16	-0,44	0,13	-0,32	0,46	0,37	0,62	0,28	-0,29
Male	0,09	0,13	0,19	0,12	0,12	0,22	0,14	0,24	0,13	0,30	0,30	0,24
hasEB	0,27	0,45	0,53	-0,53	-0,34	0,58	0,23	0,08	-0,23	0,01	-0,17	0,39
hasFPS	3,61	5,25	3,76	3,97	4,05	3,81	3,16	3,41	6,20	5,96	6,45	4,68
Age	0,01	0,01	-0,01	-0,01	0,00	-0,02	0,01	-0,01	0,00	-0,01	0,00	0,00
Efron R2	0,39	0,66	0,46	0,40	0,43	0,48	0,32	0,28	0,56	0,55	0,48	0,50
Mcfadden R2	0,38	0,59	0,38	0,36	0,39	0,41	0,31	0,22	0,49	0,46	0,41	0,46

Probit												
Constant	-5,73	5,45	-17,34	23,37	13,72	-25,16	-21,69	7,51	8,32	-1,11	-6,14	3,68
ValueChange1	3,78	-7,70	15,90	-24,67	-15,29	23,88	19,74	-8,83	-10,92	-0,55	4,07	-5,84
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Probit	-0,40	-0,28	-0,06	-0,08	-0,20	0,06	-0,13	0,24	0,15	0,28	0,12	-0,13
Male	0,06	0,07	0,10	0,08	0,08	0,11	0,08	0,13	0,07	0,14	0,14	0,12
hasEB	0,10	0,18	0,22	-0,22	-0,14	0,24	0,09	0,04	-0,11	0,01	-0,07	0,11
hasFPS	2,02	2,96	2,21	2,20	2,26	2,23	1,76	2,01	3,39	3,31	3,52	2,62
Age	0,00	0,01	0,00	0,00	0,00	-0,01	0,01	0,00	0,00	0,00	0,00	0,00
Efron R2	0,39	0,66	0,45	0,39	0,43	0,47	0,31	0,28	0,56	0,55	0,48	0,49
Mcfadden R2	0,38	0,58	0,38	0,35	0,39	0,40	0,30	0,22	0,49	0,46	0,41	0,46

FS5	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Variables						**Parameter estimate**						
Logit												
Constant	-3,04	-21,30	-1,89	58,88	22,91	5,27	-43,00	-5,60	13,48	3,20	-9,83	-12,52
ValueChange2	-0,19	17,22	-0,75	-61,29	-25,47	-8,04	40,05	3,71	-18,21	-6,95	6,37	8,35
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,86	-0,58	-0,09	-0,09	-0,49	0,12	-0,29	0,54	0,30	0,66	0,24	-0,33
Male	0,10	0,14	0,19	0,11	0,15	0,20	0,15	0,25	0,15	0,23	0,35	0,24
hasEB	0,27	0,48	0,57	-0,42	-0,30	0,57	0,22	0,06	-0,22	-0,03	-0,11	0,35
hasFPS	3,67	5,22	3,84	4,15	4,10	3,97	3,04	3,29	6,20	6,41	6,49	4,78
Age	0,01	0,01	-0,01	0,00	0,00	-0,01	0,01	-0,01	0,00	-0,01	0,00	0,00
Efron R2	0,39	0,66	0,45	0,40	0,44	0,47	0,33	0,27	0,56	0,55	0,48	0,50
Mcfadden R2	0,37	0,59	0,38	0,35	0,39	0,40	0,31	0,22	0,48	0,47	0,40	0,46

Probit												
Constant	-2,87	-9,03	-1,03	26,06	10,15	2,18	-19,58	-3,08	5,89	1,31	-4,79	-4,96
ValueChange2	1,01	6,90	-0,46	-27,44	-11,64	-3,75	17,82	1,95	-8,37	-3,35	2,91	2,78
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,41	-0,27	-0,05	-0,05	-0,23	0,07	-0,13	0,30	0,11	0,27	0,11	-0,14
Male	0,06	0,07	0,10	0,07	0,09	0,10	0,09	0,14	0,08	0,11	0,16	0,12

hasEB	0,11	0,19	0,23	-0,19	-0,11	0,22	0,09	0,02	-0,11	0,00	-0,05	0,10
hasFPS	2,04	2,95	2,26	2,29	2,28	2,31	1,70	1,97	3,36	3,48	3,55	2,66
Age	0,00	0,00	0,00	0,00	0,00	-0,01	0,01	-0,01	0,00	0,00	0,00	0,00
Efron R2	0,38	0,66	0,45	0,40	0,43	0,46	0,32	0,27	0,56	0,55	0,48	0,49
Mcfadden R2	0,37	0,59	0,37	0,35	0,39	0,39	0,31	0,22	0,48	0,47	0,40	0,46

