The Economic Journal, **109** (*June*), F360–F369. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

APPROACHES FOR ESTIMATING NONCOMPLIANCE: EXAMPLES FROM FEDERAL TAXATION IN THE UNITED STATES

Jonathan S. Feinstein

Crime and underground economic activity are a fact of life around the world. Most societies attempt to control these activities through the expenditure of resources on monitoring, prosecution, punishment, and educational programmes. Gathering statistics about who is engaging in illegal or underground activity, the frequencies with which various crimes or underground transactions are occurring, and the severity or magnitude of these acts, is crucial for making good decisions regarding our allocation of resources in this area. Unfortunately, in general it is very difficult to obtain accurate information about these issues. This is partly because the individuals engaged in such activities wish not to be identified and often take actions to avoid detection, and partly because by their very nature the activities tend not to generate economic, accounting, legal, demographic or other cultural records. The fact that it is difficult to obtain accurate information about the activities, however, does not in any way imply that we should not make a concerted effort to try to obtain the best information we can. I believe the cost associated with attempting to determine the extent and nature of these activities is in most circumstances well below the expected benefit, which includes the likelihood of improved decision-making by relevant government personnel and the general public. As a result, I also believe the debate about developing estimates of illegal or underground economic activity should primarily focus on practical issues related to how best to go about developing estimates that are as reliable as possible.

In this short article I shall discuss the measurement of noncompliance with laws and regulations, my particular area of expertise, and shall not discuss the measurement of other kinds of 'hidden' economic behaviours, such as informal contracting, unmeasured home production, and 'grey-market' activities. In particular, I shall discuss three approaches for estimating the aggregate amount of noncompliance with respect to a specific law or regulation, (i) intensive data collection for a focused subsample of the population, (ii) detection controlled estimation, and (iii) the comparision of estimates generated by distinct datasets and models. For each, I shall describe in general terms what the approach is and how it is implemented, review limitations of the approach, and present an example of the application of the approach to the measurement of tax noncompliance in the United States. My discussion shall necessarily be short and I refer the interested reader to a variety of original research publications for more details, especially about the specific examples.

1. Intensive Data Collection for a Focused Subsample

Perhaps the most straightforward way of obtaining information about noncompliance with respect to a particular law or regulation is through the use of audits or inspections to collect detailed information about the actions of a sample of potential offenders. If a taxpayer has not payed his entire tax obligation, an intensive tax audit is likely to uncover much of what he owes but has not paid; similarly an intensive inspection of a plant is likely to uncover many if not most instances of noncompliance with specific regulatory requirements. In fact, data collected through audits or inspections is utilised by many governmental organisations, including tax authorities and regulatory agencies, to develop estimates of aggregate levels of noncompliance. However, such organisations often simply use data collected during routine enforcement actions for this purpose, and this practice poses statistical difficulties for extrapolating from the audited sample to the entire population because in many cases – for example tax audits, these routine enforcement actions are the result of a careful selection process, which means that the audited group is not representative of the population as a whole. In contrast, the approach I am suggesting requires that the sample be random, so that the results about noncompliance for the sample may be extrapolated to make statistically valid estimates about noncompliance in the population at large. I note, however, that often the rate of noncompliance is known or expected to vary with certain observable characteristics – for example it is well known that certain kinds of taxpayers, such as the self-employed, pay a smaller proportion of their legal tax obligation voluntarily than other kinds, and in these circumstances sampling can be made more efficient by being stratified, oversampling from groups in which noncompliance is a bigger problem. Administratively, the requirement that the sample be random often means that a separate programme must be designed and independently funded for the purposes of collecting this kind of data.

