Research report

HMRC compliance segmentation
- Segment prediction.

Business Customer & Strategy

March 2013

Behavioural Evidence & Insight Team

About the Business Customer and Strategy Team

Business Customer and Strategy is part of Business Tax. The goal of BC&S is to maximise Business Customer compliance for HMRC at best cost for both HMRC and the customer. This is done by developing business tax strategies through customer understanding, working with teams in HMRC and across government departments.

Contents

Research requirement (background to the project)

The SME Compliance Segmentation was developed in 2009 to create an attitudinally-based compliance segmentation of SME businesses, and to provide a market representative view of SMEs' interactions with HMRC and their attitudes towards fulfilling their tax obligations. The resulting seven segments were built on the following six core dimensions:

1. Awareness
2. Acceptability of evasion
3. Perception of risk
4. Ability
5. Prioritisation of payments to HMRC
6. Ongoing cash flow difficulties
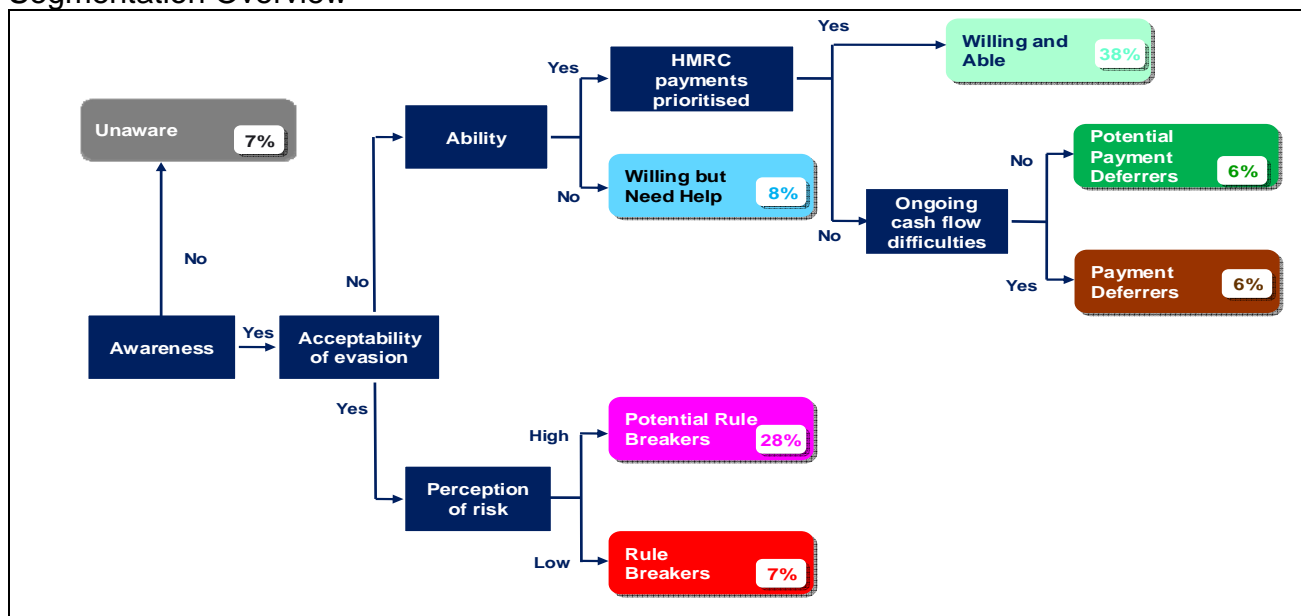
Segmentation Overview

Fig. 1

The SME Compliance Segmentation has provided HMRC with the basis from which to understand *why* businesses behave as they do and, to date, has been used primarily to underpin HMRC'S Tier 2 SME strategy.

The overall aim of the Segment Prediction work was to move from theory to practice in order to realise the potential of the SME Compliance Segmentation as a means to **direct behavioural change.**

3/E04, 100 Parliament St. London, SW1A 2BQ

While the SME Compliance Segmentation provides the *basis* for the different strands of work in this project only by looking at the *dimensions* on which the segmentation is based can the full potential be realised.

The research looked to systematically recreate the findings of the attitudinal research at an individual customer level and develop models with sufficient accuracy at the individual level to improve HMRC's targeting of communications to sole traders.