DV: FS4, 12 month dataset for ValueChange1 and ValueChange2

FS4	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Variables						**Parameter estimate**						
Logit												
Constant	20,02	26,41	-58,42	92,54	51,84	-59,80	-63,90	-36,44	-11,36	-18,56	-9,99	5,90
ValueChange1	-25,30	-32,85	53,90	-97,75	-56,57	55,55	59,60	32,94	7,51	13,80	3,52	-11,18
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,12	-0,48	-0,57	-0,59	-0,76	-0,31	-0,43	-0,44	-0,53	-0,66	-0,18	-0,46
Male	-0,02	-0,16	-0,03	-0,07	-0,05	0,17	0,09	0,13	0,10	0,06	0,34	0,15
hasEB	0,35	0,71	0,94	-0,63	0,08	1,53	0,30	1,04	0,00	0,31	0,25	0,95
hasFPS	2,96	4,65	3,84	5,89	4,45	4,69	3,35	4,36	5,50	4,04	6,85	5,36
Age	0,02	0,03	0,01	0,02	0,02	0,00	0,02	0,00	0,01	0,02	0,03	0,00
Efron R2	0,09	0,57	0,47	0,57	0,45	0,66	0,33	0,60	0,68	0,39	0,73	0,61
Mcfadden R2	0,20	0,57	0,48	0,58	0,49	0,62	0,35	0,54	0,64	0,46	0,67	0,62

Probit												
Constant	2,92	9,86	-20,43	33,13	21,55	-20,96	-23,89	-15,61	-4,17	-5,72	-4,52	0,98
ValueChange1	-5,64	-12,96	18,11	-35,76	-23,99	18,67	21,35	13,73	1,97	3,10	1,40	-3,63
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,09	-0,26	-0,31	-0,32	-0,38	-0,19	-0,18	-0,25	-0,27	-0,31	-0,08	-0,24
Male	0,00	-0,05	0,01	-0,01	-0,01	0,09	0,06	0,07	0,06	0,05	0,14	0,08
hasEB	0,12	0,27	0,39	-0,23	0,09	0,62	0,13	0,42	0,02	0,12	0,09	0,30
hasFPS	1,20	2,49	2,06	2,71	2,15	2,55	1,76	2,44	2,98	2,13	3,66	2,81
Age	0,01	0,01	0,01	0,01	0,01	0,00	0,01	0,00	0,00	0,01	0,01	0,00
Efron R2	0,07	0,57	0,46	0,53	0,43	0,64	0,30	0,59	0,67	0,38	0,73	0,60
Mcfadden R2	0,20	0,56	0,47	0,56	0,48	0,61	0,34	0,53	0,64	0,45	0,67	0,62

FS4	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Variables						**Parameter estimate**						
Logit												
Constant	-1,07	5,93	-15,48	109,36	38,66	10,25	-51,26	-29,49	-12,12	-16,48	5,94	-24,49
ValueChange2	-4,89	-11,28	11,21	-114,85	-42,95	-15,20	47,32	26,25	8,42	12,36	-12,10	18,55
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,11	-0,57	-0,57	-0,56	-0,80	-0,32	-0,39	-0,51	-0,49	-0,61	-0,12	-0,48
Male	-0,05	-0,12	-0,01	-0,07	-0,01	0,14	0,10	0,13	0,11	0,08	0,34	0,12
hasEB	0,36	0,75	1,05	-0,23	0,21	1,52	0,29	0,98	-0,05	0,26	0,23	0,78
hasFPS	2,73	4,78	3,87	6,10	4,42	4,90	3,21	4,37	5,48	3,86	6,74	5,62
Age	0,02	0,02	0,01	0,02	0,02	0,00	0,02	0,00	0,00	0,02	0,03	0,00
Efron R2	0,07	0,56	0,45	0,58	0,45	0,66	0,33	0,61	0,68	0,43	0,73	0,63
Mcfadden R2	0,19	0,56	0,47	0,58	0,48	0,61	0,36	0,54	0,64	0,47	0,67	0,63