A well known example of this first approach to estimating noncompliance in the United States is the Internal Revenue Service's Taxpayer Compliance Measurement Program or TCMP. The IRS has conducted TCMPs for both households and small businesses; each TCMP has involved intensive audits on a stratified random sample of returns. The most recent TCMP audits of households were for tax years 1982, 1985 and 1988, with the 1988 household TCMP involving audits of approximately 50,000 households (see United States Internal Revenue Service (1996)). Information from the TCMP is used both to estimate the prevalence and magnitude of noncompliance and to assist IRS personnel in developing audit selection methods. Statistics from the 1988 TCMP indicate that about 40% of U.S. households underpaid their taxes for that year, 53% paid correctly, and 7% overpaid. Most overpayments were small – the median overpayment was $158, and were presumably due to errors. Assuming that a taxpayer who makes an error is equally likely to under- as to over-estimate his true tax liability, a small fraction of the underpayments were also likely due to error; but it appears that most were intentional. Further, a sizable minority of taxpayers underpaid their taxes by significant amounts – over one-fourth of all taxpayers (nearly two-thirds of those who underpaid) were found to have underpaid by $1,500 or more. The IRS uses TCMP statistics to generate estimates of the tax gap, the aggregate value of taxes owed but not voluntarily paid; for recent years the tax gap for households is estimated to have exceeded $100 billion.[1]

Although the collection of data for a random sample of potential offenders is a sensible approach to the problem of estimating aggregate noncompliance, it suffers from a number of drawbacks. Perhaps the most important is the cost involved in collecting the data. In general, audits or inspections for estimating noncompliance tend to be significantly more costly than ordinary enforcement actions. One reason for this is that whereas ordinary enforcement actions usually focus on a few critical issues for which the potential offender is suspected of noncompliance, audits for estimating noncompliance tend to be thorough and therefore slower and more expensive. In addition, precisely because the sample is random, relatively few of those included in the sample will be found actually to be in noncompliance, as compared with the proportion of cases for which noncompliance is detected during ordinary enforcement actions; as a result the average monetary return (damages plus fines) or 'direct yield' is likely to be lower than for ordinary enforcement actions.

Standard cost-benefit analysis suggests that the size of the data sample collected for purposes of estimating noncompliance should be increased to the point where the marginal cost of collecting additional data equals the marginal benefit. In this formulation the marginal benefit is the sum of the expectation of the direct audit yield and the expectation of the social benefit associated with the improved precision with which noncompliance can be estimated using the additional data, a benefit which presumably is realised through improved public decision-making about how many resources to allocate to attempting to control the activity. The marginal cost is likely to be high and the direct yield is likely to be low, but I believe the expected benefit associated with improved public decision-making is likely to be substantial, so that in many situations collection of reasonably sized sample is appropriate. In fact, however, there are a variety of political and institutional reasons why samples for estimating noncompliance tend either not to be collected at all, or to be substantially smaller than normative economic theory indicates is appropriate. Random audits or inspections are unpopular with the population of potential offenders, for example taxpayers, who resent in principle intrusive audits that are not motivated by suspicion of wrongdoing. Further, the personnel engaged in enforcement often perceive their own career advancement to be related to the average level of direct yield on the audits they conduct and therefore dislike conducting random audits. Lastly, bureaucrats often perceive that the size of their organisation's budget can most readily be

[1] The TCMP results can also be used to estimate the amount of unreported income, which is more comparable to other measures of underground economic activity.

justified by the 'hard dollars' resulting from direct yields, and so may prefer not to expend scarce resources on random audits.

The IRS's experience with its TCMP illustrates many of the arguments I have made. TCMP audits last longer and cost more than ordinary audits. In addition, the direct yield from TCMP audits is much lower than from ordinary audits; indeed in the mid-1990s the average yield for a non-TCMP household audit was more than $5,500 in additional assessments, while for a TCMP audit for the 1988 tax year it was less than $300. Because yields are so low IRS examiners dislike being assigned to conduct TCMP audits. Many Americans also dislike the TCMP because they object to the notion of being subjected at random to an intrusive and time-consuming audit. In contrast to the last argument I made above, top officials at the IRS have consistently supported the Program and have repeatedly asked Congress for funding for a new TCMP during the last decade, including in 1995 to conduct a TCMP for the 1996 tax year. However, the high cost of the Program coupled with the fact that many Americans dislike it has influenced members of the Congress to oppose the Program, and no funding has been appropriated for a TCMP during this time period.

