HM Revenue & Customs

Research report

Attitudinal Prediction Models

Can third party data predict taxpayer attitudes?

**Behaviour Change & Customer Segmentation Research
PT Change**

6th November 2013

Behavioural Evidence & Insight Team

About this paper

This report has been written by the Behaviour Change & Customer Segmentation Research Team within HMRC, drawing heavily on advice provided by Ipsos Mori, which we would like to acknowledge.

About Behaviour Change & Customer Segmentation

HMRC carries out a variety of research to support Behaviour Change and Customer Segmentation. Such work helps understand and provide the capability to tailor HMRC's products and services - and the way HMRC works - to the needs, abilities and motivations of customers.

Contents

Research requirement (background to the project)

Background:
In 2007 HMRC commissioned large scale[1] quantitative and qualitative research into individuals' underlying attitudes towards tax compliance. This was essential to understand customers' underlying attitudes, thoughts and feelings towards tax & HMRC.

Statistical analysis of the research findings (from both 2007 & a later refresh in 2010) established that HMRC customers divide into **five distinct attitudinal segments**:
> 1) Unaware;
> 2) Willing but Needs Help;
> 3) Willing & Able;
> 4) Potential Rule Breakers and;
> 5) Rule Breakers.

Underpinning the segments are **four key dimensions** that determine which segment customers' fall into:

> 1) Awareness of their tax obligations;
> 2) Ability to understand and deliver those obligations;
> 3) Motivation to comply and;
> 4) Opportunity to be non-compliant.

While the research provided an overview of HMRC's customer population and their attitudes, it did not provide the tools to enable HMRC to identify needs, attitudes and motivations at an individual level, and it was clearly not possible to undertake a research survey for all HMRC's 40 million individual customers.

The research did however demonstrate clear links between customers' segment and certain demographic and socio-economic data, such as age, income level etc.

HMRC therefore wanted to explore the options for using demographic, socio-economic and other personal data to predict a customer's likely attitudinal segment.

When the research took place

The analysis took place between January 2012 and May 2013.

---

[1] The studies in 2007 covered c.3,000 customers, pls 42 in-depth interviews and 10 focus groups. The study in 2010 covered 3,700 customers.

Who did the work (research agency)

Ipsos Mori

Method, Data and Tools used, including Sample selection

The pilot phase was conducted on the full set of variables available from HMRC and a number of third party data providers. It was left to the model analysts to select the most powerful predictors through step-wise regression.

After the pilot however, it was decided to concentrate the analysis on a smaller subset of variables to make the best use of the available data and allow more in-depth investigations, like computing derived variables, interactions and non-linear transformations.

Statistical techniques including correlation and decision-tree analysis were primarily used to identify a subset of variables likely to be the best predictors of the dimension scores.

Correlation was preferred to regression analysis initially, to select optimal inputs variables based on their one-to-one relationship with the dimension scores. The reason for this was to bring forward to further analysis all the variables that had the potentials to improve the model and leave the selection of the most powerful ones to subsequent stepwise regression analysis, which then created the final models.

The sample used was the 3,721 survey participants from the original segmentation research, giving a baseline in terms of customer's attitudes towards HMRC & Tax.

Main Findings:

It was not possible, despite trialling a number of modelling techniques, to predict segments directly with sufficient accuracy to enable actionable results.

However, it was possible to predict a customer's propensity against the underlying segmentation dimensions: Awareness, Ability, Motivation, and Opportunity. Plus a complementary dimension: Fear of Penalty (combined with the Motivation dimension, this provides an indication into likely customer responses to stronger compliance messages).

The findings reaffirmed the segmentation research which indicated that specific demographic and socio-economic data could be powerful predictive indicators.

Over 2,000 data variables were analysed and considered during the modelling process but in the end, 80 variables were identified as having key predictive value. The chosen variables covered a range of characteristics such as:

- Basic demographics (such as age)
- Educational qualifications

- Household income;
- Experience of dealing with HMRC; and
- Understanding and use of similar services (such as insurance or pensions).

The performance of the models is measured by the extent to which they correctly predict customers likely Awareness, Ability, Motivation, Opportunity and attitude towards Penalties. All of these characteristics are modelled numerically, by generating a score based on the customer data used. The higher the score, the higher the customer's awareness (or whatever) is likely to be.

They have been purposefully designed to be flexible, and consistent with HMRC's need to be able to target those with the highest or lowest characteristics. For example, the models can identify customers who fall within the top 5% in terms of overall Ability. Equally they could identity a much wider population e.g. the top 75% in terms of overall Ability. The first example provides a very high level of accuracy but delivers a small target population. The second example provides a large population but includes customers with a wider ability range.

All the models delivered sufficiently accurate results to improve HMRC's targeting of communications relative to a random or one size fits all approach.

The models also deliver a propensity score, for example, rather than just being able to select those customers with a predicted Awareness score greater than X, it is also possible to identify those customers for whom the probability of their actual score being greater than X is at least, say, 80%. This again is likely to be helpful when making decisions about the accuracy required to deliver value in each particular real-life situation.

In summary

The models are not designed to be used in isolation but alongside other data and analysis to help build up a picture of each individual customer. This type of analysis supports HMRC in sending the right message, to the right customer, at the right time, tailored to their needs and preferences.

The results from the analysis provide a sound basis for HMRC to judge the viability of using attitudinal prediction models alongside other data and initiatives to help target and tailor services, products and processes to the needs, abilities and motivations of our customers as envisaged by the Department's Customer Centric Business Strategy.