Who did the work (research agency)

The research was conducted by Quadrangle.

When the research took place

The research programme took place between January – July 2011.

Method, Data and Tools used, Sample

Sample was drawn from the HMRC database, and a specifically designed survey was conducted amongst 3,500 respondents. The sample was skewed to reduce the size of the Willing & Able segment, therefore biasing it towards the attitudinally non-compliant. This enabled greater understanding of the SME universe.

The methodology was designed to link directly into HMRC's own data – it was only in so doing that we could ensure the overall practical application of the work and its applicability for those who work with HMRC's own systems e.g. Risk and Intelligence Service (RIS) and Knowledge, Analysis & Intelligence (KAI).

Whereas the segmentation began from the desire to be representative, this work needs to begin from data that resides within HMRC. Additionally, key to the application of HMRC's SME segmentation is customer-level attribution.

To address this we conducted a 15 minute survey to produce an enhanced set of segmentation data and to identify:
- their segment;
- their segment dimensions;
- Experian Commercial Mosaic data;
- Business demographics.

The purpose of the work was to deliver a series of predictive models to help HMRC use its own available data to predict the core dimensions of the segmentation and, if possible, to

also predict segment membership itself. The modelling work was carried out in conjunction with HMRC KAI and RIS colleagues and the approach to this is detailed in the body of this document.

Main Approach

In summary

There were seven simple steps in this project:

1. To draw a sample from HMRC database and conduct a specifically designed survey on this sample
   - The sample was skewed to reduce the size of the Willing & Able segment.
   - n=3500 respondents.
2. We then recreated the dimensions from the SME segmentation, as well as the segments themselves within the new data set, this was completed;
   - based upon the research data,
   - and using the same algorithms as for the original segmentation.
3. Subsequently we obtained HMRC and Experian data and matched this to the research respondents.
4. Once starting the modelling work, we initially built models to predict three meta-segments (these are detailed in the next section below) which was done;
   - using HMRC and Experian data as the seven SME Compliance segmentation segments lacked sufficient size to build predictive models in their own right.
5. We subsequently build models to predict the underlying segment dimensions which are actually the levers that HMRC might use to influence business behavioural change.
   - The six underlying dimensions were successfully modelled.
6. Finally, we were able to validate the model inputs and assess their predictive power
   - This exercise was completed using a 'hold-out' sample as previously defined in conversations with the RIS and KAI data teams.
7. The final stage in the process was to complete the handover of models to HMRC for an internal 'test and learn' process along with the model rules so they might be replicated on subsequent data sets.

So in summary, on completion of the work and internal consultation and development work with HMRC project teams we built nine different models. Whilst we had a significant overall sample for the research, the small segment sizes for some SME segments made the modelling very challenging. However, the final results produced a set of nine models which could be tested by HMRC to enable better targeting of likely SME behaviours and segment

membership from HMRC taxpayer data and also supplement this with Experian data to increase model accuracy.

The **meta-segment models** were:

- Willing and Able, Payment Deferrers and Potential Payment Deferrers
- Unaware and Willing but Need Help
- Rule Breakers and Potential Rule Breakers

