Incentivising Compliance with Financial Regulation

December 2016

Occasional
Paper 25

Encouraging debate among academics, practitioners and policymakers in all aspects of financial regulation.

Financial Conduct Authority
www.fca.org.uk

FCA Occasional Papers in financial regulation

The FCA Occasional Papers

The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Occasional Papers, extending across economics and other disciplines.

The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact Peter Andrews or Kevin James at peter.andrews@fca.org.uk and kevin.james@fca.org.uk

Disclaimer

Occasional Papers contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that Occasional Papers contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

Authors

Christopher Heady – University of Kent.

Gareth D. Myles – University of Exeter

Acknowledgements

We would like to thank Peter Andrews and Peter Lukacs for the significant input they have made into the writing of this report, Nick Catton for insight into tax gap measurement at HMRC, and participants at an FCA seminar who made helpful comments on our initial thoughts.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 706 60790 or email publications_graphics @fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 25 The North Colonnade, Canary Wharf, London E14 5HS.

Contents

1	Introduction	7
	Introduction	7
2	Economic Incentives	10
	Designing Incentives	10
	Incentive in Tax Compliance	12
	Behavioural Economics	16
3	HMRC Experience	18
	Organisation of HMRC	18
	The Tax Gap	19
	Large Business Service	21
	Recent Anti-Avoidance Measures	23
	Risk Assessment	24
	Litigation and Settlement Strategy	25
	Review of Process	26
4	Cooperative Compliance	29
5	Implications for Financial Services Regulation	33
	Regulating Financial Services	33
	Detecting Regulatory Avoidance in Financial Services	37
	Measuring Non-Compliance in Banking	39
6	Conclusions	40
7	Appendix	41
	A1. The Principal-Agent Model	41
	A2. Tax Compliance	42
8	References	48

Foreword

The FCA is delighted to be publishing this paper by two leading academics on the economics of tax systems, Gareth Myles and Chris Heady. We decided to consider what lessons we might learn from tax authorities because in some ways they have a harder job than financial regulators in promoting compliance. First, tax payments are a pure cost with no direct benefits for businesses making the payments. Second, tax authorities on average know less about most taxpayers than financial regulators know about regulated firms.

Further, it was intriguing to learn that in the U.K., Her Majesty's Revenue and Customs ("HMRC") had instigated a special code of compliance for banks, one requiring banks' boards to sign up to some very specific undertakings. This suggested that tax compliance in banking might be especially difficult to achieve, further implying that we could draw lessons from HMRC's strategies to promote compliance and their implications for financial regulation.

The final paper contains some interesting, thought provoking and useful material for us.

First, as suspected, we see many parallels between tax authorities and conduct regulators. In fact, some initiatives taken by HMRC to promote compliance are like those in financial regulation. For example, the HMRC Large Business Service for individual relationship management with the largest corporates is like the FCA's "fixed" relationship supervision. Similarly, Senior Accounting Officer rules, introduced in 2009 to assign personal responsibility and non-compliance penalties to a senior manager in a taxpaying firm, has many parallels with the FCA's Senior Managers Regime for financial firms. Lessons can clearly be learned both ways.

Second, tax authorities undertake genuinely random audits of firms' returns, in addition to targeting specific firms identified through sophisticated statistical methodologies. Using a range of data about firms that were and were not found to be engaging in tax evasion, these methodologies identify combinations of characteristics that signal whether a firm with unknown tax behaviour is more likely than most to be a tax evader. These predictive analytics are considered materially to increase the pay-off from the scarce resources available for auditing activities. Similarly, the U.S. securities regulator, the Securities and Exchange Commission, has already used a system like this (text analytics) to target firms that misreport their financial condition. Overall, the results were so positive that the Securities and Exchange Commission is deploying such techniques in a wide range of applications, and we are now carrying out a pilot.

Third, the authors highlight an important difference between tax administration and financial regulation. Broadly speaking, only the aggregate tax payment of the firm matters. However, in financial services, mis-selling is likely to occur in individual client transactions. For the kind of detection described above, financial conduct regulators, such as the FCA, likely need to gather and use detailed, transaction-based data.

Fourth, introducing a "presumptive" ban on tax avoidance schemes is interesting. This deals with attempts to bypass the rules rather than violating them explicitly. Under the General Anti-Abuse Rule introduced in 2013, a scheme which reduces tax due is deemed abusive tax avoidance unless there is some commercial justification for it other than reducing tax liability. This approach might help in financial regulation to deter abuse of unintended loopholes. For example, firms may react to a ban or restriction on a particular class of products by introducing new products that legally do not quite fall under the definitions in the relevant rule but have the same economic purpose as the banned product.

Fifth, HMRC has collaborated with professional bodies such as those overseeing accountants and lawyers to set mutually reinforcing reporting requirements regarding tax schemes. External advisers on such schemes are given an incentive to report them to HMRC through the similar reporting requirement imposed on professional staff in banks who work on these schemes. And vice versa.

Sixth, while Myles and Heady focus on improving compliance through improved deterrence, they observe that even tax evasion can bring psychological costs of non-compliance whether or not it is detected. Feelings of shame and guilt are a psychological “punishment” for tax evasion, and these are likely to be stronger in societies where benefits of tax compliance are widely appreciated and evasion is considered unacceptable. When strong, these feelings reduce tax evasion. These ideas relate to “tax morale,” something authorities seek to build as it leads to strong voluntary compliance.

One aspect of the paper by Myles and Heady is interesting and thought-provoking and although it is not immediately evident how this would apply in the different context of conduct regulation. Broadly speaking, HMRC uses macroeconomic and other data to estimate how much tax ought to be paid and compares the result with the amount of tax paid. The difference is the “tax gap,” a measure of the HMRC’s success in incentivising and enforcing the payment of taxes due. Our initial thinking is that we have no comparable metric to use to assess our performance. We do, though, continue to consider this issue and are formulating plans to increase how we measure the practical results of our interventions.

Peter Andrews

Chief Economist, Financial Conduct Authority

Summary

The regulated U.K. financial services sector is a large and dynamic industry. In 2015, it consisted of approximately 56,000 firms and 125,000 approved persons. Moreover, around one thousand new firms were authorised by the Financial Conduct Authority (FCA) in 2015. By some estimates, the industry contributes 12 percent of U.K. gross domestic product. The insurance industry alone held £1.9 trillion of invested assets.

The FCA is responsible for regulating the financial services firms that supply to retail and wholesale consumers in the U.K. The FCA, along with the Prudential Regulation Authority, was established in 2012 to replace the Financial Services Authority. From 1 April, 2014, the role of regulating the consumer credit industry was transferred to the FCA from the Office of Fair Trading. The FCA's main roles are to maintain and ensure the market's integrity, to regulate financial services firms so they provide consumers with a fair deal, and to ensure that the financial services market is competitive. The FCA is not a government body; it is accountable to HM Treasury but its funding derives from the fees paid by the firms it regulates.

The financial services industry is clearly in need of regulation. The failure of firms to comply with regulations is exemplified by many well-known episodes of mis-selling of financial products. There have also been many other instances of non-compliance. In 2013/14 the FCA imposed 46 penalties totalling £425m and secured criminal convictions against several individuals. The perceived magnitude of non-compliance has led to the view that financial services firms are characterised by a culture of non-compliance and that only a cultural shift will improve the regulatory outcome.

At the heart of any regulatory process is providing incentives that encourage regulated firms to comply. This paper reviews the basic economic theory of incentives and how they can be used to enhance compliance. It considers incentives theory as applied to tax compliance since it has many similarities to compliance with financial regulations and a much longer history of effort by authorities to improve it. This starts from the standard perspective of individual economic rationality and progresses into the additional insights offered by behavioural economics. The approach of HMRC to reducing non-compliance is then reviewed with a focus on recent initiatives aimed at culture change in large businesses. Introducing relationship managers and developing cooperative compliance are both explained as methods of shifting corporate culture and incentivising compliance.

The analysis of HMRC activities is then used as a basis to explore developments in enforcement and supervision at the FCA. The FCA faces similar issues to HMRC and it has implemented many similar programmes. Securing culture change in financial services firms is important to the FCA. The Senior Managers Regime is a central component in the culture change strategy but there are also programmes aimed at ensuring the operational structures of firms deliver good conduct. These programmes indirectly address compliance failures, such as mis-selling, by encouraging compliance. What they do not do is provide direct control of mis-selling. Many past episodes of mis-selling have continued for long periods before problems have been recognised. This has led to many customers being affected and compensation processes that take years to complete. Real-time monitoring of financial services transactions could directly reduce the extent of mis-selling if technological challenges can be overcome. Mis-selling is typically individual-specific (i.e., it depends on whether a particular product is suitable for an individual consumer, given the consumer's circumstances), so it requires the analysis of a large body of information about each transaction. This creates greater complexity than in individual tax compliance and poses a continuing challenge to regulation.

An important innovation of HMRC was to introduce the "tax gap" as a measure of non-compliance. The tax gap indicates the HMRC's success at achieving compliance and, since it is a consistent measure over time, can indicate compliance trends. It also provides insight into the sources of non-compliance. The tax gap can be apportioned to tax instruments and to taxpayer groups. Each of these indicates where compliance resources should be focused. There is no similar aggregate indicator of compliance with financial regulation. If the FCA developed a compliance index, it would provide much clearer information about the state of compliance than judging it from the number and level of fines.

1 Introduction

Introduction

The financial industry has recently been through some difficult periods. The financial crash of 2008-9 was triggered by several prominent financial institutions collapsing and widespread fear of systemic failure. Only very drastic action, most notably by the U.K. and U.S. governments, prevented a catastrophic outcome. The financial crash highlighted on a grand scale the consequences of inappropriate selling of financial products. Low-income households in the U.S. were sold mortgages that they could not support, and the mortgages were securitized and dispersed through the system. The securitized instruments were rated as safe but proved anything but once mortgage defaults began to accumulate. The U.K. has also had significant instances of mis-selling. The payment protection insurance (PPI) and the Allied-Dunbar pension episodes affected millions of U.K. consumers and severely affected confidence in the financial services sector.

These events point to two perspectives from which the financial services sector is regulated.[1] First, the financial sector's stability across the system is the key focus. The central issue is to ensure prudential regulation of banks and other financial institutions to ensure that they hold sufficient reserves to meet potential liabilities. Second, at the micro level, regulation seeks to ensure the appropriate market conduct of financial institutions. This form of regulation seeks to ensure that institutions do not exploit market power or supply financial services inappropriate for the consumer or misrepresented when sold. As the financial crash so startlingly illustrated, financial stability and market conduct are not independent: stability can be threatened by the consequences of mis-selling.

The financial regulation developed in the U.K. reflects these perspectives. From 1997 through to 2013, the responsibility for regulation was held by the Financial Services Authority (FSA) with the Bank of England also expected to monitor system stability. The failure of both institutions to prevent the financial crises led to abolishing the FSA and reallocating responsibilities. From April 1, 2013, the Prudential Regulation Authority (PRA) was charged with monitoring the financial condition of major financial service firms. At the same time, the FCA became the regulator of market behaviour. Finally, the Financial Policy Committee at the Bank of England became responsible for overseeing the financial system's stability with authority to issue instructions to the PRA and FCA.

The FCA is independent of the U.K. Government and is financed by charging fees to financial services industry members. It regulates all firms that provide (most types of) financial services to consumers at both retail and wholesale levels. Its mission is to maintain the financial market's integrity. The FCA can impose minimum standards on financial products and has the power to investigate organizations and individuals. It can also ban the sale of financial products and instruct firms to modify them or remove them from sale. From April 1, 2014, the FCA took on the additional role of regulating the consumer credit industry. The practical challenge faced by the FCA is to adequately monitor the thousands of firms operating in the financial services market and to provide the firms with an incentive structure that generally ensures compliance.

Since the financial crash, a considerable literature has developed on prudential regulation.[2] The issues faced by a prudential regulator are clear. Financial institutions must satisfy capital

[1] There is also a third: the efficient markets perspective

[2] See Galati, G. and Moessner, R. (2013).

adequacy requirements to prevent collapse in adverse circumstances. Moreover, a prudential regulator must be aware of systemic risk caused by externalities between institutions both directly and via the positive correlation of adverse risks. Questions remain unresolved about how to quantify systemic risk in a network[3] or institutionally[4] but the theory of prudential regulation is a developed and coherent body of literature.

In contrast, the theory of regulating financial services to maintain the market's integrity has received much less attention. This is despite the continuing revelation of malpractice within the financial services industry.[5] The challenges facing the FCA are significant. The number of companies[6] and products, allied to continuous innovation in products, is itself sufficient to create considerable difficulties in monitoring the market. These difficulties are compounded by a market culture in which disregard for the rules is perceived in some areas to be the norm. In addressing the challenge of incentivising the financial services industry to comply with the rules, the FCA can undertake costly monitoring of firms and products, and impose reporting costs upon firms. However, firms that flout regulations may adopt a similar attitude to honest reporting. Imposing regulations generates both benefits and costs. How rigorous regulation should be, and what level of compliance costs are justified, is a question that can only be answered by a proper cost benefit analysis.

The issues faced by the FCA have many parallels in attempts by HMRC to encourage tax compliance. Collecting taxes is essential for government financing. If the tax system is poorly designed or system administration is flawed, then taxes can also cause immense frustration and resentment. Few taxpayers enjoy paying tax but most, at least in the U.K., will attempt to abide by the rules. Other taxpayers will attempt to push the rules to minimize tax obligations and some who simply ignore the rules. The task facing HMRC, and indeed revenue services worldwide, is to implement a process of administration that makes compliance as painless as possible for the majority, incentivises the tax minimizers to moderate their behaviour, and detects and punishes the rule-breakers. The range of tax instruments and the legislation's complexity make this a very demanding task. In response, HMRC has developed policies and interventions targeted at incentivising increased compliance.

This paper will review recent developments in HMRC compliance strategy focusing on the approach taken to large businesses. It reports and explains the strategies with reference to relevant economic theories. The intention is to draw out the underlying themes so the strategies can be given a context. This should make it possible to see what may be usefully applied in other areas where improved compliance with regulations is desired, particularly in financial services.

The paper begins in section two with a review of economic incentives. The first section considers the general problem of incentives in economics and describes the principal-agent model that captures the fundamental features of all incentive-design problems. This is followed by analysing incentives in regulatory compliance and the legal system's interaction with the incentive system. The final part of the section reviews recent developments in behavioural economics and how these enhance our understanding of behaviour. Behavioural economics is best seen as a complement to traditional economic analysis, increasing the set of factors relevant in decision-making. A new set of incentive tools has also emerged from applying behavioural economics and the most important of these are described. The analysis of this section is supported by a formal presentation of the material in the appendix.

The third section of the paper reviews the recent actions taken by HMRC to improve compliance so lessons can be drawn for financial regulation. The issues involved in financial regulation have strong parallels with those faced by HMRC in incentivising tax compliance: there are many taxpayers, some of which have a very poor attitude toward compliance, and it is costly for HMRC to observe their actions. HMRC's strategies are continually evolving as new challenges emerge

[3] See Gong, R. and Page, F. (2016).

[4] See Danielsson, J., James, K.R., Valenzuela, M., and Zer, I. (2016).

[5] A long list of FCA cases are presented at www.fca.org.uk/news/speeches/financial-services-regulation-enforcement.

[6] The FCA regulates approximately 50,000 firms (www.fca.org.uk/static/fca/documents/factsheet.pdf).

so there is considerable recent innovation to discuss. The first part provides a brief description of HMRC to give institutional context. The second part introduces the “tax gap,” a key tool to assist HMRC with understanding the issues that they face and for monitoring of performance. Components of the compliance strategies are then discussed including the Large Business Service, the “general anti-abuse rule” and the Litigation and Settlement Strategy.