Probit												
Constant	-3,06	-1,87	-6,47	40,47	13,26	3,12	-21,65	-11,64	-4,83	-5,39	1,33	-8,50
ValueChange2	0,13	-0,85	4,23	-43,21	-15,56	-5,65	19,30	9,84	2,71	2,95	-4,34	5,67
Value	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

P	-0,08	-0,29	-0,30	-0,30	-0,41	-0,18	-0,17	-0,28	-0,26	-0,30	-0,06	-0,25
Male	-0,01	-0,04	0,01	-0,02	0,02	0,08	0,06	0,07	0,06	0,05	0,14	0,07
hasEB	0,12	0,29	0,41	-0,11	0,14	0,60	0,13	0,40	0,00	0,11	0,08	0,27
hasFPS	1,15	2,53	2,08	2,81	2,14	2,63	1,70	2,44	2,96	2,06	3,63	2,87
Age	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,00	0,00	0,01	0,01	0,00
Efron R2	0,06	0,55	0,44	0,54	0,42	0,64	0,31	0,59	0,68	0,40	0,73	0,61
Mcfadden R2	0,19	0,56	0,46	0,56	0,47	0,60	0,35	0,53	0,64	0,46	0,67	0,62

DV: FS5, 6 month dataset for ValueChange1 and ValueChange2

FS5	Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Okt	Nov/Dec
Variables			**Parameter estimate**			
Logit						
Constant	-23,45	-21,10	24,06	-50,53	13,07	-11,08
ValueChange1	19,63	18,73	-26,60	49,04	-17,53	7,51
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,56	-0,13	0,00	0,38	0,25	0,10
Male	0,14	0,19	0,19	0,25	0,15	0,27
hasEB	0,27	-0,15	-0,04	-0,01	-0,16	-0,16
hasFPS	4,73	5,99	5,15	3,31	7,56	6,23
Age	0,01	0,00	0,00	-0,01	0,01	0,00
Efron R2	0,55	0,55	0,55	0,28	0,57	0,43
Mcfadden R2	0,48	0,46	0,46	0,23	0,50	0,36

Probit						
Constant	-11,00	-9,91	13,26	-26,54	5,97	-5,63
ValueChange1	8,95	8,54	-14,70	25,62	-8,34	3,68
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,28	-0,07	0,00	0,19	0,08	0,04
Male	0,08	0,09	0,10	0,14	0,08	0,13
hasEB	0,12	-0,07	-0,02	-0,01	-0,07	-0,07
hasFPS	2,72	3,30	2,94	1,98	3,94	3,43
Age	0,00	0,00	0,00	0,00	0,01	0,00
Efron R2	0,56	0,55	0,55	0,28	0,57	0,43
Mcfadden R2	0,48	0,46	0,46	0,23	0,49	0,36

FS5	Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Okt	Nov/Dec
Variables			**Parameter estimate**			
Logit						
Constant	-15,27	0,96	19,66	-43,19	8,91	-6,60
ValueChange2	11,63	-3,50	-22,03	42,08	-13,15	3,39
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,58	-0,11	-0,05	0,43	0,18	0,09
Male	0,14	0,18	0,22	0,26	0,17	0,31
hasEB	0,28	-0,15	0,00	0,00	-0,16	-0,12
hasFPS	4,74	6,08	5,19	3,17	7,50	6,24
Age	0,01	0,00	0,00	-0,01	0,01	0,00
Efron R2	0,55	0,55	0,55	0,29	0,57	0,43
Mcfadden R2	0,48	0,46	0,46	0,24	0,49	0,35

Probit						
Constant	-7,12	0,50	9,34	-22,97	3,72	-3,35

ValueChange2	5,17	-1,96	-10,71	22,26	-5,95	1,58
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,29	-0,05	-0,04	0,21	0,05	0,04
Male	0,07	0,09	0,11	0,14	0,09	0,15
hasEB	0,13	-0,07	-0,01	-0,00	-0,07	-0,05
hasFPS	2,72	3,35	2,95	1,91	3,91	3,44
Age	0,00	0,00	0,00	0,00	0,00	0,00
Efron R2	0,56	0,55	0,55	0,29	0,57	0,43
Mcfadden R2	0,47	0,46	0,46	0,24	0,49	0,35