The **dimension models** were:
- Awareness (of obligations to HMRC)
- Acceptability of evasion
- Ability (to meet obligations to HMRC)
- Perception of risk (of non-compliance with obligations to HMRC)
- Payment prioritisation (whether HMRC would always be paid on time and ahead of suppliers)
- Ongoing cash-flow difficulties (whether or not the business was affected by this situation).
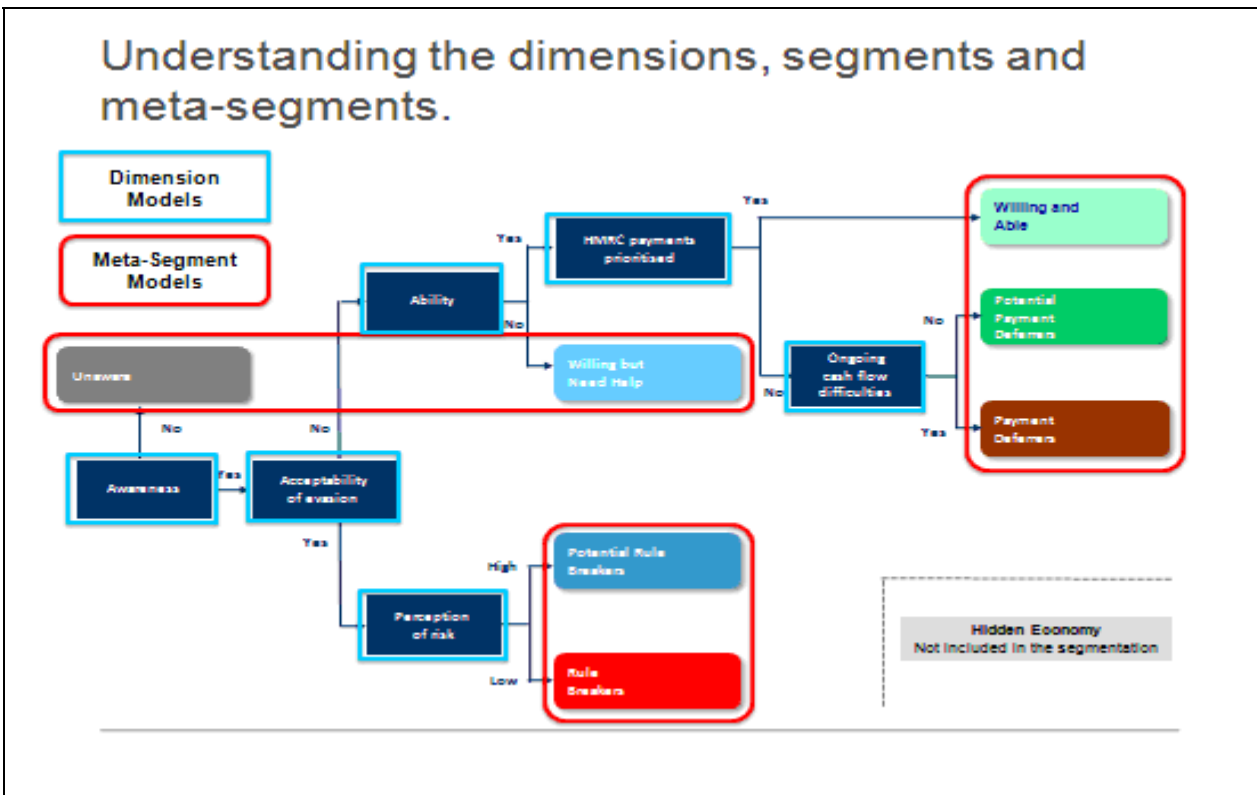
Determining the success of the modelling

The work always had an exploratory dimension to it, until we got into the work it was not 100% possible to understand the relationships between HMRC and Experian data and the attitudinal data collected through the sampled survey.

The models, as a rule, perform well. Naturally some do better than others but all results are broadly positive. It is worth cautioning that using customer data to predict attitudes is notoriously difficult and particularly this form of customer data which is primarily derived from HMRC Self Assessment forms and Third Party Experian data. In terms of the work we defined success as being able to predict a segment or dimension considerably better than random.

Understanding the limitations in the modelling work

There are several limitations in the models produced and these can be summarised as:
- Most models require the provision of some Experian data (for optimal performance).
- While the meta-segment models generally perform well it is not possible to build the seven segment models (due to sample size considerations as detailed above).
- As with all models of this nature they are better at predicting extremes than exact scores.

Operationalising this insight: thinking about potential uses.

On completion of the work, the work also determined a number of areas in which HMRC might consider applying the insights generated through the work. The model uses are limited only by the need to match to a **Unique Taxpayer Reference** for wider application and usage.

Thus any activity whereby the customer is known provides a potential opportunity to deploy the models in addition to existing tools, for example to inform;
- Campaigns via direct mail e.g. risk or support targeting;
- In-bound calls e.g. understanding a need for 'enhanced support';
- Prioritised support levels and channel allocation e.g. directing call traffic to lowest cost, most appropriate support options (supporting move to online self-serve as a preferred option).
- Targeting of interventions e.g. feed in to risk or potential Debt Management & Banking (DMB) targeting work.