A new approach to the relationship between large corporate taxpayers and revenue services is discussed in section four. Tax administrations across the world face similar compliance issues to those faced by HMRC. International bodies such as the Organisation for Economic Cooperation and Development (OECD) promote initiatives for improving tax administration. One current initiative intended to increase compliance by large business is the promotion of cooperative compliance.

The purpose of the review of HMRC strategies is to draw out practices that may transfer to other areas of regulation. The FCA faces challenges that, in some dimensions, are like those faced by HMRC. The regulations it supervises have been put in place with the intention to protect both sides of the market. However, if compliance with regulations is costly, then some firms will choose not to comply. The FCA needs to incentivise the compliance of these firms with the regulations. The discussion in section five reviews the implications of HMRC experience for the regulatory role of the FCA.

2 Economic Incentives

Designing Incentives

Economics is often described as the science of making choices when faced with scarcity of resources. This description captures the observation that consumers, firms, and governments need to take choices but have limitations, or “constraints,” upon the options that can be chosen. These limitations can be due to what is affordable with a given income or available for a fixed access to resources. They can also include subtler constraints such as limits on the availability of information or the capacity for processing the available information.

To make a choice, a person (used here to refer to a consumer, or a firm or a government) determines the feasible choices permitted by constraints, confronts the feasible choices with preferences, and selects the preferred option. This is the process of “maximization subject to constraint.” The choice will depend on both preferences and the range of feasible choices. If either of these change, the choice will likely change. Certainly, if the range of feasible choices expands to include new and better options, then the old choice will be dropped and one of the new options will be selected. Similarly, if the set of options changes only so the original choice is no longer available, then a different (but less preferred) option must be chosen. These ideas about how choices can be changed by altering the set of feasible alternatives are central to understanding how incentive schemes operate.

In this discussion, an incentive is any intervention into the choice process that raises the relative benefit of choosing one option, X, relative to another option, Y. A positive incentive will raise the benefit from choosing X. For example, an additional benefit may be given if X is chosen or the cost of X may be reduced. Either of these will reward the choice of X. A negative incentive will reduce the benefit from choosing Y. For example, the cost of Y can be increased or a punishment can be inflicted with some known probability if Y is chosen. Implementing either a positive or negative incentive will increase the likelihood that X is chosen over Y. By making the incentive sufficiently large, it is possible to guarantee that X will be preferred to Y. However, a large negative incentive (such as a large punishment with a high probability) might cause Z to be chosen instead of X or Y,[7] whereas a large positive incentive will be costly to finance.

Incentives are offered to get a person to choose one option instead of another. The reason for wishing to see a particular option chosen is an external benefit arising from the person making the preferred choice. There are many examples of situations with external benefits: the owners of a firm want the managers they employ to make choices that benefit the owners rather than the manager, or the regulators of an industry want the firms in the industry to behave in a socially-beneficial way rather than just being driven by profit.

Reflecting on these ideas has led to the theory of incentive contracts in economics.[8] This theory explores how incentive contracts should be designed and the limitations of what can be achieved. The basic idea is usually that we want to motivate the agent but cannot directly order them to do one thing rather than another. We cannot simply order the behaviour we wish to see because some element of the agent’s actions is unobservable or context-specific. In the context of financial regulation, the auditors employed by a financial services firm have the responsibility for “gathering sufficient and appropriate evidence to form an independent opinion about the

[7] In terms of the principal-agent problem discussed below, “Z” could be the agent choosing to work for another firm.

[8] The theory is reviewed in detail in Laffont and Martimort (2001) and various applications are contained in the January 1999 issue of the Review of Economic Studies.

regulated firm's assertions on those financial statements."[9] Although the auditors may be closer to the firm than the regulator, actions can still be unobservable to the auditor. Hence, auditors cannot eliminate the incentive problem.[10] It is impossible to monitor an unobservable action or to write a contract which requires the unobservable action to be set at a given level.

These ideas have been formalised in the context of the principal-agent problem. A principal-agent relationship arises when one party wishes another to undertake an act on their behalf. If the principal cannot directly observe that act undertaken and its consequences can only be observed with some random error, then moral hazard can enter the relationship between the principal and agent. That is, the agent can hide behind the randomness and take an action that they find less costly but which yields a lower expected return to the principal. As noted in Mirrlees (1974a, 1975/1999) such relationships potentially arise in any economic situation with contingent contracts. For example, they arise between owners and managers of firms, between insurance companies and the insured, and between regulators and regulated firms. The problem facing the principal is to structure the contract offered to the agent so it gives an incentive to the agent to take the action (conditional on the realization of the event that is unobservable to the principal but observable to the agent) that yields the best possible pay-off to the principal.

The logic can be explained by considering the example that motivated the literature. The owner of a firm (the "principal") employs a manager (the "agent") to run the firm. The manager can supply a low-effort level or a high-effort level. The owner of the firm cannot observe the effort level. The owner does observe the profit level of the firm which can be high or low. The level of profit is determined by random factors but if the manager supplies a high-effort level it increases the probability of securing a high profit and reduces the probability of low profit. The owner must choose the best contract to offer the manager.

The only variable on which the contract can condition payment is the observable profit level. Consider what would happen if the owner were to offer the manager a fixed salary no matter what profit level was observed. In this case, the manager would choose a low-effort level and the probability of high profit would be at its lowest level. This is not to the benefit of the owner. Clearly, it is necessary to pay the manager more when profit is high and less when it is low to provide an incentive to supply high effort. In choosing the amount, the owner must confront two constraints. First, if the payments are not high enough, the manager will refuse the contract and work somewhere else. Second, if the payments are too high, the owner will earn a lower return than is possible. The choice of payment level must balance these two competing constraints.

A third constraint is involved. It is standard to assume that the manager does not like to face risk in payments. That is, the manager would prefer to be paid a constant amount rather than an amount that varies according to the level of profit. The expected payment that the owner makes therefore must increase the more they differentiate the payment with low profit from the payment with high profit. Consequently, the two payments should be differentiated just enough to encourage the manager to supply high effort rather than low effort. Finally, the owner then must ensure that adopting payments satisfying these three constraints is better than accepting that the manager will supply a low effort level and paying just sufficient to ensure they do not move to an alternative employer. If this is the case, the payments then constitute the optimal incentive contract.

The logic of this example forms the basis for the economic analysis of incentives. The key elements are that: contracts must be conditioned on publicly observable or verifiable information and that some relevant facts are not publicly observable. The asymmetry of information between the agent and the principal creates the incentive problem.

It is worthwhile at this point to consider how this model relates to the regulation of financial services. The model is designed to cut out unnecessary detail so it can focus on the essence of the problem—namely, how to provide incentives to motivate actions that cannot be observed. The

[9] www.fca.org.uk/static/documents/finalised-guidance/fg13-03.pdf

[10] There is also a further issue about the incentives confronting the auditor. See OECD (2004).

actual situation of financial regulation is considerably more complex than the model described, so interpretation involves condensing many details of reality into the limited number of concepts in the model. The principal is the regulator and the agents are the financial service companies. The principal benefits if the companies behave per the agreed rules of market conduct. In this sense, the “profit level” used as the objective of the principal in the standard principal-agent model is replaced by a measure of the goodness of conduct (or degree of compliance) perceived by the regulator. The actions of the agents summarise all the factors that affect the goodness of conduct: the financial products traded, the treatment of customers, and the competitiveness of behaviour. No literal payment is made by the regulator in the financial market. However, the payment made by the principal to the agent can be understood as an aggregate of the compliance costs imposed on financial services firms, the tightness of imposition of rules, the frequency and intensity of financial scrutiny, and (from the competition perspective at least) the permissible level of returns. The asymmetry of information arises for financial services because the regulator cannot feasibly observe all actions of all firms.

A further issue that arises in principal-agent relationships is determining conditions that guarantee payment increases as the observed output increases. This happens easily with only two possible profit levels, such as the high and low profit levels in the example. Surprisingly, with three or more levels of output, there can be no guarantee that the optimal incentive contract will have a payment that increases with output — see Grossman and Hart (1983). Mirrlees (1976) showed that if costlier actions for the agent make more profitable outcomes more likely, then payments will increase with output. In interpreting this model for financial services, the profit levels are the degrees of compliance as viewed by the regulator.

This brief review of incentives has observed that choices are restricted through cost or resource availability and that an incentive changes the relative benefits from alternatives. A successful incentive contract ensures that a choice desired by the principal is selected by the agent. For financial services, this will involve the financial services firms complying with regulations to the extent that the regulator finds acceptable. The central challenge in designing incentive contracts is the principal’s inability to observe information relevant to the outcome. This asymmetry of information restricts the incentive contracts that can be implemented since the contract can only depend on observed information. Consequently, a perfect incentive scheme is rare. For example, for the original problem of a manager supplying effort, it is impossible to construct a contract that gives them the incentive to supply the most efficient effort level in every eventuality. Similarly, it implies that financial services firms cannot be incentivised to achieve perfect compliance. However, it is possible to provide incentives for the agent to make a preferred choice over what would have been made without the incentive scheme. This is the benefit of the incentive scheme perceived by the regulator.

Incentive in Tax Compliance

The economic theory of tax compliance views the taxpayer’s decision on whether to accurately declare their income as one of choosing between a certain net income (that would arise if they declared their income correctly) and an uncertain net income (if they underdeclared their income and faced the possibility of detection and penalty). This was first formalised as a problem of decision-making under uncertainty by Allingham and Sandmo (1972). This led to a long series of papers by economists and psychologists analysing both theory and evidence. The initial literature was developed using the traditional economic model of a self-interested individual. Researchers realised, however, that this model was inadequate to describe observed behaviour and the results of tax compliance experiments. This led to adoption of a behavioural perspective on the tax compliance decision. The behavioural perspective does not change the basic assumption of maximising behaviour but focuses on generalising the objective of the taxpayer and the factors that affect the objective. The behavioural perspective is discussed in detail below.

A taxpayer choosing not to report all income to the revenue service faces the possibility that they will be subjected to an audit that will discover some, or all, of the unreported income. This is a similar situation to a financial services firm choosing not to comply with regulations and taking the risk of being subjected to an investigation. The taxpayer does not know if an audit will take place, so can do no more than assess how likely an audit will be. Formally, the taxpayer will assign a probability to the event of an audit taking place. The choice of how much income to report then considers the probability of an audit and the corresponding probability of avoiding an audit. Consequently, from the perspective of the taxpayer the outcome of a false report is random. How the taxpayer behaves is determined by the attitude toward accepting risks, which is a feature of their preferences. Some taxpayers will be quite willing to take on risk so will be very non-compliant. Others will strongly prefer to avoid risk so will engage in very limited non-compliance. Although the chosen level of compliance will differ between taxpayers, the direction of response to a more rigorous compliance system (a higher probability of audit or a greater fine if discovered to be non-compliant) is always toward increased compliance. The constraint upon the taxpayer is determined by how much income they earn. The amount of income not declared must fall between zero (complete declaration of all income) and total income (no declaration or a declaration of zero income). The amount not declared is chosen between these limits to achieve the highest ranking in the preferences of the taxpayer.

What matters to the taxpayer is the sum available to spend. If there is no audit, the amount is given by income before tax, less the tax due on the amount of income declared. When an audit occurs, the amount available for spending is equal to income less the total tax due less the fine levied as a penalty on the attempted tax evasion. The policy variables potentially available to incentivise better compliance are the tax rate, the rate of fine, and the probability that an audit is conducted. For financial services regulation, the analogue of the probability of audit is the intensity of supervision. How changes in these variables affect the compliance choice of the taxpayer depends on the precise details. In the economic model of compliance formalised by Allingham and Sandmo (1972), the penalty was a multiple of income not declared. This was changed by Yitzhaki (1974) to make the penalty a multiple of the evaded amount of tax, which is approximately the practice in the U.K. and U.S. In the modified version, the amount of non-compliance is reduced by an increase in the size of the penalty and by an increase in the probability of audit.

This model of the compliance decision has been used to explore the design of incentive schemes to increase the degree of compliance. The two central variables that can affect incentives are the probability of an audit taking place and the rate of fine levied on unpaid tax. In terms of the earlier discussion, both are negative incentives. The first discussions of this question assumed that the revenue service could control both the fine and the probability. Observing that the fine is costless to increase but the probability is costly (more tax inspectors must be hired) leads to the conclusion that the fine should be increased without limit, and the probability reduced to close to zero. In the words of Kolm (1974), the incentive scheme should be to “hang tax evaders with probability zero.” This finding is extreme but it helpfully illustrates the need to think carefully about the constraints that operate in the design of an incentive scheme. It is unrealistic to assume that the revenue service controls the fine levied for non-compliance. From a narrow perspective, U.K. tax legislation places an upper limit of the rate of fine that HMRC can levy. Taking a broader perspective, the penalty for tax evasion is one component of the criminal justice system in the U.K. The economics of crime explains why punishments should be graduated according to the seriousness of each crime, and there are good reasons for limiting the penalty for evasion. Consequently, the correct model of how HMRC can set incentives should take penalties as hard to change and focus mainly on how to select audits.

The FCA can also levy fines and has levied some very large penalties for misconduct. Barclays were fined £284m in 2015 and UBS £234m for foreign exchange failings. Fines of similar magnitude have been levied on Deutsche Bank, Citibank, JPMorgan Chase Bank, RBS and HSBC. These fines and publishing offending parties’ names provide a clear signal that the FCA levies significant punishments and has greater flexibility than HMRC in choosing the level of punishment.

The quantity of resources available to a revenue service to conduct audits is limited, so it is impossible for all tax returns to be audited. The revenue service decides how to allocate auditing effort across tax returns. The simplest analysis of this issue assumes that taxpayers differ in income and that the only information on a tax return is a declaration of the income level. The problem facing the revenue service is that a return with a very low declaration may come from a taxpayer who truly has a low income or from a high-income taxpayer attempting a significant level of evasion. The revenue service would like to catch the latter, but does not wish to waste audit resources looking at the former. This is the trade-off that is faced in the auditing decision.

How the trade-off is resolved depends upon the choice of the revenue service to commit to an audit strategy or to undertake a responsive audit strategy. Commitment would involve a clear statement of the audit strategy and a belief among taxpayers that the announced strategy will be followed. In this situation, Scotchmer (1987) showed that the optimal incentive policy takes the form of a cutoff rule: income reports below the cutoff level stand a chance of being audited but reports above the cutoff are not audited at all. If taxpayers believe this strategy will be implemented, it is rational for all taxpayers with incomes above the cutoff to file a report at just above the cutoff, safe in the knowledge that they will not be audited. The revenue service then has an incentive to renege on its announcement and to audit the taxpayers who have announced the cutoff as their level of income. Only if there is commitment not to do this can a cutoff rule be announced and believed. A more likely scenario is that the authorities will reduce the probability of audit for people declaring higher incomes, but taking account of other characteristics of the taxpayer. Also, in practice, most authorities do not commit to a public audit strategy.

The situation for the regulation of financial services falls between the two alternatives. The FCA has made a public announcement of its supervisory strategy,[11] and the extent of supervision depends upon the “conduct classification” of the firm. The classification is determined by combining two indicators: metrics that measure the firm’s potential impact on FCA objectives, and the number of retail customers served by the firm.

