Using big data to model farmers' behaviour

Determinants of non-compliance with the Nitrates Regulations

Abstract

Compliance with environmental regulations is an important area of application for economic psychology. We study adherence to the Nitrates Regulations in Ireland, which have been in place since 2006. Even though they led initially to a substantial improvement in farmers' behaviour, around 2,000 farmers (1.5-2% of all Irish farmers) are still found to be in breach of the regulations each year. This occurs despite financial penalties for non-compliance (ranging from 1 to 200% deduction from the EU Basic Payment Scheme). This pattern suggests that financial incentives may not be effective in changing farmers' behaviour and that other psychological factors may be at play. In order both to better enforce the regulations and to assist compliance, it is necessary to understand what these other factors are.

We make use of "big data", provided by the Nitrates section of the Department of Agriculture, Food and Marine (DAFM), totalling 1.65 million observations compiled from multiple data files. We build a statistical model to shed light on motivations for (non-)compliance. In addition, we use the model to predict who the most likely violators are, with the aim of targeting interventions to increase compliance. To ensure that our targeting is as accurate as possible, we also apply machine learning algorithms to our dataset and compare performance against our statistical model.

This dataset contains farm-level data on nitrate emissions, penalties incurred, farm and farmer characteristics (e.g. farm size, legal characteristics, age). It also has data on the conditional "derogation", which farmers can apply for early each year and which sets their nitrates limit higher in return for compliance with additional nutrient management rules. Hence, decisions relating to obtaining a derogation and to compliance are intertwined. Thus, we perform not only logistic regressions to model likelihood of being in breach of the regulations, but also Heckman selection models to model selection into the derogation simultaneously.

We find that compliance is less likely among farms that are smaller, had higher nitrate emissions the previous year, display greater volatility close to the regulatory limit and have breached more often in the past. Quite surprisingly, we find no significant relationship between penalty size the previous year and the likelihood of breach in the current year. If anything, the relationship is positive. An increased likelihood of getting a derogation is strongly associated

with bigger farms, younger farmers and larger and more volatile nitrates levels. The derogation decision also has a self-reinforcing element – those with a derogation the previous year are more likely to obtain it again. These findings are not predicted by standard economic theory and open a discussion on the motivations for compliant behaviour.

Predictive performance of our models was tested by predicting violators for 2017 (the newest year for which data is available) and comparing predictions with actual data. The same procedure was used to assess predictive performance of the machine learning algorithms – random forest and gradient boosting machine (both based on decision trees). Parameters were fine-tuned and corrections made to solve the class imbalance problem (the fact that only 2% of our sample are violators). Performance was compared between our statistical models, the machine learning algorithms and an "old" targeting rule used in DAFM, which is a rule-of-thumb that correctly captures around 60% of violators when selecting more than 16,000 farmers. Our models improve on this substantially. To "catch" the same proportion of violators, we need to target 3,000 less farmers, saving administrative costs. Alternatively, we can target the same amount of farmers, but correctly predict around 7 percentage points more violators, hence capturing more than one-in-six who previously escaped targeting. Our statistical models and the machine learning algorithms had closely similar predictive performance.

In summary, compiling and analysing a very large administrative dataset can lead to better enforcement of regulations. It offers a deeper understanding of farmers' behaviour, more precise targeting of potential violators and hence the opportunity to design more effective interventions. It is also a case study that demonstrates the usefulness of behavioural science in regulatory environment.