DV: FS4, 6 month dataset for ValueChange1 and ValueChange2

FS4	Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Okt	Nov/Dec
Variables			**Parameter estimate**			
Logit						
Constant	-2,21	-32,58	36,11	-26,02	-6,79	-9,55
ValueChange1	-2,72	27,95	-40,65	21,98	2,42	3,51
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,58	-0,56	-0,37	-0,31	-0,48	-0,28
Male	-0,09	-0,07	0,08	0,15	0,10	0,30
hasEB	0,66	-0,06	0,93	0,66	0,11	0,20
hasFPS	4,46	4,65	4,89	3,99	5,25	6,69
Age	0,02	0,03	0,01	0,02	0,02	0,03
Efron R2	0,53	0,54	0,65	0,50	0,65	0,68
Mcfadden R2	0,52	0,52	0,60	0,43	0,59	0,62

Probit						
Constant	-4,40	-12,44	16,28	-9,06	-2,77	-4,34
ValueChange1	1,78	10,00	-18,65	6,82	0,37	1,39
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,28	-0,28	-0,20	-0,16	-0,23	-0,13
Male	-0,03	-0,01	0,06	0,08	0,05	0,12
hasEB	0,24	-0,01	0,39	0,27	0,05	0,08
hasFPS	2,41	2,54	2,70	2,32	2,92	3,60
Age	0,01	0,01	0,01	0,01	0,01	0,01
Efron R2	0,52	0,54	0,64	0,49	0,65	0,68
Mcfadden R2	0,52	0,52	0,59	0,43	0,59	0,62

FS4	Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Okt	Nov/Dec
Variables			**Parameter estimate**			
Logit						
Constant	3,50	-14,43	15,79	-26,41	-8,84	5,37
ValueChange2	-8,15	10,03	-20,21	22,64	4,70	-11,12
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,57	-0,55	-0,45	-0,30	-0,45	-0,22
Male	-0,08	-0,06	0,13	0,15	0,11	0,30
hasEB	0,64	-0,02	1,03	0,65	0,08	0,19
hasFPS	4,58	4,68	4,83	3,91	5,19	6,58
Age	0,02	0,02	0,01	0,02	0,02	0,03
Efron R2	0,54	0,54	0,64	0,51	0,65	0,68
Mcfadden R2	0,52	0,51	0,59	0,44	0,59	0,62

Probit						
Constant	-2,30	-6,05	4,77	-10,44	-3,79	1,59
ValueChange2	-0,24	3,69	-7,12	8,32	1,49	-4,43
Value	0,00	0,00	0,00	0,00	0,00	0,00
P	-0,28	-0,27	-0,23	-0,16	-0,22	-0,11
Male	-0,02	-0,01	0,08	0,08	0,06	0,12
hasEB	0,24	0,00	0,43	0,27	0,04	0,07
hasFPS	2,43	2,55	2,67	2,28	2,90	3,57
Age	0,01	0,01	0,01	0,01	0,01	0,01
Efron R2	0,52	0,53	0,64	0,50	0,65	0,68
Mcfadden R2	0,52	0,51	0,58	0,43	0,59	0,62

Joakim Mossberg I06

Fondbyten inom fondförsäkring – Agerar försäkringstagarna som rationella investerare?

Syftet med examensarbetet var att undersöka faktorer som påverkar om och varför folk byter mellan olika fonder inom sin fondförsäkring. Detta gjordes dels för att få en överblick hur, vem, varför och när bytena sker? Samt dels för att kunna uppskatta vad effekten av bytena blir på framtida kassaflöden.