The basic model of tax compliance described above has been developed in many different directions. At a theoretical level, some have analysed how taxpayers construct beliefs about the risk of detection, different ways in which the penalty might be related to the extent of underdeclaration and different approaches to how people make decisions when facing uncertainty. Alongside the theory, evidence analysis has made substantial use of laboratory experiments in which individuals (usually students) were asked to behave as if they were taxpayers deciding whether to underdeclare their income and, if so, how much income to hide. Much of the theoretical analysis concerns modifications that would enable the model to more accurately replicate what was already known about the extent of tax evasion. Meanwhile, the laboratory experiments confirmed the theoretical prediction that increasing the size of penalties for evasion would reduce the amount of evasion. Surprisingly, weaker empirical support exists for the theoretical prediction that a higher probability of a tax audit (and thus of detection) would reduce the amount of evasion (Choo, Fonseca and Myles, 2016).

The basic model can be developed in several ways to help think about how tax compliance could be improved. Developments of this model also suggest how an adaptation can be made to address regulatory compliance more generally. The potential developments include[12]:

- third-party reporting of income
- the design of efficient audit strategies
- a wider view of penalties for evasion

[11] See www.fca.org.uk/static/fca/documents/factsheet.pdf.

[12] See Hashimzade et al. (2013) and Slemrod and Yitzhaki (2002) for further details.

The U.K.'s system of tax administration, like those of many other countries, makes extensive use of information provided by those who make payments to the taxpayer. These include reports of earned income submitted by employers, reports of interest income by banks and building societies, and reports of dividend payments from corporations. Evidence from audits in the United States shows that compliance is very much higher for income sources subject to third-party reporting than for other sources of income. This allows audits to be more focused on people who receive income not subject to third-party reporting.

For financial services, third-party reporting would involve the consumer transmitting information about the purchase to the FCA but the match is not exact. With taxation, there are only two relevant facts: that a transaction occurred, and the value of that transaction. The situation is more complex with financial services, since much more detail is required about the information flows between the financial services firm and the customer when the transaction was agreed. This point is discussed in detail below. However, the main point is always true: the regulated outcome is improved when more information can be received and processed. Knowing that information is flowing to the regulator deters non-compliance, and increased information permits improved focus of supervisory activities.

Generally, it is possible to improve the targeting of audits by undertaking a random audit programme to identify the characteristics of taxpayers who evade. It is important that these audits are random, so the sample is representative and thus able to both estimate the prevalence of evasion among groups of taxpayers and identify unexpected types of evaders. Thus, several countries (including the U.K.) conduct a random audit programme alongside the main targeted audit programme. The results of these audits can be used to further refine the identification of "high risk" taxpayers. The same arguments can be used to support supervisory visits in financial services regulation to randomly-selected firms. The key is the randomness – this ensures the information collected is statistically valid as a sample of the data without systematic bias.

Turning to penalties, the economic model normally assumes a financial penalty and, indeed, most countries (including the U.K.) mainly use administrative financial penalties to deter and punish tax evasion. However, there can be formal prosecution if the taxpayer refuses to settle or the case is extremely serious. In extreme cases, a prosecution can result in a prison sentence.

In addition, feelings of guilt, shame and stigma can be psychological "punishments" for tax evasion; and some can arise even if the evasion is not detected. This raises the question whether the names of taxpayers who have been shown to evade taxes should be made public, even if they have not been prosecuted. This could possibly improve compliance. These feelings for the taxpayer can be translated into potential reputational damage for a financial services firm if its products are not compliant. This point has been clearly recognised by the FCA and culture change has been key initiative to encourage compliance. In the words of Martin Wheatley (FCA chief executive until September 2015): "Culture change begins at the top, and to drive this we must make sure senior individuals in positions of responsibility are held accountable for the decisions and behaviour carried out by themselves and the people that they are managing."[13]

Further, these psychological "punishments" are likely to be felt much more strongly in a society in which the benefits of paying taxes are widely appreciated, and tax evasion is seen to be socially unacceptable. The extent of social consensus against tax evasion is often called a country's "tax morale." The parallel to this argument for financial services is the attitude of the firms to regulatory compliance. The evidence from tax literature is that a respected and supported regulation system will induce increased compliance. The same is almost certainly true in financial services, and a change in "regulation morale" would likely lead to a demonstrable increase in compliance. If correct, these arguments provide solid support for the FCA drive to secure culture change.

[13] www.fca.org.uk/your-fca/documents/second-annual-public-meeting-martin-wheatley-speech

The idea of tax morale has interesting implications, as it is reasonable to suppose it would be higher in countries with a higher level of compliance. Indeed, it could be argued that more tax evaders should be prosecuted and face the stigma that follows from that, even if the costs of prosecution are higher than the tax that will be recovered and the fines that will be levied. The rationale for this argument in favour of raising the level of enforcement beyond what a simple cost-benefit calculation would support is that the higher profile of punishments for tax evasion can raise tax morale by reinforcing the social consensus against tax evasion and making it clear to the public that tax evaders face serious punishment.

Behavioural Economics

Behavioural economics is a varied collection of ideas linked by the common theme of bringing new factors into the analysis of decision-making. It draws on ideas from psychology and sociology, reflecting evidence – often experimental – on choice behaviour. The insights of behavioural economics suggest opportunities for policy interventions.

The traditional model of choice assumes that the consumer cares only about the quantity of each consumed good. Behavioural economics has relaxed this in several ways. First, reference dependence assumes that it is not consumption quantity per se that matters, but the quantity relative to a reference point. The reference point may be some previous consumption level or some target consumption level. The consumer then cares about gains or losses relative to the reference point. Second, habit formation involves current preferences depending on past levels of consumption. This could, for example, capture smoking behaviour, in which past consumption makes future consumption a necessity. Third, external effects can link the preferences of one consumer to the consumption patterns of others. This can include network effects or envy.

The behavioural models are frequently described as capturing what some choose to call “irrationality”. This irrationality may come from myopia, so the consumer does not assign adequate weight to consequences. When the myopia involves failing to see all the present consequences of an action it is labelled an “internality.” A typical example is consuming food containing excessive fat or sugar, ignoring the health effects.[14] These models of irrationality lead to the idea of the consumer having two sets of preferences: one set of preferences are responsible for the myopic short-term choices and the other set are the “true” long-term preferences. Bernheim and Rangel (2005) call these, respectively, the decision utility and the experienced utility. The models are also perceived by some to justify calls for paternalistic government policies to help guide consumers to make decisions the Government believes are good for them. For example, immediate needs may seem so pressing for many consumers that they do not pay adequate attention to saving for retirement. Ensuring that consumers are automatically enrolled in a pension scheme and need to actively opt-out is a paternalistic policy that exploits inertia (opt-out will not take place) to raise the level of pension saving.

Behavioural economics has also provided new perspectives on decision-making. The common starting-point is the observation that a consumer (or any economic decision-maker) faces limitations on the information available when decisions are made and has limited capability in determining the consequences of a choice (“bounded rationality”). In either case, the outcome is that not necessarily the best choices will be made. The consumer may follow some rule-of-thumb or mimic the behaviour of peers. Decisions may be made based on beliefs about future events that are subjective rather than objectively determined through evidence. If beliefs are subjective (e.g., “I am an exceptionally good driver”) they have the potential to be very different from the truth (e.g., “I am an average to poor driver”) and lead to mistaken decisions (e.g., “I do not need comprehensive car insurance”).

[14] O’Donoghue and Rabin (2006) model the policy implications of this situation.

The ideas of behavioural economics are motivated by the large range of apparent “anomalies.” Anomalies are choices that cannot be explained by standard economic reasoning. As examples, people who know that smoking is bad for them continue to smoke, people spend money when they know they should be saving, and lottery tickets are purchased when the expected pay-off is negative. A behavioural explanation for the observation is always obtained either by appealing to non-traditional preferences or to the process for choice. No single, general and consistent solution explains all anomalies. We are left with several situation-specific behavioural models but not a general theory. Currently, the best response is to empirically explore each market with apparent anomalies to determine the most appropriate behavioural explanation.

The new models of behavioural economics have implied new motivations for economic policy and new ways of incentivising behaviour. As already noted, the fact that individuals may not make decisions in their own interests creates a role for paternalism. The Government (or a regulator designing an incentive scheme) can modify incentives so choices are made that perform better when measured using experienced utility. For example, introducing a “fat tax” would discourage the purchase of food with a high fat content which decreases decision utility but raises experienced utility.

The fact that decisions may be based on rules-of-thumb or subjective information can also be used to engineer incentives that change behaviour. Nudges – a small incentive that leads to a large change in behaviour – are frequently proposed as an incentivising technique. The reason that nudges might work is that they can affect the decision process in a significant way if this is not based on objective reasoning. For example, the likelihood of suffering a tax audit depends, objectively, on the full set of information revealed on a tax return and reported to the revenue service. However, experiments have shown that having to sign the tax return before completion leads to better compliance than signing after completion.[15] Bringing the signature to the start of the process is the nudge, and the effect might be explained by the fact that every subsequent false statement is violating an initial promise of honesty. A second example is the wording used in letters sent by the revenue service to taxpayers. Cabinet Office (2012) reports several trials of nudges involving small variations in the content and tone of letters and messages.

The relevant question for regulation of financial services firms is the extent to which these behavioural ideas apply to firms. This question has been analysed in detail in Armstrong and Huck (2010) who conclude, in the context of antitrust policy: “Behavioral economics may sometimes pose at least as many questions as it answers, and complicate antitrust debate. But it also sheds light on important market phenomena, and if competition policy is to reflect market realities, behavioural economics cannot be ignored.” Although the context is different, the basic observation of Armstrong and Huck that a behavioural perspective applies carries over to the regulation of financial services firms.

[15] The experiment is reported in Shu et al. (2012).

3 HMRC Experience

HMRC summarizes its role in this way:[16]

"Our purpose:

- we make sure that the money is available to fund the U.K.'s public services
- we also help families and individuals with targeted financial support

Our vision:

- we will close the tax gap, our customers will feel that the tax system is simple for them and even-handed, and we will be seen as a highly professional and efficient organisation"

The strategy it adopts to pursue the purpose and vision involves a diverse set of elements. The strategy has evolved over time and is supported by analytical teams that research improved techniques. HMRC is also part of the Forum on Tax Administration, an international collaboration of tax administrations that share their methods through working parties and conferences.[17] Further assistance with research into administrative methods is provided by international organizations, most notably the International Monetary Fund and the OECD. This reflects the benefits of international collaboration to improve regulation in an increasingly globalised world.

This section of the paper describes four elements of the HMRC strategy that are potentially informative for compliance in the financial sector. It begins with a brief review of the status and organisation of HMRC. This is followed by a discussion of the tax gap, which provides a summary measure of HMRC compliance success. Developments in large business, risk assessment, and litigation strategy are then reviewed. The section is completed by assessing the current situation.

Organisation of HMRC

HMRC is a non-ministerial department of the U.K. Government. It was formed by merging the Inland Revenue and Her Majesty's Customs and Excise and established by the Commissioners for Revenue and Customs Act (CRCA) 2005. The current structure involves a chief executive who is also permanent secretary and an executive committee that includes the second permanent secretary. The total staffing of HMRC is currently about 56,000. This number has fallen steadily since the merger of Customs and Excise and the Inland Revenue.[18] Much of the reduction has been achieved through closing local offices.

HMRC leads on maintaining and implementing tax policy. HMRC is not involved directly in formulating tax policy, which is the responsibility of the Treasury, but does conduct research to support policymaking in HM Treasury. A notable recent example of this is the HMRC (2012) report on the revenue effectiveness of the 50 percent tax rate. The report concluded that the 50 percent rate had significant disincentive effects that caused a reduction in revenue, and this finding played a major role in the final policy decision. HMRC and the Treasury work in close collaboration and the Financial Secretary to the Treasury is the minister responsible for HMRC.

[16] Downloaded from www.gov.uk/government/organisations/hm-revenue-customs/about#responsibilities on 9/10/15.

[17] Further details can be found at www.oecd.org/tax/forum-on-tax-administration/

[18] The staffing levels in 2004 were 23,000 at Customs and Excise and 68,000 at Inland Revenue (en.wikipedia.org/wiki/HM_Revenue_and_Customs)

HMRC collected £515 billion for the Treasury in 2014/15 with a total administrative cost of £1bn. The largest single tax was income tax with £171bn, followed by VAT with £105bn. Corporation tax and capital gains taxes raised £40bn and £4bn respectively.

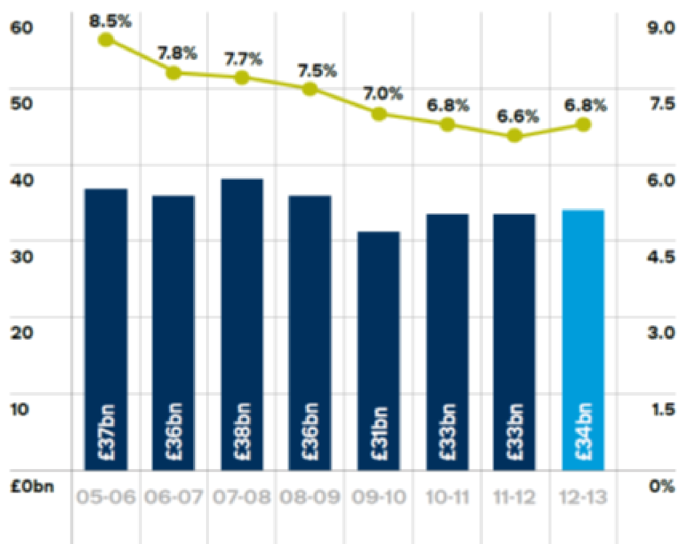
The Tax Gap

The central measure of the success of HMRC in improving compliance in tax payment is the tax gap. This is defined by HMRC as:

> “The difference between the receipts HMRC actually collects and the amount of tax that should be collected if all taxpayers complied with the letter and spirit of the law.”

HMRC has been publishing overall estimates of the tax gap since 2009. A small number of other countries also publish a similar measure (Denmark, Mexico and the U.S.) and more countries, including Australia, intend to publish one soon. The tax gap calculation has operational value for HMRC and its importance is made clear by the fact that it is listed first among the six strategic objectives of HMRC.[19]

First, the tax gap is an indicator of the long-term performance of HMRC in the compliance sphere and it displays the trend in non-compliance. The measure has been refined over time, and after each improvement, the tax gap measure is recalculated for earlier years. This ensures a consistent time series for the measure computed using a single technique. As shown in figure 1 the U.K. tax gap steadily declined from 2005-06 until 2011-12.[20] It increased slightly in 2012-13 to a value of £34bn, which was equivalent to actual tax collection being 6.8 percent less than what should have been collected. Second, the measure indicates where compliance resources can be usefully applied. Resources devoted to increasing compliance should be allocated across categories so the “marginal pound” of compliance expenditure obtains the same return in all categories. It is important to observe that the marginal impact of expenditure on compliance is needed for effective allocation of resources. Breakdowns of the tax gap by sector and by tax instrument are shown in figure 2. It is surprising to observe that VAT was responsible for £12.4bn (or 36 percent) of the tax gap since its design (collection of a fraction of the tax payment at each stage of the production process) is often thought to mitigate against non-compliance. Finally, HMRC frequently adopts initiatives aimed at reducing parts of the tax gap. The change in the relevant component of the tax gap can then be used to demonstrate the success or failure of these initiatives in tackling the tax gap. For example, fact that 44 percent of the tax gap is due to the activities of small and medium enterprises (SMEs) motivates the focusing of HMRC compliance efforts upon this sector.