Sannolikhetsmodell

En samling försäkringstagare fondportföljer undersöktes och information angående vilka fonder som försäkringstagarna har bytt till och från samlades. Utöver detta så lagrades kontrakt specifik information angående försäkringstagaren och kontraktet, t.ex. försäkringsvärde, försäkringstyp och värdeutvecklingen för fonderna inom försäkringen de senaste månaderna. Från det här urvalet söktes en modell som beskriver sannolikheten att en försäkringstagare gör ett specifikt fondbyte under en viss period. Sannolikheten att en försäkringstagare utför ett fondbyte modelleras med en typ av binär regression sk. Generlized Linear Model (GLM). I vanlig binär regression antas sannolikheten att ett binärt utfall inträffar vara lika med en linjärkombination av de förklarande variablerna. I de sk. GLM modellerna så antas sannolikheten att ett binärt utfall inträffar vara lika med en funktion av en linjärkombination av de förklarande variablerna. Funktionen som används brukar vara en fördelningsfunktion eftersom sannolikheter ska ligga mellan 0 och 1, samma värden som en fördelningsfunktion antar. I detta fall sannolikheten att en viss kund gör ett vist fondbyte är då lika med en funktion av de kontraktspecifika faktorerna. Tre stycken fördelningsfunktioner undersöktes och den fördelning som valet tillslut föll på var standard logistic fördelningen.

Faktorerna

En av de kontraktspecifika parameterna som samlas in är huruvida kontraktet är knuten till en s.k. Fondplaceringstjänst.

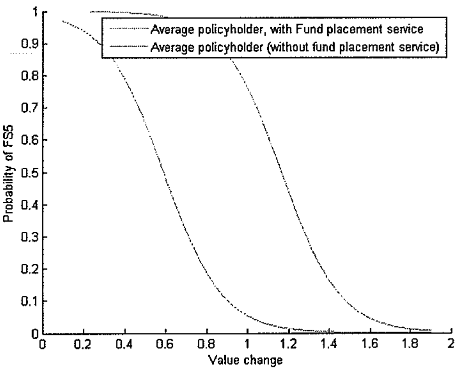
En fondplaceringstjänst är en tjänst som erbjuds till ett pris av ca 1% av försäkringsvärdet till försäkringstagarna. Det finns ett flertal olika fondplaceringstjänster på den svenska marknaden som kunden kan välja att ansluta sig till. Om kunden har anslutit sig till en sådan så lämnar försäkringstagaren alla investeringsbeslut som berör försäkringen till tjänstleverantören. Tjänsteleverantören bestämmer således vilka fonder som kunden har, samt när och till vad ett eventuellt byte ska ske. Samtliga försäkringstagare som är knutna till samma fondplaceringstjänst delar således samma fondallokering.

Värdeutvecklingen på försäkringskontraktet tros ha en ha betydande påverkan på sannolikheten att en försäkringstagare utför något typ av fondbyte, om fonderna som försäkringstagaren har valt, eller låtit någon annan välja, underpresterar så tros försäkringstagaren reagera på detta faktum och byta fonder.

Ett antal andra faktorer som ålder, försäkringstyp, kön etc. insamlades.

Resultat

Det visade sig att de flesta kunderna ligger väldigt stilla i sina portföljer och sällan byter fonder. Av de försäkringstagarna som faktiskt gör något typ av byte under det senaste året har ca 60 % av de kontrakten en fondplaceringstjänst knuten till sig. En snitt försäkringstagares attributs beräknas, vilken typ av försäkring, kön, ålder, värde på kontraktet etc. som snitt kunden har. Denna kund är inte kopplad till någon fondplaceringstjänst, sannolikheten för att en kund med snitt attribut utför ett byte under en viss period ses i figur 1.



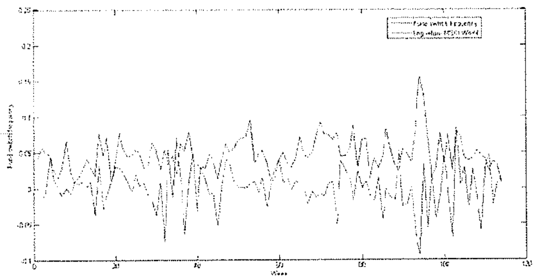
Figur 1 - Snittkund (Röd), snittkund fast med fondplaceringstjänst (Blå)

Av de över 11 miljoner observationer som analysen är baserad på så träffar modellen rätt i över 92 % av fallen, baserat på att om sannolikheten att man gör ett byte är över 0.5 så anses då bytet ha skett. Detta jämförs sedan med om bytet faktiskt har ägt rum. Från den röda linjen i figur 1 som bygger på en snitt kund så kan det ses, för att sannolikheten ska gå över 0.5, och att kunden då antas göra ett byte så måste värdeutvecklingen vara ca 0.6. Dvs. att värdeutvecklingen på kontraktet är minus 40% på en månad! Bland de undersökta kontrakten finns det ingen med en sådan hastig nedgång, dock finns det exempel på andra nedgångar som får anses som stora, men ingen i den här magnituden. Men passiviteten bland försäkringstagarna som saknar en fondplaceringstjänst, lyser starkt. För de kunder som har en fondplaceringstjänst vilket är ungefär 13 % av kunderna så ser vi att värdeutvecklingen på försäkringen är en mer drivande faktor än för de som saknar en. En rimlig förklaring borde ju vara att leverantörerna för dessa tjänster agerar mer rationellt, dvs. att agera utifrån de förutsättningar som finns på marknaden. För de kunder som saknar en sådan tjänst kan vi se att sannolikheten att de gör något typ av byte är lågt oavsett vad man har för andra attribut. Av de kunder som är aktiva med fondbyten har ca 60 % någon typ av fondplaceringstjänst, anledningen till varför att de resterande försäkringstagarna byter fonder är däremot svårare att förstå.

Att modellerna personers beteende av olika slag är svårt, vilket även detta visar. Målet med denna modell var att senare gå vidare och titta på vad effekten från bytena blir på den framtida intjäningen. Eftersom det verkar lättare att förstå hur rationella investerare agerar (fondplaceringstjänster) än hur individuella försäkringstagare agerar så baserades nästa del av arbetet på icke kontraktspecifika faktorer, t.ex. variabler som beskriver det ekonomiska klimatet på de olika värdepappersmarknaderna.

Framtida intjäning

Eftersom de flesta av bytena som sker görs av fondplaceringstjänster så kan man tänka sig att investerarna bakom agerar rationellt utifrån de förutsättningar som finns på marknaden. Baserat på ens förväntningar angående läget på vissa marknader så skulle alltså de fondbyten som sker kunna uppskattas. Figur 2 visar fondbytesfrekvensen, dvs. andel av försäkringstagarna som gör något typ av byte i aktuell vecka. Data från över 110 veckor från 2009 till 2011 insamlades och plottates mot avkastningen från ett globalt aktieindex från samma period. Här finns det ett tydligt mönster att när avkastningen från aktiemarknaderna är negativ och stor, så ungefär en vecka senare ses det en ökande frekvens av kunder som utför fondbyten. Vi ser det mönster vi trodde hela tiden, att om aktiemarknaden presterar dåligt kommer kunderna (stor del fondplaceringstjänster) att agera och byta fonder i sina portföljer.



Figur 2 – Fondbytesfrekvens (Blå), logavkastning världsindex (röd)

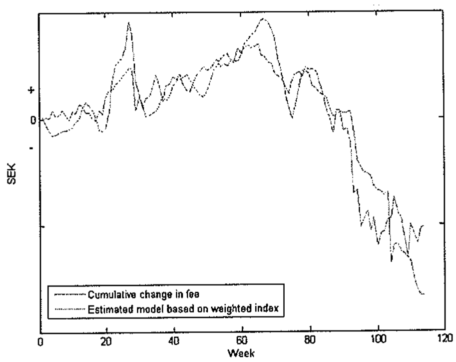
Fondbytesfrekvensen ser alltså ut att vara negativt korrelerad med en förskjutning av avkastningen. Eftersom vi är mer intresserade av vad dessa fondbyten har för effekt på framtida intjäningen istället för hur stor del av kunderna som utför ett byte, så modelleras avgiftsförändingen från bytena istället.

Intjäning

Varje fond förvaltas av ett fondbolag, detta fondbolag tar sedan ut en årlig förvaltningsavgift, på värdet i fonden. Förvaltningsavgiften skiljer sig mellan olika fonder och fondtyper. Räntefonder dvs. fonder som placerar sitt kapital i olika obligationer eller andra räntebaserade tillgångar, har ofta en lägre förvaltningsavgift än aktiefonder som placerar sitt kapital på de olika aktiemarknaderna runt om i världen. Försäkringsbolaget har förhandlat till sig en procentuell "kick-back" från fondbolagen, intjäningen för försäkringsbolaget kommer alltså från stor del i en procentuell kick-back på förvältningsavgiften. Denna kick-back förhandlas individuellt och skiljer sig mellan olika fonder och fondbolag.