[19] As listed on www.gov

[20] Figures 1-3 are taker

Figure 1: U.K. Tax Gap

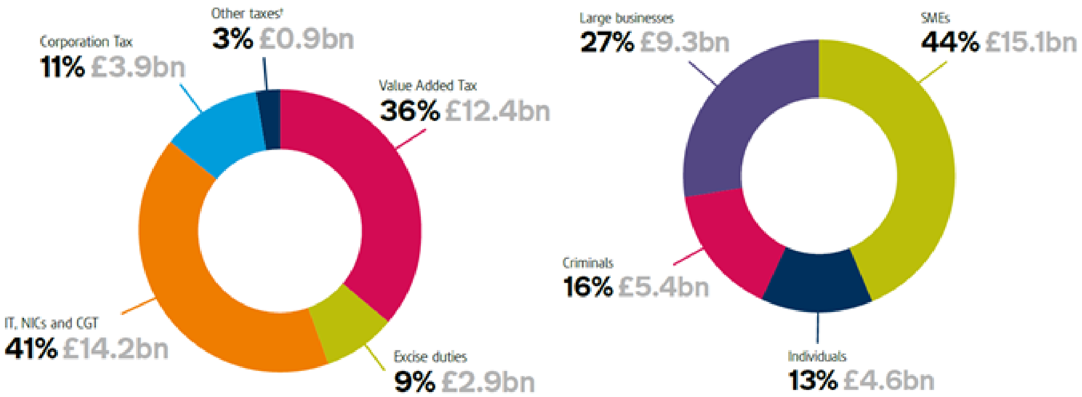


Figure 2: Tax gap by tax and sector

The measurement of the tax gap raises several methodological issues because measures what is not observed or directly recorded. The tax gap must be measured indirectly from the traces that the unrecorded activity leaves in other economic indicators and, due to the diverse nature of the tax system in the U.K., no single methodology can be applied uniformly to all taxes. Instead, HMRC selects the most appropriate methodology for each part of the tax gap and seeks, where possible, to use two entirely different measures of the same part of the tax gap to enhance validity. For example, there is a "top down" measure of the VAT tax gap that takes the Office of National Statistics measure of aggregate output for the economy and computes the VAT liability. The difference between the liability and actual receipts is the VAT tax gap. Second, there is a bottom up measure, formed from the summation of HMRC internal estimates of the tax gap for different VAT product categories. The two measures are then verified against each other to confirm the final measure. The tax gap for income tax combines results from the random enquiries programme and modelling methods that capture the income tax gap from groups such as moonlighters and "ghosts" (individuals with no official tax record). "Management information" – estimates based on the experience of HMRC staff – supplements these separate measures.

HMRC is generally regarded as having a rigorous and developed tax gap measurement processes. A recent review of the methodology by the International Monetary Fund was supportive of HMRC's work on the tax gap:

> "The HMRC's tax gap analysis program is comprehensive in tax coverage, effectively addresses its multiple dimensions, and work is ongoing to enhance its support to HMRC management. Tax gaps are estimated for most of the taxes administered by HMRC. In this regard, HMRC produces one of the most comprehensive studies of tax gap estimates internationally."
>
> (International Monetary Fund, 2013, p. 8)

The tax gap reflects the fact that if a behaviour is to be successfully regulated then it must first be measured. No incentive scheme can expect to succeed if it operates without an evidence base. The ability to slice the aggregate gap in different ways provides valuable perspectives on issues relevant to operations. The measure also provides an externally-verifiable indicator of performance. As figure 3 shows the methodology is not yet finalised. Although two thirds of the tax gap is measured using established techniques, the remaining third is based on developing or experimental techniques.

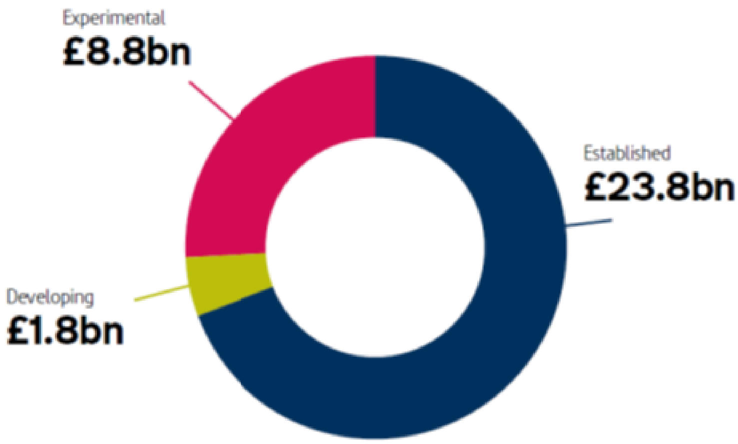


Figure 3: State of development

The success of the tax gap for HMRC suggests that a similar measure should be constructed for financial service regulation. This measure would be designed to capture the degree of non-compliance with regulations. Exactly parallel with tax non-compliance, this is not a directly observed quantity and would have to be inferred from the indirect observations. The FCA holds evidence on the results of company visits, the number of complaints about financial products, the number of cases of identified mis-selling and poor conduct, and the compensation payments that follow successful complaints. These items of information can be used to build a measure of the "compliance gap" that would inform the development of regulatory activities.

Large Business Service

The Large Business Service of the HMRC was formed following the Hartnett Review (2001). The role of the service is to manage the HMRC relationship with the largest companies and partnerships and to develop relationships so the two sides – HMRC and the business – share understanding and trust. This is intended to reduce a potentially confrontational attitude to tax payment and tax collection. The Large Business Service currently deals with the tax affairs of the 2,100 largest businesses in the U.K.

The method of operation is that each large business is appointed an HMRC Customer Relationship Manager (CRM) who maintains responsibility for that business. The role of the relationship manager with respect to the business is to: "engage with them directly to develop open and transparent working relationships."[21] Ideally, a two-way flow of information between HMRC and the business will exist so any tax issues can be resolved as they arise and not after a tax declaration has been submitted. This real-time working is designed to reduce the likelihood of tax disputes arising, which should benefit both HMRC and the business.

The Senior Accounting Officer (SAO) rules contained in Finance Act 2009 also had significant implications for the companies falling under the Large Business Service. A qualifying company had to identify its SAO, who must be a director or officer of the company and the SAO must take

[21] www.gov.uk/large-businesses

"reasonable steps to ensure the company has appropriate tax accounting arrangements."[22] The position must be certified annually to HMRC with supporting evidence provided of steps taken. A key element of the rules is that the SAO risks a personal penalty if they fail to comply. Penalties can be imposed if a qualifying company fails to inform HMRC of the name of the SAO, if the SAO fails to deliver on the main duties required of the position, or if the SAO fails to provide the annual tax certificate. HMRC has stated that the requirement that the SAO take reasonable steps to ensure that the company establishes and maintains appropriate tax accounting arrangements will in general

> "involve the SAO having in place mechanisms for identifying, on an ongoing basis, the risks which might result in the tax returns not being accurate in all material respects and ensuring that processes and controls are in place for managing and monitoring these risks."[23]

HMRC expect the SAO to understand the likelihood and potential impact of a risk being realized, and that the SAO should use this information to determine the extent to which a risk should be mitigated. The SAO is expected to discuss the appropriate level of mitigation with the CRM, although the ultimate responsibility for risks remains with the SAO.

The SAO regulations are generally viewed as increasing taxpayer transparency to HMRC. They are also significant for the increased importance they place upon taxation, and tax-risk management, within companies. Misutka and MaxEachern (2013) emphasise that the SAO regulations should raise the place of tax on the corporate agenda and may affect the corporate culture around tax compliance. If this is achieved, it will be a considerable benefit to HMRC.

The introduction of the Large Business Service and of SAOs was supported by a change in the HMRC risk model applied to corporations. The risk model changed in focus from considering all compliance issues to looking instead only at issues of higher value and risk to tax collection. The risk rating of a company determines the level of HMRC intervention. HMRC (2015) describes the features of a business that contribute to receiving a low risk rating.[24] Three groups of factors are involved: corporate governance, delivery of tax information and tax strategy. The High Risk Corporates Programme (HRCP) has existed since 2006. It works with the highest risk cases, which frequently involve multiple avoidance schemes (House of Commons Treasury Committee, 2012). The programme selects cases by considering the scale of open risks across all taxes and the company's tax and relationship strategies with HMRC.

The aims of HRCP are to[25]:

- resolve the tax issues of very large businesses by agreement or litigation
- reduce the scale of risk for the future
- improve the relationship between HM Revenue and Customs (HMRC) and its large business customers

A key element of the programme is Board-to-Board engagement, so contact under the programme is conducted at a high level on both the company and HMRC sides. The programme also involves a very high level of governance under the monthly cross-directorate HRCP Board. Any case is subject to rigorous investigation and involves face-to-face debate of technical issues with company representatives, leading in successful cases to an agreed and detailed project plan. The programme is also intended to involve effective partnership working across the directorates of HMRC and with external counsel. The desired outcome of the programme is resolution decisions fully in line with the Litigation and Settlement Strategy described further below.

[22] www.legislation.gov.uk/ukpga/2009/10/schedule/46
[23] www.hmrc.gov.uk/large-businesses/sao-script.pdf
[24] www.hmrc.gov.uk/manuals/tcrmanual/tcrm3370.htm
[25] www.gov.uk/large-business-the-high-risk-corporates-programme

The internal processes of HMRC for the review of large tax cases vary based on three factors: the value of tax involved, the complexity of the tax issues surrounding the case, and whether the case is under the HRCP. The largest and most sensitive cases are referred to the HRCP Board for review. The very largest cases must be signed off by HMRC Commissioners. Up to 2012, the practice was that two commissioners needed to sign off a settlement. In 2012, HMRC introduced a new assurance role at HMRC Commissioner level that involved the explicit purpose of challenging the decision-making on cases. This assurance commissioner has no engagement with taxpayers, nor any line management responsibility for case-workers. Consequently, decisions on the largest cases are now taken by three commissioners. These are the assurance commissioner, the director-general for the line of business involved, and a third commissioner from a different work area.

HMRC have also implemented a system of non-statutory business clearances through which tax questions can be directly addressed.[26] A non-statutory clearance is written confirmation of the view that HMRC takes of the application of tax law to a specific transaction or event. Since April 2008, such clearances have been provided for all taxes affecting corporations. A non-statutory clearance can be provided both before a transaction takes place and after a transaction. HMRC promises to provide a response within twenty-eight days or, for complex questions, provide information on how long a reply will take. The benefit to the corporation is that if they act based on the reply they can expect this will be the basis used for the final tax calculation. However, this is not guaranteed and some circumstances, such as legal rulings between the advice and the tax calculation, can result in the final tax demand differing from the corporation's expectation.

The role of the Large Business Service can be understood in the context of incentive theory as an attempt to reduce the asymmetry of information between the principal and the agent. If this were achieved, the scope for moral hazard arising is reduced, and the agent can be incentivised to perform correct actions at lower cost. The practical difficulties arise because the relationship becomes personal and individual personalities will matter. There is always the risk of a CRM failing to maintain the correct perspective by aligning too closely with the corporation's interests. It is important to find practical ways to mitigate this risk. Experience has demonstrated that there is such a narrow line between a sweetheart deal and an acceptable compromise that outsiders (especially the media and politicians) are unable to distinguish between the two.

The Code of Practice for Taxation of Banks was introduced in 2009 and applies to banks and building societies. It was designed to change banks' attitudes to tax avoidance and to encourage beneficial behaviour. The code encourages good governance within banks and a transparent relationship with HMRC. Banks were encouraged to sign up to the code, and the signatories were made public to provide social pressure. By the time of the 2014 Budget, 283 banks had signed the code and its 2015 annual report judges the code to be a success because of the reduced number of avoidance schemes disclosed.[27] Further, the annual report notes a perception within HMRC of a changing culture in banks. Financial regulators may well draw inferences for regulatory policy from HMRC's experience that an industry-specific code was warranted.

Recent Anti-Avoidance Measures

HMRC has been confronted by large-scale avoidance of taxation with an industry (including many banks) devoted to constructing and promoting tax avoidance schemes. This failure in compliance mirrors the extensive mis-selling of products in the financial services sector: both are not clearly illegal contravention of the regulations but an attempt to bypass or exploit the regulations. The important factor is that permissibility under the regulations must be tested on a costly and time consuming case-by-case basis. In past few years, two new policies have been introduced to deter

[26] The current guidance is given here: www.hmrc.gov.uk/Manuals/nbcgmanual/index.htm

[27] www.gov.uk/government/uploads/system/uploads/attachment_data/file/488869/The_Code_of_Practice_on_Taxation_for_Banks_-_Annual_Report_2015.pdf

what HMRC regards as “abusive” tax avoidance measures. Generally, abusive tax avoidance consists of complex artificial arrangements whose purpose is solely to reduce tax liability.

The first to be introduced, in 2004, were disclosure rules, known as Disclosure of Avoidance Schemes (DOTAS). These rules require promoters of avoidance schemes to disclose these schemes, so HMRC can decide whether they can be legally used and whether legislation should be used to prohibit them.

Schemes registered with HMRC are allocated a scheme reference number. Users of a scheme must report the scheme reference number in their tax return. Meanwhile, promoters of the scheme must report which of their clients have purchased each of their schemes. This dual reporting requirement makes compliance with the disclosure rules more likely, as nondisclosure by one party (e.g., the purchaser) could be discovered from the disclosure made by the other party (e.g., the promoter).

More recently, in 2013, a General Anti-Abuse Rule (GAAR) was introduced, following the example of several other countries.[28] This rule overturns the previous legal presumption that a taxpayer may use any lawful method to reduce their tax liability. It outlaws avoidance schemes that “cannot reasonably be regarded as a reasonable course of action.” This has become known as the double reasonableness test. In more basic language, an action is judged as abusive tax avoidance if it has no commercial justification other than the reduction of tax liability.

The policy objective of the GAAR is to provide a deterrence to taxpayers from entering tax arrangements that might be found later to be “abusive” and, equally, to provide a deterrence to the introduction and marketing of such arrangements. The key feature of a tax avoidance arrangement, which locates it under GAAR, is the fact that it is judged abusive. The GAAR legislation is designed to recognise that a taxpayer is often faced with several courses of action that have different tax implications. If an action is reasonable (as judged by the test) then it will not fall under GAAR. The difficulty facing a taxpayer is to determine when an action is reasonable, and when it is not. An analogy for consideration in financial services regulatory policy is development of products whose economic purpose is the same as that of a banned, penalised, or restricted product but whose legal form is that of a product not affected by tough regulation.

Risk Assessment

Identifying and estimating risks are a significant activity of HMRC. The risk assessment activities are designed to provide a much more detailed picture that allows identification of the largest risks among taxpayers. The description of incentives in tax compliance made clear the role that audits play in increasing the incentive for compliance and recovering tax for HMRC that would otherwise have been lost. The discussion of audit strategy was developed within a simplified setting in which the only indicator was the income declaration on the tax return. In practice, HMRC has considerably more information than this. For example, for a small company, it will know the history of taxpaying performance, the outcome of previous audits, the industry in which the firm operates, the data on previous returns, and the typical structure of a tax return from such a firm. The role of risk assessment is to combine these indicators, and most likely many more, into a quantitative assessment of the risk.

Neither the actual methods nor the data used by the risk assessor in HMRC are public information. What is known is that risk assessments are broadly based on the methods of credit scoring that are used in the financial sector.[29] Such credit scores are routinely applied in the finance industry and form the basis of decisions to award loans but are not faultless; the

[28]The most recent guidelines for GAAR are given here: www.gov.uk/government/uploads/system/uploads/attachment_data/file/399270/2__HMRC_GAAR_Guidance_Parts_A-C_with_effect_from_30_January_2015_AD_V6.pdf

[29] www.statsoft.com/Textbook/Credit-Scoring, and the printed version, Hill and Lewicki (2007).

significant number of defaults during the financial crisis made many question the application of the methodology.[30] Although any statistical predictions will be imperfect what really matters is the size of the type 1 and type 2 errors.