Modell för avgiftförändring

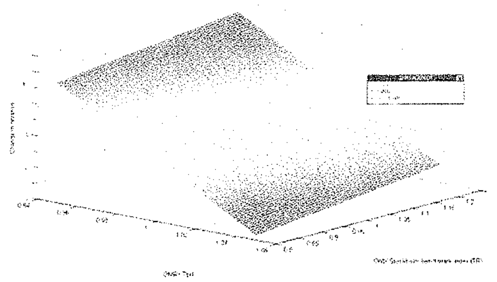
Den kumulativa avgiftsförändring som de veckovisa bytena bidrar till, modeleras med en indexserie. En indexserie som matchar den kumulativa avgiften viktas fram och en linjär regressions modell skattas där den kumulativa avgiften antas vara lika med en konstant term plus en faktor multiplicerat med den viktade indexserien. Figur 3 visar den kumulativa avgiftsförändringen samt modellen som baserats på det viktade indexet.



Figur 3 - Kumulativ avgiftsförändring (Blå) Skattad modell baserad på viktad indexserie (Röd)

Modellen passar den kumulativa avgiftsförändringen bra. I den viktade indexserien ingår två stycken olika index serier. Den första är ett svensk aktieindex och det andra är ett svenskt kortränteindex. Utifrån ens förväntningar på den svenska aktie- och (kort)räntemarknaderna så kan prediktioner om hur avgiften till försäkringsbolaget kommer att förändras utföras. Även stresstester baserat på "worst-case scenario" och "best-case scenario" kan göras.

Figur 4 visar en yta som beskriver hur intäkterna från avgiften ändras baserat på hur de två indexserierna utvecklas.



Figur 4 - X-axeln: Ränteindex, Y-axeln: Aktieindex, Z-axeln: Estimerad intäktsförändring

Resultat

Om kortränteindexet faller samtidigt som aktiemarknaden ökar kraftigt kommer det att bidra till den största positiva avgiftförändringen till försäkringsbolaget (Det mest röda hörnet i figur 4). Det beror på att i ett sådant läge kommer försäkringstagarna att fly från räntebaserade fonder och gå över mer i aktiefonder, dvs. de går från billigare till dyrare fonder vilket då i sin tur *ökar* intjäningen till försäkringsbolaget. Motsatsen, att räntemarknaden går kraftigt uppåt medan aktiemarknaden kraschar är det sämsta möjliga scenariot för försäkringsbolaget då förändringen i avgifterna är som mest negativ (Det mest blåa hörnet i figur 4). Det beror på att försäkringstagarna då flyr aktiefonder och går över i mer räntebärandefonder, dvs. de går från dyrare till billigare fonder vilket i sin tur *minskar* intjäningen till försäkringsbolaget.

De slutsatser som kan dras av detta är att de flesta kunder som gör byten och är aktiva i sin försäkring, är något "semi-aktiva". De har överlåtit investeringsbesluten till något annan och är dessutom bereda att betala ett premium för detta. För dessa kunder så verkar värdeutvecklingen på kontraktet spela roll, vilket stämmer med den teorin att de som placerar fonder för fondplaceringstjänsterna agerar som rationella investerare. För de försäkringstagare som ej har anslutit sig till en sådan tjänst så är de flesta passiva med fondbyten. De kunder som trots allt gör byten själva, dvs. att de som saknar en fondplaceringstjänst och gör byten, deras beteende är svårare att fånga upp med en modell som använts här.
Värdeutvecklingen verkar inte spela en så betydande roll som hade anats, de övriga faktorerna som insamlades hade inte heller någon jätte stor inverkan i sannolikheten att observera ett fondbyte.

I och med att de flesta bytena som sker görs av en fondplaceringstjänst så kunde ett mönster i frekvensen av byten ses med relation hur det går på de stora väldsmarknaderna. Från detta utarbetates en modell som hade till uppgift att förutspå hur förändringen av intjäningen går som kan härledas från fondbyten, relaterat till olika index från värdepappersmarknaderna.

Master's Theses in Mathematical Sciences 2012:E6
ISSN 1404-6342

LUTFMS-3188-2012

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/