The process adopted is described in HMRC (2015) as follows:

> "As Tax and Audit Specialists identify potential risks as a result of their Risk Assessment activity, they should enter those they believe to be worthy of team discussion onto the Customer Relationship Management Module which will calculate a Priority Risk Score (PRS). The PRS is used to determine whether a risk is significant enough to be raised with the customer based on the value, probability and impact of the potential risk. Further information on the Customer Relationship Management Module and Priority Risk Scores can be found in the Customer Relationship Management Module guidance."[31]

Academic research on predictive methods use in tax compliance has tested the risk assessment methods using computer-based simulations. Hashimzade and Myles (2015) show that using risk assessments can raise revenues by approximately 8 percent compared when compared to random audits. But the key finding is that the audit strategy that emerges looks like the cutoff rule identified in the analysis of incentivising compliance with audits. That is, audits are focused on taxpayers reporting the lowest level of income and above a certain level, or cutoff, virtually no taxpayers are audited.

The application of the credit scoring methodology is in its early days at HMRC. At present, the methods seem to be broadly accepted but not completely believed. The advent of "big data," and methods for its analysis, has opened up the possibility of considerable refinement in the production of risk assessment. Particularly beneficial could be the linking of data across government departments to build a complete data picture. One area that risk assessment cannot help with is the "ghosts" who do not register at all for tax purposes. The lack of tax data prevents a statistical risk assessment so they must be addressed using other methods, as do those who carry on regulated financial activities but who do not seek authorisation from the FCA.

Litigation and Settlement Strategy

The Litigation and Settlement Strategy (LSS) is a process designed to ensure better outcomes for HMRC. The LSS was first published in 2007 and was refreshed in 2011 and 2013.[32] At the heart of the LSS is the recognition that HMRC has finite resources and that these are better focused on cases for which there will be a revenue gain for HMRC. In economic terms, it provides explicit recognition that HRMC staff should consider the marginal cost of obtaining additional revenues and focus effort where the marginal cost of an additional unit of revenue is lowest. The FCA may be able to use metrics for redress or detriment in a similar way. The LSS applies to all taxes and taxpayer types within HMRC.

The LSS encourages HMRC staff to work to minimise the scope for disputes between HMRC and corporate taxpayers and to always seek non-confrontational solutions to tax disputes. Further, it confirms that the selection of cases to investigate and the handling of those cases should be based on what best closes the tax gap. That is, the additional revenue from each case and the ease of obtaining those revenues should be what guides case selection. In implementing this guidance, the LSS does place limitations on flexibility since it requires the resolution of tax disputes in a manner consistent with HMRC's considered view of the law. It does not suggest that the law should be set aside to ease the generation of revenue. The LSS also requires HMRC staff to ensure that the potential revenue flow from a case is sufficient to make any dispute worthwhile.

[30] www.bloomberg.com/bw/stories/2008-02-06/credit-scores-not-so-magic-numbers

[31] www.hmrc.gov.uk/manuals/tcrmanual/TCRM4300.htm

[32] www.gov.uk/government/publications/litigation-and-settlement-strategy-lss

The summary guidance in HMRC (2013) for its enforcement strategy distinguishes between strong cases (settle for the amount a court would determine, otherwise litigate), “all or nothing” cases (do not split the difference), and weak or non-worthwhile cases (concede).

The success of HRMC in resolving disputes under the LSS is the subject of an annual report by the Tax Assurance Commissioner,[33] which gives an overview of the distribution of cases by taxpayer type. HMRC (2015) reports on a review of disputes that were settled in 2013-14. In total, 402 cases were reviewed, and in 80 percent of these cases there was evidence of full adherence to governance procedures. A further 18 percent had evidence that governance processes were followed but there was some scope for improvement in compliance with processes. Finally, 2 percent had “one or two aspects” not meeting governance standards.

The LSS represents a clear process for both parties to a tax dispute. More importantly for this review, it also contains the correct economic approach to disputes. Revenues should be pursued where the marginal benefits exceed the marginal costs and these factors should be explicit in the decision to progress with a dispute. What the LSS does not do is consider the indirect effects that follow any enforcement activities since it is focused entirely on costs and direct revenues. Some enforcement activities may have wider incentive effects than others but these are not considered. The FCA has implemented a strategy with equally clear aims and processes. FCA (2014) describes the broad structure of this strategy and the organisational restructuring that supports the strategy.

Review of Process

The primary objectives of HMRC are to raise revenue where it is due and reduce the tax gap. These objectives are pursued subject to the legal requirement that all taxpayers are treated fairly and that taxpayers are to be believed except where there is evidence that they have provided false information. The pursuit of these objectives is also subject to the limited resources within HMRC.

Two competing factors are at work here. First, the staffing of HMRC has fallen significantly in recent years as reported above, and the number of local offices has been reduced, which can limit the exploitation of local expertise. Second, HMRC has received generous settlements in recent spending plans.[34] However, a case can be made that, on balance, the combination of objectives and resources has led HMRC to make decisions that prioritise short-term revenue raising over the punishment of offenders. This, indeed, is precisely how LSS can be interpreted.

The claim that short-term revenue raising is given priority is supported by the following two observations:

- HMRC compromises in settling tax disputes if the costs of litigation would be high, especially if the outcome is uncertain.
- HMRC sometimes offers “incentivised tax disclosure schemes” under which taxpayers who have deliberately evaded can pay the tax due with a relatively low fine imposed, if they declare their true tax position before it is discovered by HMRC.

The risk to HMRC of pursuing approaches that prioritise revenue over punishment is that they can reduce “tax morale” – the general public’s sense of fairness of tax administration and policy – by giving the impression that tax evaders are not punished severely enough. The academic literature has found that tax morale has a significant impact on compliance behaviour[35] and that a fall in morale will reduce compliance and future tax revenues. HMRC must achieve a balance

[33] Currently Sir Edward Troup, second Permanent Secretary.

[34] www.theguardian.com/politics/2015/jul/08/budget-boost-for-hmrc-in-new-push-on-tax-evasion.

[35] Evidence on the effect of tax morale can be found in Torgler (2001) and Torgler and Schaltegger (2006).

between doing the correct thing by collecting revenue and being seen to do the correct thing by punishing non-compliance.

The interpretation of this concept in financial services relates to the attitude of financial services firms to regulation. “Regulation morale” would be high if the firms generally accept that the regulatory framework is fair and justifiable. If the parallel with taxation applies, high regulation morale would result in voluntary compliance with regulations. In contrast, if regulations are considered unfair and unjustifiable, then the response should be a culture of non-compliance and disregard for regulations. These ideas could potentially be tested using a similar approach to the tests of the impact of tax morale.

The FCA is pursuing an initiative to develop “sustainable regulation,” designed to build support from financial services firms.[36] The potential benefits of this approach can be understood by adapting the tax morale argument. A regulation strategy that balances rules, supervision, and enforcement and is developed with support of financial services firms will encourage additional compliance through increased morale. Conversely, an adversarial approach lacking firms’ support will generate low morale and reluctance to comply.

A related challenge for the FCA results from its ambition to use insights from behavioural economics where it is sensible to do so. These insights can require very detailed rules, for example specifying disclosures and documentation in ways shown by trials to be more effective than alternatives. This, however, will tend to lengthen rule books and lengthy requirements, through sheer volume, can make compliance difficult in practice. This in turn may damage “regulation morale.”

Discussing tax morale emphasises how important public perceptions of tax administration are for sustaining a high level of compliance. Public perceptions are important because the tax system relies heavily on what is little more than voluntary compliance: the number of audits conducted is low and there is no third-party reporting for many taxes. For example, capital gains tax requires disclosure of transactions about which HMRC will have little or no information.[37]

If the public feel that the tax system’s administration is unfair, then this could possibly reduce their degree of compliance. Similar observations apply even more strongly to financial services. The number of firms visited is low and, moreover, little external evidence exists from which to judge compliance. HMRC can compare actual tax receipts to expected tax receipts as an initial indicator of compliance. There is no similar “first look” indicator for compliance with financial service regulations.[38] From this perspective, decisions by HMRC not to prosecute tax evaders – such as in the recent HSBC Geneva case[39] – are harmful to public perceptions. This is enhanced when the decisions receive extensive adverse publicity in the media or are made the subject of political debate. Similarly, the cases of multinational companies paying apparently very low rates of tax in the U.K. have received a very poor public reception to the extent that they have been the cause of public demonstrations.

A survey conducted by the Oxford Centre for Business Taxation reveals the business perspective on HMRC’s recent strategies.[40] Several important observations can be drawn from the survey. First, 68 percent of the companies surveyed held the opinion that the Business Risk Review resulted in a more focused approach from HMRC. This figure can be further divided into 72 percent of firms dealing with the Large Business Service (LBS) agreeing, but only 53 percent of

[36] The FCA model is described here: www.fca.org.uk/news/towards-sustainable-model-of-regulation

[37] A good example is the purchase and sale of a second property. The transactions may be public information via the Land Registry database but nothing automatically signals that the sale is of a second home (on which any capital gain is taxable) rather than a first home (on which the gain is not taxable). Even if the sale was observed, HMRC would still not have information on any of the allowable deductions.

[38] Careful analysis of complaints data might provide some insights but these are likely to be undercut by consumers’ lack of information or understanding, especially in complex situations. Possibly monitoring of informal discussions on a “Trip Adviser” for financial services would yield useful insights.

[39] Reported here: www.theguardian.com/business/2016/jan/13/hmrc-admits-to-winding-up-inquiry-into-hsbc-tax-evasion-claims.

[40] The report can be downloaded from www.sbs.ox.ac.uk/sites/default/files/Business_Taxation/Docs/Publications/Reports/HMRCs-relationship-with-business.pdf

large and complex (LC) firms. Second, 64 percent believed that it was impossible for large complex multinationals to ever be classifiable as low risk (LBS: 67 percent, LC: 53 percent). Third, most firms surveyed perceived a benefit in moving from high to medium risk but not from medium to low. The lack of benefit arose from the costs of implementing the processes necessary to be judged low risk.

Further responses from the survey are summarized in Table 1. In general, they support the work of HMRC. The only exception is the discretion that HRMC staff could exercise.

Statement	**Percent in agreement**
HMRC has not taken forward disputes which could not be justified in terms of revenue at stake	82
HMRC adopted collaborative working practices	77
HMRC staff were familiar with the law underlying a dispute	90
HMRC staff were familiar with the facts underlying a dispute	78
HMRC staff with whom respondent negotiated had sufficient discretion	52

Table 1: Assessment of HMRC

Public perceptions and tax morale are important for sustaining compliance in a system where much tax payment is little more than voluntary. The reaction of the public to some recent HMRC actions has been very strong and very negative. This was amplified by the nature of the political debate. Such perceptions can limit how HMRC can behave, or at least be seen to behave. In contrast, the business perception of recent HMRC strategies is much more favourable and indicates that these strategies have moved the relationship between HMRC and business in the correct direction.

4 Cooperative Compliance

The OECD Forum on Tax Administration concluded its report into aggressive tax planning by encouraging large corporate taxpayers and revenue services to develop relationships based on cooperation and trust (OECD, 2008). This was termed “the enhanced relationship” and two pillars were suggested as the basis of this relationship:

- In dealings with taxpayers, revenue services demonstrate understanding based on commercial awareness, impartiality, proportionality, openness through disclosure and transparency, and responsiveness.
- In dealings with revenue services, large corporate taxpayers provide disclosure and transparency.

These ideas have been adopted by many countries alongside the adoption of tax control frameworks as a tool for disclosure and transparency. Table 2 provides a summary list of the countries with formal adoption or pilot projects. In OECD (2013) the proposal was made to rename the concept as “cooperative compliance” to capture the idea of a close working relationship between the corporation and the revenue service.

Formal Adoption	Pilot Projects
Australia	Austria
Canada	Finland
Denmark	France
Ireland	Norway
Japan	Russia
The Netherlands	
New Zealand	
Singapore	
Slovenia	
South Africa	
Spain	
Sweden	
UK	
USA	

Table 2: Adoption of cooperative compliance, based on Table 2.1 OECD (2013)

Cooperative compliance essentially involves a less confrontational approach to the relationship between the revenue service and large corporate taxpayers. The system should be designed to ensure that the level of confrontation is reduced by both parties – the corporation and the revenue service – involved in tax payment. To achieve this, there must transparency in the tax behaviour of the corporation and the tax demands of the revenue service. This obliges the revenue service to develop a better understanding of the corporation's business activities and risk profile. The same features are integral to the FCA sustainable model of regulation: clear and shared understanding, willingness to review rules, and constant and open engagement.

A cooperative compliance arrangement should also be viewed as a long-term commitment by both parties. This permits trust between parties to be built which should further improve transparency and increase the extent of information sharing. The increase in trust is facilitated by implementing horizontal monitoring as part of cooperative compliance. Horizontal monitoring is the idea that contact between the corporation and the revenue service is conducted at a high level – preferably at board level – between the two organizations. This is the idea behind the SAO role implemented in the U.K. by HMRC.

Further benefits that should follow from a successful cooperative compliance agreement are increased speed of decision-making and reduced potential work backlogs. Ideally, the decision-making should be undertaken in real time and the interventions of the revenue service should move from repressive to preventive. These improvements follow as consequences of the improved information sharing and the revenue service's enhanced knowledge of the corporation's activities. In an ideal implementation, tax issues should be solved in advance of the submission of a return. How far the actual systems achieve these ideals has not yet been fully assessed.[41]

Implementing a cooperative compliance arrangement requires that the corporation and the revenue service agree formally to the arrangement. As part of the arrangement, a statement of basic principles describes how the two parties should behave. The arrangement will also place commitments on both parties and include expectations of how the parties should behave. The duration of the arrangement will also be defined, and the frequency for re-evaluation stated. The HMRC Code of Practice on Taxation for Banks allows for naming banks that refuse to sign up to the agreement. A significant number of banks declined to sign the code when implemented, but by March 2015 virtually all banks had done so.[42] Importantly, HMRC felt the need to adopt special treatment for the banking sector. This emphasises that the financial sector poses unique compliance difficulties.

The example of implementation of cooperative compliance agreements by the Netherlands Tax and Customs Administration (NTCA) is instructive. The vision of the NTCA in implementing cooperative compliance is to maintain and improve voluntary compliance through influencing behaviour. As such, the agreements are considered an incentive device to secure an improved outcome. The structure of agreement in the Netherlands is not fixed. Instead, the NTCA adopts responsive regulation. In practice, this means that the form and intensity of the supervision is determined by (a) the willingness of the corporation to comply and (b) the ability of the corporation to comply.

The starting-point for developing the supervision process is a client profile. From this the choice of enforcement instruments is made to ensure that there is a balanced mix of repressive and preventive actions (corresponding to supervision and enforcement in FCA terminology). Among the preventive instruments are assessing rules and legislation, communication between the NTCA and the corporation, tax services provided by the NTCA, and visits by NTCA officials to corporations. The repressive tools and corrective actions include the assessment of tax returns, filed audits, fraud investigations and administrative fines. The outcome of this process is a

[41] OECD (2013) reports the responses to a survey on the success of cooperative compliance but does not provide any quantitative analysis.

[42] HMRC, The Code of Practice on Taxation for Banks – Annual Report 2015.

strategic supervision plan designed specifically for each corporation that enters into a cooperative compliance agreement.

The steps in the process are described in Table 3 with the assignment of responsibility. This described the actors with the NTCA and large business (LB) responsible for the implementation of each step.

	Step	Actors/Responsibilities
0	Up-to-date client profile	Account team NTCA
1	Introduction to horizontal monitoring	Management NTCA and management LB
2	Compliance scan	Account team NTCA in cooperation with the LB
3	Resolution of pending tax issues	Account team NTCA in cooperation with the LB
4	Compliance agreement	Management NTCA and management LB
5	Analysis and improvement of tax management and monitoring	LB
6	Form and intensity of supervision adjusted Strategic Supervision Plan	Account team NTCA

Table 3: Implementation of cooperative compliance at NTCA

Implementing a cooperative compliance agreement places obligations on both parties. The corporation is obliged to ensure that the tax strategy is transparent and that tax issues are embedded throughout the corporation. The Forum on Tax Administration (FTA, 2010) has described an ideal structure for how the corporation should treat tax matters in a tax control framework.

The six building blocks of a tax control framework as identified by the OECD/FTA can be summarized in the following points:

1. Tax strategy.

The tax strategy should be clearly documented and owned by senior management of the corporation at board level.

2. Comprehensive.

The tax control framework needs to be able to govern the full range of the corporation's activities and should be embedded in the day-to-day operations of the corporation.

3. Responsibility.

The role of the tax department should be recognised within the corporation and implementation of the tax compliance framework should be fully resourced.

4. Governance.

There should be a system of rules and reporting to ensure that tax activities compare with expected normal tax activities and to ensure that the potential risks of non-compliance are identified. This procedure should be explicitly documented.

5. Testing.

Compliance with the policies and process in the tax control framework should be the subject of regular monitoring, testing and maintenance.

6. Assurance.

The tax control framework should be capable of providing assurance to internal and external stakeholders that tax risks are subject to appropriate control and that outputs of the process, such as tax returns, can be relied upon.

The SAO concept described in section 3.3 above incorporates most of these recommendations.

The discussion above has focused on the potential benefits of an ideal cooperative compliance system. The OECD has initiated the development of cooperative compliance systems and has argued very strongly in support of them. However, a potential drawback needs to be raised. The process of cooperative compliance, and the system of customer relationship managers implemented by HMRC, by nature requires developing a close working relationship between officials at the revenue service and at the corporation. This puts the process at risk of what economists call "regulatory capture": the regulator coming to identify with the corporation more than with the revenue service.

In practice, regulatory capture would see the revenue service official becoming too keen on reaching an agreed outcome with the corporation. There is then the possibility that this can lead to lenient taxation, with the interests of the revenue service being sacrificed relative to those of the corporation.[43] This returns the argument back to behavioural economics and applying psychological ideas to explain the motivations of economic agents. There is anecdotal evidence that relationships between the revenue service and corporations have become too close but no scientific testing has been attempted.

[43] An example that can possibly be explained in these terms is the agreement reached by David Hartnett with Goldman Sachs. See www.theguardian.com/commentisfree/2013/may/28/dave-hartnett-one-sweetheart-deal-too-many

5 Implications for Financial Services Regulation

Currently, financial institutions in the U.K. appear to have a poor record of compliance with the regulations, at least compared with taxpayer compliance. This raises the question whether financial regulators can learn anything from tax officials.

In considering this question, it is important to recognise some differences between the requirements of tax enforcement and financial service regulation. Key differences include:

1. The tax authorities are only looking at one aspect of behaviour (paying the correct tax amount); the financial services regulator is policing a much wider range of behaviours, including such disparate behaviours as market manipulation and mis-selling of financial products to individuals.
2. Many aspects of financial misbehaviour apply to individual transactions rather than any aggregate measure of behaviour analogous to total tax payment. This requires the regulator to look at the activities of financial services firms in much greater detail to discover evidence of malpractice.

These differences imply that considerable care is needed in applying tax administrators' experience to financial regulation. Nonetheless, it is worthwhile considering the extent to which HMRC practices could be adapted to meet some of the FCA's needs, and we have flagged some ideas in preceding sections.

Regulating Financial Services

The FCA has a range of supervision and enforcement tools available to achieve its regulatory objectives. It can specify minimum standards for financial services and place requirements on products. To ensure compliance with regulations it has the power to investigate both organisations and individuals. The FCA possesses the power to ban financial products, to instruct firms to retract or modify products, and to publish such decisions. The FCA Annual Report provides a record of its pro-compliance activities undertaken in each year.[44] The compliance reports are notable for openly naming individuals and organisations against which compliance action has the taken.

The FCA Handbook[45] provides detailed guidelines that firms selling financial services must follow. The topics include High Level Standards, Prudential Standards and Business Standards. Also included are details on Regulatory Processes and Redress. At the heart of the Handbook are sets of Principles that apply to each area. As stated in the Handbook, "The Principles are a general statement of the fundamental obligations of firms under the regulatory system" (PRIN 1.1.2). The FCA sees the principles as setting the minimum standard that each firm must achieve. In this respect, they are not a set of rules that must be followed with specific test and penalties. This is explained in the following: "Breaching a Principle makes a firm liable to disciplinary sanctions. In determining whether a Principle has been breached it is necessary to look to the standard of conduct required by the Principle in question. Under each of the Principles the onus will be on the appropriate regulator to show that a firm has been at fault in some way" (PRIN 1.1.7).

[44] www.fca.org.uk/your-fca/documents/corporate/annual-report-14-15
[45] www.handbook.fca.org.uk/handbook

The highest principle on the treatment of customers by firms requires that “A firm must take reasonable care to ensure the suitability of its advice and discretionary decisions for any customer who is entitled to rely upon its judgment” (PRIN 2.2.1). If this Principle is violated, it does not follow that the customer can take direct action to secure compensation and redress. This is stated in PRIN 3.3.4: “A contravention of the rules in PRIN does not give rise to a right of action by a private person.” Consequently, the Principles give protection to the customer but cannot be directly enforced, which undermines a potentially important market discipline.

A key regulatory role for the FCA is to act on the mis-selling of financial services to customers. There is no direct definition of what constitutes mis-selling in the FCA Handbook but other FCA documents do provide guidance. In “Risks to consumers from financial incentives”[46] the following information is provided:

> “What do we consider ‘mis-selling’?
>
> We use the term ‘mis-selling’ in this document to refer to a failure to deliver fair outcomes for consumers. These outcomes include:
>
> - customers are treated fairly
> - customers understand the key features of the product and whether they are being given advice or information
> - customers are given information that is clear, fair and not misleading – information that enables them to make an informed decision before purchasing a product or service or before trading
>
> and
>
> - customers buying on an advised basis are recommended suitable products”

This description is informative but not entirely precise without further specification of terms such as “fair” or “key features.” However, it does provide an adequate basis for the following discussion.

The FCA’s reports make it clear that there have been numerous incidents of significant mis-selling in the financial services industry. The episodes of mis-selling have included, among a long list, advice to nurses, armed forces personnel, and miners to exit from index-linked pensions into private pensions, endowment mortgages, PPI and credit card protection. In 2013/14 alone, the FCA imposed 46 penalties totalling £425m.[47] It also secured criminal convictions against several individuals.

The frequency of major mis-selling episodes raises the question of why so much violation of the FCA Principles takes place in a closely regulated industry. A potential explanation is that the culture of the financial services industry is to blame. This explanation accords with the FCA striving to develop a model of sustainable regulation to drive culture change. If correct, the explanation is consistent with having appropriate rules in place but with the financial services firms pursuing objectives that make compliance with the rules of secondary importance. A parallel can be drawn again with tax compliance. The tax rules in the U.K. and in Italy are similar in rates and structure. Levels of enforcement are also similar. However, the culture with regard to tax compliance is different, which results in considerably lower compliance in Italy than in the U.K.

[46] www.fca.org.uk/static/fca/documents/finalised-guidance/fsa-fg13-01.pdf

[47] From FCA Annual Report 2013-14: “In 2013/14 we imposed 46 penalties (totalling £425m), five public censures and 26 prohibitions. We have also taken action against firms and individuals operating a variety of unlawful schemes, such as boiler room frauds, land banking scams, and other “get rich quick” investment schemes. This has included freezing assets, closing down unlawful schemes and pursuing civil and criminal action in appropriate cases. We have warned the public about the risks of other types of investment scheme such as carbon credit and graphene scams, which often fall outside of our jurisdiction but nevertheless pose a risk to consumers. We secured criminal convictions against Michael Lewis, Gary Hexley and John Cooper – who, among other offences, breached the general prohibition (i. gave financial advice without being authorised to do so) – and against Benjamin Wilson for defrauding investors of over £21m, after pleading guilty to fraud, forgery and operating a collective investment scheme without authorisation. We also secured a criminal conviction for market abuse.”

The FCA's activities demonstrate recognition that mis-selling can occur as a byproduct of the environment within a firm and not necessarily as a directly intended consequence. Firms are complex organisations with many layers of responsibility and different incentives operating for different employees. A decision made with good intentions at one level (such as the provision of a new insurance product) can have a negative impact when implemented at a different level. In such cases, the observed culture of the firm toward compliance reflects the outcome of the interplay of incentives within the firm rather than being the consequence of a single guiding authority. A culture that emerges through this route must be changed by ensuring that the internal incentives align with good conduct.

The importance of internal firm incentives is clearly recognised by the FCA. The report "Risks to customers from financial incentives – an update"[48] develops from the observation that the structure of financial incentives within a firm influences the decisions of employees. If the financial incentive scheme is poorly designed, it can mistakenly incentivise mis-selling. For example, if a sales person on a short-term contract is offered a bonus for selling then they will missell – knowing that they will not be with the firm when the consequences are realised. The report confirms that bonuses paid for the sale of PPI are recognised as one of the major instances of poor incentive structures. Financial incentives are not the only cause of mis-selling. In a second report, the FCA consider the possibility of mis-selling arising due to the general process of performance management.[49] Clearly, a workforce under pressure to achieve sales targets will be more likely to sell inappropriate products to customers.

The difficulties faced by the FCA in combating mis-selling can now be addressed. It is instructive to begin by comparing the determination of when a financial product has been missold to the judgements made by HMRC on whether the correct level of tax has been paid. HMRC receive a report of the activities of the individual or company at the end of each financial year. This information can be checked against the legislation at that point to judge whether the correct report is being made. HMRC use the supplied information to calculate the tax liability so there is no issue about the liability given the reported information. What is potentially in doubt is whether the information is correct.

Information can be incorrect for a variety of reasons. First, there can be deliberate misreporting to reduce tax liability through understatement of income or overstatement of expenses. Second, there can be errors by the taxpayer. Third, there can be misunderstanding of the tax rules. The sum of these three gives the omitted income of the taxpayer (which can be negative if errors or misunderstanding cause income to be overstated). The role of the audit process is to uncover this omitted income and ensure that the correct level of tax is paid, plus a penalty in some circumstances. In summary, a set of tax rules is in place for a given tax year. In principle, any set of activities has a clearly defined tax liability given these rules. The assessment is undertaken ex post using information on activities completed within the tax year. Granted this, the HMRC's task is to determine that the correct information has been declared on the tax return.

For the clear majority of personal income taxpayers, there is no doubt about what the correct level of tax payment should be. The circumstances of most taxpayers are sufficiently simple that the rules can be applied without question to the activities of the taxpayer and the tax liability determined. Regarding the theory and evidence discussed in section 2, the theory has normally been developed in the context of individuals deciding whether to evade personal income tax. In addition, almost all the compliance evidence from experiments and audits comes from personal income taxpayers.

For companies, the situation is less straightforward. Businesses, including financial institutions, can have various legal forms. This is, of course, an issue for the tax authorities as well as financial regulators. In terms of the theory of tax compliance, the nature of the risk is broadly like personal income tax. However, it can be argued that there are some differences. The finance

[48] www.fca.org.uk/your-fca/documents/thematic-reviews/tr14-04

[49] www.fca.org.uk/your-fca/documents/guidance-consultations/gc15-01

literature suggests that widely-held businesses with several (or very many) owners are likely to be less risk-averse than private individuals. Also, the penalties that arise when non-compliance is detected tend to be borne by the owners of the business, rather than any employee deciding on whether to break the rules. Arguably, this makes larger businesses even less risk-averse.

Numerous complications also arise from corporate structure and through international activities that make direct application of rules difficult. In these cases, the correct level of taxable income is a question of interpretation that may require a process of negotiation. The HMRC LBS is designed to deliver these negotiations in real time rather than though after-the-fact litigation. The implication of this is that larger businesses need stronger enforcement, not only because of the damage that might be caused by any risk-taking on their part, but also because they are more likely to take risks. As discussed above, HMRC does apply stricter supervision to larger businesses.

A further recent policy initiative to encourage a change in culture in financial regulation is the Senior Managers Regime in force since March 2016.[50] This mirrors the SAO scheme implemented by HMRC. It identifies senior individuals holding key roles within financial services firms upon whom identified obligations fall. Under the terms of this regime firms need to:

- ensure each Senior Manager has a statement of responsibilities setting out the areas for which they are personally accountable
- produce a firm responsibilities map that knits these together
- ensure that all Senior Managers are pre-approved by the regulators before carrying out their roles[51]

The “duty of responsibility” introduced by the U.K. Government in the Bank of England and Financial Services Act 2016 means that senior managers must take reasonable steps for a person in that position to prevent a regulatory breach from occurring. If such steps are not taken, individuals (rather than firms as was previously the case) can now be held to account and punished. Introducing individual accountability is another step in changing the culture of financial services firms.

Determining when mis-selling has taken place is a major judgement the FCA must make in conducting its enforcement and supervision. This task is complicated by the nature of financial services: many financial services are risky, so the outcome is unknown at the time at which the customer makes a purchase. Mis-selling cannot therefore be determined on an ex post basis by reviewing the outcome of the purchase. Riskiness implies there will always be a possibility for the occurrence of outcomes detrimental to the customer. A service is not necessarily missold because the ex post outcome is poor.

The correct determination of mis-selling must be made from an ex-ante perspective on the same basis that the original sale was made. The FCA description of mis-selling provides the key points that form the basis of the judgement. A service is correctly sold when the customer is given sufficient information to understand the product they are buying. The customer must also understand whether the seller is giving advice or providing information. Finally, any customers being advised must be sold suitable products. These conditions make it necessary to have an individual-based test to determine whether mis-selling has occurred because the judgement is relative to the characteristics of the customer. For example, PPI may be a perfectly suitable product for a customer in employment but is not a suitable product for a self-employed customer when self-employment disqualifies a claim.

Reflecting on the situation reveals that mis-selling is more difficult to identify than non-compliance with tax rules. In practice, mis-selling is hard to assess because of the ex-ante judgement involved. There is a need to know what took place at the point of sale regarding the information that flowed between the seller and the customer. This requires correct and verifiable record

[50] www.fca.org.uk/your-fca/documents/policy-statements/ps16-06

[51] www.fca.org.uk/firms/senior-managers-certification-regime#sthash.Bifyy9kf.dpuf

keeping by both parties. The product must be suitable for the customer but suitability may involve a subjective test, rather than an objective one. PPI provides an example of an objective test: selling an insurance product to a customer who is a member of a group that cannot claim under that policy is an objective example of an unsuitable product. In contrast, selling a risky product to a customer who does not wish to make a risky investment becomes a subjective judgement: was the product too risky given the customer's preference? It would require very detailed information to assess how the customer's risk preferences were assessed at the point of sale. Riskiness raises a further question: should the product be judged in isolation or viewed in its interaction with other investments in the customer's portfolio?[52] The assessment of suitability may be very different from the two perspectives. Any data on sales does not cover deals that did not happen, so failure to recommend a product could equally be mis-selling but will never be observed. Further, much mis-selling can occur before it is fully recognised by the regulator. The case of PPI is the prime example. The mis-selling was not initially detected because at the time of sale, oversight for PPI fell administratively between the FCA and the Office for Fair Trading.

The extent of these difficulties is made apparent by the dispute between the U.S. Department of Labor and Robert Litan.[53] The dispute centres on the possibility that brokers do not act in clients' best interest but instead are swayed by commission fees. The U.S. Department of Labor conducted a cost benefit analysis and concluded that commission-based sales should cease. Litan argued that the cost benefit analysis of the U.S. Department of Labor was flawed because it did not consider the costs that would be imposed on small investors by the new rules. If brokers could not earn commission, then they would need to charge fees. Faced with a fee for advice small investors might then choose not to pay for advice and consequently make poor investment decisions. The message here is that a full assessment must look beyond the immediate effects to see ultimate consequences.

The ideal situation for regulation of mis-selling would be real-time monitoring of the sales of financial services firms. This would permit mis-selling to be identified as it occurred and prevent the accumulation of many cases that must be compensated ex post. HMRC have implemented real-time monitoring for large businesses because the number of firms is low. In contrast, the number of financial transactions is very high so implementing real-time monitoring is more of a technological problem than a methodological problem. There is evidence of success in cases where it has been applied, such as stock exchange transaction monitoring.

Detecting Regulatory Avoidance in Financial Services

This sub-section considers the compliance measures used by HMRC, the encouraging extent to which analogous measures are applied by the FCA and whether there is scope for greater use in financial regulation. The main measures identified in the preceding two sections lead to the following suggestions:

1. Making punishments sufficiently severe that they cost the offending business more than the business has gained by violating the regulations. The system of tax penalties automatically ensures this. The FCA does have the power to set a penalty, and organise compensation, but it is not clear that these are always high enough to produce real deterrence, even though the FCA publicly names offenders and HMRC does not. There is a case for considering whether the penalties could be modified to achieve a greater deterrent effect.
2. Adopting rules like HMRC's SAO rules (see section 3.3), so a named person is responsible for submitting truthful reports, and who would face a personal penalty if any reports are misleading. The FCA does recently introduced a similar regime, the "Senior Managers

[52] A standard result in finance theory is that combining risky assets with negatively correlated returns can produce a portfolio with less risk than either of the assets (and in the limiting case can produce a risk-free portfolio).

[53] www.fortune.com/2015/10/05/elizabeth-warren-robert-litan-u-s-department-of-labor/

Regime," under which managers can be punished for certain failings. It would be useful for this new scheme to be assessed after some experience of how it operates in practice, with the possibility of extending its scope and reviewing the levels of punishment.

3. Systematically using information from third parties on financial institutions' behaviour. This could involve developing regulations that require financial firms to report all transactions they make, to help verify financial statements. Currently, auditors conduct third-party confirmations on a sample basis and the audited financial statements are the basis for regulatory returns. Given the very large number of transactions conducted, transaction reporting could become onerous if it were extended to a larger sample and is likely to be unaffordable with the FCA's current budget. It is, therefore, worth investigating whether some aggregated data could be used to identify possible areas of concern in particular businesses.
4. Using random audits to identify the characteristics of businesses most likely to violate the regulations, and using the results of those audits to improve the targeting of future audits of financial institutions. This is practiced in the United States by the Securities and Exchange Commission and has been proposed for the FCA. This could prove to be cost-effective and deserves serious consideration.
5. Developing a system of "cooperative compliance" for financial institutions that are particularly large or particularly important in the financial system. The FCA is already doing this for "fixed portfolio" firms and it would be useful to evaluate this experience and consider its extension to a larger group of businesses.
6. Developing programmes like HMRC's LBS. The FCA already conducts something similar but its effects could usefully be reviewed, with the possibility of improving it.
7. Developing programmes like HMRC's HRCP programme, so surveillance is focused on those financial institutions that pose the greatest risks. The FCA already has a programme along these lines.
8. Adopting rules like HMRC's anti-avoidance measures: DOTAS and GAAR.

The first measure is basic. HMRC deals with non-compliance by both recovering the additional tax due and applying a fine that is a substantial fraction of that additional tax. The second measure supplements this by providing an incentive for senior management to ensure that their regulatory reports are accurate.

The third and fourth measures could be used to identify misreporting among the whole population of financial firms. In addition, techniques for identifying misreporting in the financial services sector, such as textual analysis of reports (as used by the U.S. Securities and Exchange Commission in the form of Latent Dirichlet Analysis), could be used for the same purpose.

The fifth and sixth measures are potentially suitable for dealing with large and important financial institutions. These would involve developing a close relationship between FCA officials and specific firms, and so would only be practicable for a fairly small number of firms. While these arrangements could bring improved compliance, there is also a risk of "regulatory capture," where the FCA official could identify too much with the firm they are responsible for monitoring. It is therefore important to have a clear system of governance for managing these close relationships.

Last, the seventh and eighth measures could be suitable for dealing with firms that appear to pose serious compliance problems. As discussed in section 3, the seventh involves contacting senior management in risky firms to try to resolve problems by agreement or, if that fails, by enforcement action. The eighth would be particularly designed for dealing with attempts to get around financial regulations by disguising the true nature of complex transactions. Some equivalent of the Disclosure of Avoidance Schemes could help identify such practices, while legislation like that used to establish GAAR could help in forcing firms to stop using regulatory avoidance measures. It is particularly important that both the promoter of an avoidance scheme and its users should have a disclosure obligation.

Measuring Non-Compliance in Banking

As discussed in section 3.2, HMRC attempts to measure the tax gap: the difference between taxes legally due and taxes collected. This is a difficult task and is bound to be subject to error. Nonetheless, it is valuable to HMRC for two reasons:

1. It provides HMRC with a measure of how well it is doing in collecting taxes and whether it is getting better at it.
2. By breaking down the tax gap into its component parts, the tax gap for each type of tax, it can identify where it is having difficulty in collecting taxes. This can help in deciding where to focus the available resources to carry out enforcement activity, although it should also take account of the resource cost of improving collection in different areas.

This naturally raises the question whether the FCA should measure an equivalent "regulatory gap." It could use the outcomes of the various detection methods, outlined above, to estimate the prevalence of regulatory avoidance in much the same way as HMRC estimates the prevalence of tax avoidance. It would clearly require careful reflection on what is to be measured and how the available information can be summarised usefully to produce a "regulatory gap" measure like HMRC's tax gap.

The difficulty is that, while HMRC has a simple way of measuring tax compliance (money), there is no similarly obvious way of aggregating the various types of regulatory non-compliance into a single measure. Certainly, simply counting the estimated number of regulatory violations is not particularly useful, because some violations are much more serious than others. What is needed is a measure of the seriousness of each violation. To some extent, presumably, the size of the fines imposed for regulatory violations could be used, if the fines are a reasonable estimate of the damage caused by the violation. It is unclear whether this is currently the case. Perhaps an alternative would be to set up a set of criteria by which the seriousness of each violation can be assessed, and use expert judgement (combined with any available empirical evidence) to produce seriousness index. This could be combined with measures of market efficiency to provide an overall index of regulatory success. Also, as previously mentioned, complaints data and a "Trip Adviser" for financial services might have roles to play. Over time, the index could be modified as experience of its use accumulates.

6 Conclusions

This paper has outlined the economic theory of compliance and the strategies it suggests for improving compliance. It has also outlined the key aspects of how HMRC operates, emphasising the methods used to improve compliance and measuring the tax gap (an indicator of compliance). The compliance activities of HMRC were interpreted relative to the theories of incentives and tax compliance. It then proceeded to discuss the regulatory activities of the FCA and of the many parallels with the compliance practices of HMRC. In both cases, there is a concern with the prevailing culture with an effort being made to push the culture to one of compliance.

The key elements identified in the paper are:

- Ensuring that the penalty for non-compliance is large enough to provide a real incentive to comply, and recommending that at least some of the penalty should be borne by a senior executive of the firm. The fines levied by the FCA appear large but need to be given greater context when reported since their size in relation to the estimated benefits from non-compliance is unclear.
- Methods for identifying non-compliance among the whole population of financial institutions, such as an auditing strategy and third-party reporting. HMRC have moved to real-time monitoring of large businesses' tax compliance. The FCA has the data necessary to undertake real-time monitoring but the complexity of the task creates a technological challenge. This is an area in which both could enhance their future compliance activities. Real-time monitoring would be particularly beneficial for the early identification of mis-selling. The application of "big data" analytics should be a priority area for development at the FCA.
- Methods for focusing on large or particularly important financial institutions through closer engagement. The process of sustainable regulation should be developed further, but with monitoring of its impact on culture and compliance. HMRC has been accused of making "sweet heart deals," and this downside of close relationships must be avoided.
- Methods for dealing with financial institutions at high risk of regulatory non-compliance, by high-level engagement and the mandatory disclosure of regulatory avoidance schemes, including, for example, products designed to avoid their natural and appropriate regulatory treatment.

Finally, the paper has considered whether it would be practical to devise a measure of the "regulatory gap" to indicate the overall level of compliance. Such a measure would allow the success of regulatory activity to be tracked over time and provide a guide to focus the allocation of enforcement and supervision resources.

7 Appendix

A1. The Principal-Agent Model

The principal-agent model captures the essential features of all incentive problems. The model has an asymmetry of information so the principal, who designs the incentive contract, cannot observe all the actions of the agent. If there were complete information, a contract could be implemented that would ensure the agent did precisely what the principal wished. There is also a partial conflict between the interests of the principal and the agent. If interests were directly aligned, then the agent would not need an incentive to act in the best interest of the principal.

The model can be illustrated by considering the owner of a firm who must motivate a manager to run the firm in the interest of the owner. The information restriction is that the owner cannot observe the effort level of the manager. What can be observed is the profit level of the firm, so the reward for the manager must be conditioned upon this. Profit is related to effort, but only imperfectly. The owner needs to reward the manager for making effort but not wish to be excessively generous.

A formal representation of this situation is as follows. The agent can choose low effort (e_l) or high effort (e_h). When low effort is supplied the probability of earning high profit, π_h, is P_l and when high effort is supplied the probability is $P_h > P_l$. The corresponding probabilities of low profit, π_l, are $1 - P_l$ and $1 - P_h$. If profit level i occurs ($i = l, h$) the pay-off to the agent is determined by the wage received and the level of effort expended.

$$U(w_i, e_j) \qquad \text{(A1)}$$

The principal must determine the profit-dependent wages $\{w_l, w_h\}$ to offer the agent. The principal's objective is to maximize the expected level of profit net of the wage paid to the agent

$$E\pi = P_j [\pi_h - w_h] + [1 - P_j][\pi_l - w_l]. \qquad \text{(A2)}$$

To induce the agent to supply high effort, it must be the case that the expected pay-off of the agent for high effort is no lower than for low effort

$$P_h U(w_h, e_h) + [1 - P_h] U(w_l, e_h) \geq P_l U(w_h, e_l) + [1 - P_l] U(w_l, e_l) \qquad \text{(A3)}$$

This is the incentive compatibility constraint that ensures high effort. The wage offer must also ensure that the agent chooses to work for the principal. If the pay-off to the agent in the outside option is U_0, the participation constraint is that

$$P_h U(w_h, e_h) + [1 - P_h] U(w_l, e_h) \geq U_0 \qquad \text{(A4)}$$

If this constraint were not imposed, the principal would be able to incentivise the agent by setting w_l very low as a punishment for low profit. The participation constraint limits the extent to which such negative incentives can be imposed.

The optimal incentive contract is a pair of wages $\{w_l, w_h\}$ that maximise (A2) subject to the constraints in (A3) and (A4). This optimization can be written as

$$\max_{\{w_h, w_l\}} P_h[\pi_h - w_h] + [1 - P_h][\pi_l - w_l]$$

$$+ \lambda [P_h U(w_h, e_h) + [1 - P_h] U(w_l, e_h) - P_l U(w_h, e_l) - [1 - P_l] U(w_l, e_l)], \qquad \text{(A5)}$$

where λ and μ are the Lagrange multipliers. The necessary conditions are

$$-[1 - p_h] + \lambda[[1 - p_h]U_1(w_l, e_h) - [1 - p_l]U_1(w_l, e_l)] + \mu[1 - p_h]U_1(w_l, e_h) = 0 \qquad \text{(A6)}$$

$$-p_h + \lambda[p_h U_1(w_h, e_h) - p_l U_1(w_h, e_l)] + \mu p_h U_1(w_h, e_h) = 0 \qquad \text{(A7)}$$

The analysis is made more straightforward if we assume separability with

$$U(w_i, e_j) = v(w_i) - g(e_j). \qquad \text{(A8)}$$

It then follows that the optimal wages solve

$$\frac{v'(w_l)}{v'(w_h)} = \frac{\lambda\left[1 - \frac{p_l}{p_h}\right] + \mu}{\lambda\left[1 - \left[\frac{1 - p_l}{[1 - p_h]}\right]\right] + \mu}. \qquad \text{(A9)}$$

With decreasing marginal utility of income, $v''(w_i) < 0$, so (A9) implies that

$$w_h > w_l. \qquad \text{(A10)}$$

The agent is incentivised to provide high effort by rewarding the observation of a high level of profit.

This analysis shows how the optimal incentive contract can be obtained by considering the constraints upon the choice of the principal. Information asymmetry limits the options of the principal. If the principal could observe the effort of the agent, then the wage payment would be conditioned directly on effort and the possibility would never arise of rewarding a low-effort agent for a lucky draw of high profit or punishing a high-effort agent for drawing low profit.

A2. Tax Compliance

The decision to evade taxation is an example of choice under risk. Not all tax evaders are caught by the tax authorities so a taxpayer who evades stands a chance of successfully concealing income. They also face a chance of being caught and punished. In the basic model of Allingham and Sandmo (1972), Srinivasan (1973) and Yitzhaki (1974), the taxpayer chooses the amount of tax to evade, subject to the probability of being caught and punished, to maximise expected utility. A diagrammatic presentation of this model can be found in Hindriks and Myles (2013).

The taxpayer receives income *M* which is known to the taxpayer but not to the revenue service. Declared income, *X*, is taxed at a constant rate t. If the taxpayer is caught evading, which occurs with probability *p*, a fine *F* > 1 is placed upon evaded tax. The taxpayer chooses *X* to maximise expected utility. This decision problem can be written as

$$\max_{\{X\}} EU = [1-p]U(M-tX) + pU(M-tX-Ft[M-X]), \tag{A11}$$

where $Ft[M-X]$ is the total fine paid when caught evading and it is assumed that *U'* > 0 and *U*" < 0. The statement of the fine assumes that an audit discovers the true level of income. In practice, this will rarely be the case. Defining $Y = M - tX$ and $Z = M - tX - Ft[M-X]$, the first- and second-order conditions for maximising (A11) are

$$p[F-1]U'(Z) - [1-p]U'(Y) = 0 \tag{A12}$$

and

$$S \equiv t[[1-p]U''(Y) + [F-1]^2 pU''(Z)] \leq 0 \tag{A13}$$

For tax evasion to take place, the solution to (A12) must have *X* < *M*. When (A13) holds for all X, this will occur when the derivative $\frac{\partial EU}{\partial X}$ is negative when evaluated at *X* = *M*, implying that *X* must be reduced to arrive at the optimum. Calculating the derivative at $X = M$, tax evasion will occur whenever $pF < 1$.

This condition is significant. In practice, the rate of fine, F, will never exceed 3 (a value of $F = 1$ restores the correct level of tax payment and a value of $F = 2$ adds a fine equal to 100 percent of avoided tax). Even if $F = 3$, the model predicts that a taxpayer will choose not to declare all income when $p < 1/3$. This is a level of auditing that no revenue service succeeds in achieving. So if the model correctly captures behaviour, then it provides the strong conclusion that all taxpayers who have an opportunity to evade will do so.

The model can also determine how the level of tax evasion is affected by changes in the key variables. This is necessary as a prelude to understanding how to structure incentives for compliance. Four variables are of interest: the income level M, the tax rate t, the probability of detection p, and the fine rate F. The effects are obtained through comparative statics analysis using (A12) and (A13). Totally differentiating (A12) with respect to X and p gives

$$\frac{dX}{dp} = -\frac{[F-1]U'(Z) + U'(Y)}{S} > 0, \tag{A14}$$

so an increase in the probability of detection raises the level of income declared and reduces evasion. This is a clearly expected result, since an increase in the likelihood of detection lowers the pay-off from evading and makes evasion a less attractive proposition. For a change in the fine rate upon evaded tax

$$\frac{dX}{dF} = -\frac{pU'(Z) - p[F-1]U''(Z)t[M-X]}{S} > 0 \tag{A15}$$

As expected, an increase in the fine leads to a reduction in the level of tax evasion. There is therefore no ambiguity about the effects of the two punishment variables upon the level of evasion.

Differentiating (A12) with respect to X and M and using the Arrow-Pratt measure of absolute risk aversion, $R_A(I) = -\frac{U''(I)}{U'(I)}$, determines the effect of an increase in income upon evasion as

$$\frac{dX}{dM} = \frac{FtR_A(Z) - [R_A(Z) - R_A(Y)]}{FtR_A(Z) - t[R_A(Z) - R_A(Y)]} \quad \text{(A16)}$$

If absolute risk aversion decreases as income increases, then $R_A(Z) - R_A(Y) > 0$. Since $t < 1$, it then follows that that $\frac{dX}{dM} < 1$ and hence that declared income rises less than actual income implying that the amount of tax evaded rises with income.

The final variable is the tax rate. The effect of the tax rate upon tax evasion is given by

$$\frac{dX}{dt} = -\frac{[1-p]U'(Y)[X[R_A(Z) - R_A(Y)] + F[M - X]R_A(Z)]}{S} \quad \text{(A17)}$$

The surprising conclusion is that when absolute risk aversion decreases as income increases, a higher tax rate will lead to a *greater* income declaration and a reduction in evasion. This result has provoked considerable discussion because it is counter to the intuitive expectation that an increase in the tax rate should provide a greater incentive to evade. The explanation for the result can be found in the fact that the penalty for evasion is $Ft[M - X]$ so an increase in the tax rate increases the penalty rate. This has the effect of deterring evasion.

The basic model provides clear predictions about taxpayer compliance behaviour and how this will change as the parameters are varied. However, criticism points out that the model overpredicts the amount of evasion and that the effect of the tax rate is counter to evidence. The evidence itself is mixed and not entirely compelling (see, for example, Feinstein 1991). In any case, numerous extensions claim to provide improved models of the compliance decision. These are surveyed in Hashimzade *et al*. (2013) and Sandmo (2005).

A3. Incentivising Compliance using Audits

The first analysis of incentivising compliance in Kolm (1973) considered how best to jointly set the probability of audit and the fine levied when evasion is detected. The basic question was to see what the outcome would be if the two instruments could be freely varied. The predictions of the basic model are that both an increase in the probability of detection and of the fine will reduce evasion. As noted by Allingham and Sandmo (1972), this makes the two instruments substitutes with respect to reducing evasion since a reduction in one can be compensated for by an increase in the other.

In relating the following discussion to incentive problems in general, it is worth remarking that the true income of the taxpayer is unknown to the revenue service. This is the asymmetry of information that gives the incentive problem content. What is different is that the participation constraint is not normally considered. However, there are at least two ways a participation constraint could be invoked. First, the possibility the taxpayer could move to a new tax jurisdiction if the audit and punishment scheme are too onerous. Second, the taxpayer could choose to become a “ghost” and cease involvement in the formal economy.

The optimal values of *p* and *F* are straightforward to determine. Assume that detection is costly in the sense that resources are required to finance the audit process. An increase in *p* therefore requires additional expenditure to be made by the revenue service. In contrast, alternative levels of *F* have no cost implications so, effectively, increases in *F* are costless to produce. Given this, the optimal combination of *p* and *F* to incentivise compliance should be to set *p* as close to zero as possible and increase *F* without limit. This structure provides maximum deterrence at zero cost. This result is supported by Christiansen (1980) who shows, with a slightly different specification, that when the detection probability and the fine are adjusted to keep the expected gain from tax evasion constant, tax evasion will be reduced by an increase in the fine so the fine is the more efficient deterrent. Further development of this result, introducing non-proportional taxes, can be found in Koskela (1983).

In practice, it is more likely that there is a distinction between the agencies that determine the fine and the audit rate. Punishments are set by the legal system to reflect the perceived seriousness of an offence, and the fine for evasion will fall within the legal framework. Consequently, it is more appropriate to analyse determining the optimal rate of detection in isolation, taking as given the rate of tax and the fine that will be levied. Specifically, the revenue service can be modelled as setting audit strategy to maximize the amount of revenue collected less the cost of enforcing the audit policy.

Reinganum and Wilde (1985) observe that the key element of the audit process is that each taxpayer is required to make an income report to the revenue service and it is through this report that the revenue service makes the decision whether to investigate. The reported income is a *signal* from the taxpayer about true income and the revenue service must extract information from this signal. Given their observed income, which is private information, each taxpayer files an income report and the revenue service decides to audit, or not, based on this report. The reported income is chosen to maximise the welfare of the taxpayer and the audit probabilities are chosen to maximise the net revenue of the revenue service.

Without pre-commitment by the revenue service the audit policy must be optimal given the choices of the taxpayers. This simultaneity in choices makes the Nash equilibrium concept appropriate. Reinganum and Wilde (1986) consider an audit policy as determining the probability of audit, p , as a function of the declared income, X .

Given a reported income X , the function $p(X)$ determines the probability with which an audit will take place. Given a true income level of *M* the taxpayer chooses *X* to maximise

$$EU = p(X)[M[1-t]-tf[M-X]]+[1-p(X)][M-tX] \tag{A18}$$

It is again assumed that an audit is entirely successful in discovering all undeclared income. Given $p(X)$, the maximisation in (A18) determines a report as a function of true income. This function is denoted $r(M)$, hence the report is given by $X = r(M)$.

Following receipt of an income report, the revenue service forms a belief about the true income level of the taxpayer filing the report. Let these beliefs be captured by a function $\tau(X)$, so an income of X is believed to arise from a taxpayer with true income $\tau(X)$. If the cost of achieving a probability ρ of detection is $c(\rho)$, the expected revenue from a taxpayer making report X is given by

$$R(X,\rho;\tau) = \rho[t\tau(X)+ ft[\tau(X)-X]]+[1-\rho]tX - c(\rho) \tag{A19}$$

Given its beliefs, the revenue service chooses ρ to maximize $R(X,\rho;\tau)$.

The equilibrium is assumed to be a separating equilibrium so taxpayers with different incomes file different reports. The report function $r(M)$ will therefore be monotonically increasing in M. A separating equilibrium can then be defined as a set of functions $\{\overline{\tau}(X), \overline{p}(X), \overline{r}(M)\}$ such that $\overline{r}(M)$ is monotonically increasing in M and

(i) Given $\overline{\tau}(X)$, $\overline{p}(X)$maximises $R(X,\rho;\tau)$;

(ii) Given $\overline{p}(X)$, $\overline{r}(M)$ maximises EU;

and

(iii) $\overline{\tau}(\overline{r}(M)) = M$.

Condition (i) ensures that the audit strategy maximizes revenue for given beliefs. Condition (ii) ensures that the reports maximize expected utility given the audit strategy. Condition (iii) is the consistency condition for the beliefs of the revenue service and ensures that they cannot be consistently wrong in their assessment of the income levels from the reports they are given.

Reinganum and Wilde (1986) demonstrate that the optimal audit strategy and taxpayer reporting strategy satisfying these conditions satisfy (i) $\overline{p}'(X) < 0$ and (ii) $0 < \overline{r}^{-1}(X) < 1$. (i) implies that higher income reports are less likely to be audited. The important point is that the relationship is monotonic, so the very lowest reports are audited most and the probability declines as the report increases. This is a standard structure for an audit scheme since it provides the marginal incentive to always report a slightly higher level of income. If the audit probability were non-monotonic – for example, very low reports were rarely audited – this would give an incentive for taxpayers to report where the probability was lowest. These results have the implication that the equilibrium effort at verification diminishes the higher the level of reported income. In addition, the identity $r^{-1'}$ = 1/r' and the result 0< $r^{-1'}$ <1imply that $r'(M) > 1$ and hence that tax evasion declines with income since *M - r(M)* is decreasing in *M*. The second of these two conclusions may seem surprising but if evasion increased with income the revenue service would have an incentive to raise the audit probability on high incomes.

Turning now to the analysis with pre-commitment, somewhat different results are obtained. Allowing pre-commitment permits the adoption of strategies by the revenue service which involve a zero probability of auditing some income levels. In fact, as shown by Scotchmer (1987), the optimal policy takes the form of a cutoff rule in which income reports above a certain level are not audited at all. Since this leads all taxpayers with an income above the cutoff to file a report at the cutoff, this can only be an equilibrium when pre-commitment is allowed because if there were no commitment the revenue service would gain by reneging on the zero-audit probability and catching all the evaders at the cutoff point. This optimal strategy is clearly an exaggeration of the decreasing audit probability shown to be optimal without commitment.

The consequence of the cutoff rule is that the official tax which is proportional at rate *t* becomes regressive since actual tax payment is constant on incomes above the cutoff point and the effective tax function is regressive. This observation illustrates the general point that with tax evasion the properties of the effective tax function may well be very different from those of the official tax function.

8 References

Allingham, M.G., and Sandmo, A. 1972. “Income tax evasion: A theoretical analysis.” Journal of Public Economics, 1, 323 – 338.

Armstrong, M. and Huck, S. (2010) “Behavioral economic as applied to firms: a primer,” Competition Policy International, 6, 1 – 44.

Bernheim, B.D. and Rangel, A. (2005) Behavioral public economics, www.stanford.edu/~bernheim/Behavioral%20Public%20Economics%20Final.pdf

Cabinet Office (2012) “Applying behavioural insights to reduce fraud, error and debt,” London: Cabinet Office Behavioural Insights Team.

Choo, L., Fonseca, M.A., and Myles, G.D. (2016) "Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment," Journal of Economic Behavior and Organization, 124, 102 – 114.

Christiansen, V. (1980) "Two comments on tax evasion," Journal of Public Economics, 13, 389 – 393.

Danielsson, J., James, K.R., Valenzuela, M., and Zer, I. (2016) “Model risk of risk models,” Journal of Financial Stability, 23, 79 – 91.

Feinstein, J. S. (1991) "An Econometric Analysis of Income Tax Evasion and Its Detection," RAND Journal of Economics, 22, 14 – 35.

FCA (2014) Our Strategy, London: Financial Conduct Authority.

FTA (2010) Information Note: Tax Compliance and Tax Accounting Systems, Paris: Forum on Tax Administration.

Galati, G. and Moessner, R. (2013) “Macroprudental policy – a literature review,” Journal of Economic Surveys, 27, 846 – 878.

Gong, R. and Page, F. (2016) “Shadow banks and systemic risk,” LSE Systemic Risk Centre, DP 55.

Grossman, S.J. and Hart, O.D. (1983) “An analysis of the principal-agent problem,” Econometrica, 51, 7 – 45.

Hartnett Review (2001) Review of Links with Business, webarchive.nationalarchives.gov.uk /20070101084523/hmrc.gov.uk/pbr2001/businesslinks.pdf

Hashimzade, N., Myles, G.D., and Tran-Nam, B. (2013) “Applications of behavioural economics to tax evasion,” Journal of Economic Surveys, 27, 941 – 977.

Hashimzade, N. and Myles, G.D. (2015) “The behavioral economics of tax enforcement: an agent-based approach,” Public Finance Review, (online).

Hill, T. and Lewicki, P. (2007) STATISTICS: Methods and Applications, Tulsa: StatSoft.

Hindriks, J. and Myles, G.D. (2013) Intermediate Public Economics, 2nd edition, Cambridge: MIT Press.

HMRC (2012) The Exchequer effect of the 50 percent additional rate of income tax, http://webarchive.nationalarchives.gov.uk/20130129110402/http:/www.hmrc.gov.uk/budget2012/excheq-income-tax-2042.pdf

HMRC (2013) Resolving tax disputes: Commentary on the litigation and settlement strategy, London: HM Revenue and Customs.

HMRC (2014) Measuring tax gaps 2014 edition: Tax gap estimates for 2012-13, London: HM Revenue and Customs.

HMRC (2015) How we resolve Tax Disputes The Tax Assurance Commissioner's Annual Report 2014-15, London: HM Revenue and Customs.

House of Commons Treasury Committee (2012) Closing the Tax Gap: HMRC's record at ensuring tax compliance, Twenty-ninth Report of Session 2010–2012.

International Monetary Fund (2013) Assessment of HMRC's Tax Gap Analysis, Washington: IMF.

Kolm, S.-C. (1973) "A note on optimum tax evasion," Journal of Public Economics, 2, 265 – 270.

Laffont, J.-J. and Martimort, D. (2001) The Theory of Incentives: The Principal-Agent Model, Princeton: Princeton University Press.

Litan, R. and Singer, H. (2015) “Good Intentions Gone Wrong: The Yet‐To‐Be‐Recognized Costs Of the Department Of Labor’s Proposed Fiduciary Rule,” Economists Incorporated.

Mirrlees, J.A. (1974a) "Optimal allocation under uncertainty," in J. Dreze (ed.) Allocation Under Uncertainty, London: Macmillan.

Mirrlees,. J.A. (1975/1999) "The Theory of Moral Hazard and Unobservable Behaviour, Part I," (Mimeo., Nuffield College, Oxford), published in the Review of Economic Studies, 1999, 66, 3 – 21.

Misutka, D. and Eachern, M.L. (2013) “The CRA's tax transparency initiatives: implications for a corporate governance response,” Canadian Tax Journal, 61, 829 – 855.

O'Donoghue, T. and Rabin, M. (2006) “Optimal sin taxes,” Journal of Public Economics, 90, 1825 – 1849.

OECD (2004) Corporate Governance: A Survey of OECD Countries, Paris: OECD Publishing.

OECD (2008) Study into the Role of Tax Intermediaries, Paris: OECD Publishing.

OECD (2013) Cooperative Compliance: A Framework, Paris: OECD Publishing.

Reinganum, J. and Wilde, L. (1985) "Income tax compliance in a principal-agent framework," Journal of Public Economics, 26, 1 – 18.

Reinganum, J. and Wilde, L. (1986) "Equilibrium verification and reporting policies in a model of tax compliance," International Economics Review, 27, 739 – 760.

Sandmo, A. (2005) “The theory of tax evasion: a retrospective view,” National Tax Journal, 58, 643 – 663.

Scotchmer, S. (1987) "Audit classes and tax enforcement policy," American Economic Review, 77, 129 – 136.

Shu, L.L., Mazar, N., Gino, F., Ariely, D., and Bazerman, M.H. (2012) “Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end,” PNAS, 109, 15,197 – 15,200.

Slemrod, J. and Yitzhaki, S. (2002) “Tax evasion, avoidance and administration” in A.J Auerbach and M. Feldstein (eds.), Handbook of Public Economics, Amsterdam: North Holland.

Srinivasan, T.N. (1973) "Tax evasion: a model," Journal of Public Economics, 2, 339 – 346.

Torgler, B. (2001) "What do we know about tax morale and tax compliance?," International Review of Economics and Business, 48, 395 – 419.

Torgler, B. and Schaltegger, C.A. (2006) "Tax morale: A survey with a special focus on Switzerland," Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 142, 395 – 425.

Yitzhaki, S. (1974) "Income tax evasion: a theoretical analysis," Journal of Public Economics, 3, 201 – 202.