The second drawback of estimating noncompliance by collecting data about the actions of potential offenders is that, even though the data collection is done in a thorough manner, the data are still likely to fail to include many instances of noncompliance, biasing downwards resulting estimates of noncompliance. Again the IRS's TCMP illustrates this point. The IRS has for many years recognised that TCMP audits fail to uncover all tax evasion and has developed multipliers that 'scale up' the noncompliance actually detected during TCMP audits when computing estimates of the tax gap. The average value of these multipliers is approximately two, and they vary substantially across line items. In the case of data collected through audits or inspections one way to address this problem is to analyse the data using detection controlled estimation, the method I discuss next.

Finally, this approach is difficult to apply for crimes for which there is no way to collect the relevant data about noncompliance, such as theft, violent crimes, and victimless crimes. For these kinds of crimes audits of potential offenders are not likely to turn up direct evidence about the crime – indeed the police usually try to identify perpetrators of such crimes from evidence gathered at 'the scene of the crime.' For theft or violent crimes an alternative approach for estimating noncompliance is through interviews with victims.

2. Detection Controlled Estimation

For most laws and regulations the fundamental reason why it is difficult to estimate aggregate noncompliance is that many violations of the law or regulation remain undetected; as a result these undetected violations are not recorded in compliance data and are not taken into account, either explicitly or implicitly, when estimates of noncompliance are constructed. I have developed an econometric model called detection controlled estimation to address the issue of nondetection (Feinstein, 1990). The model is based on the idea of incorporating the detection process into the statistical analysis of compliance data, thereby explicitly allowing for the possibility of incomplete detection. In particular, the model includes two expressions, one referring to potential offenders and specifying the probability of a violation, and the other referring to monitors and specifying the probability of detection, conditional on a violation having occurred. The two expressions can be jointly estimated and the resulting parameter estimates used to construct an estimate of the proportion of violations that remain undetected. In this section I shall review a simple example of detection controlled estimation, describe how it can be used to estimate undetected noncompliance, illustrate its use in an application to the estimation of tax noncompliance, and finally discuss strengths and weaknesses of the method, including the problem of identification.

Suppose data have been collected for a random sample of potential offenders, that a specific monitor or inspector has been assigned to audit or inspect each potential offender and that the results of the audit or inspection are also available. Note that in the example I am presenting detection controlled estimation is meant to be applied to a random sample, just as for the method discussed in Section 1. The detection controlled methodology can be extended to ordinary enforcement data, by appending an audit selection equation to the model to correct for the fact that a non-random sample of potential offenders is selected for audit. The methodology can also be applied to aggregate crime data.

For a representative potential offender – monitor pair, denoted i, the detection controlled model consists of two expressions. The first expression refers to the probability of a violation and is represented by the following expression:

$$\begin{aligned} Y_{1i} &= \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \epsilon_{1i} \\ L_{1i} &= 1(\textit{violation}) \;\; \textit{if} \;\; Y_{1i} > 0 \\ L_{1i} &= 0(\textit{compliance}) \;\; \textit{if} \;\; Y_{1i} \leqslant 0, \end{aligned} \tag{1}$$

where $\boldsymbol{X}_{1i}$ is a vector of characteristics for the potential offender for the ith case, $\boldsymbol{\beta}_1$ is a vector of parameters, and ϵ_{1i} is a mean zero random disturbance that is drawn from the distribution $F(\cdot)$. This first expression can be generalised in a straightforward fashion to a tobit-like formulation for situations in which the magnitude of the violation is important, for example tax evasion. The second expression refers to the detection process. Conditional on L_{1i} being equal to one (a violation),

$$\begin{aligned} Y_{2i} &= \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \epsilon_{2i} \\ L_{2i} &= 1(\textit{detection}) \;\; \textit{if} \;\; Y_{2i} > 0 \\ L_{2i} &= 0(\textit{no detection}) \;\; \textit{if}\, Y_{2i} \leqslant 0, \end{aligned} \tag{2}$$

where $\boldsymbol{X}_{2i}$ is a vector of characteristics for the detection process for the ith case, $\boldsymbol{\beta}_2$ is a vector of parameters, and ϵ_{2i} is a mean zero random disturbance that is drawn from the distribution $G(\cdot)$. For situations in which violations differ in magnitude, this second expression can also be generalised, to a model of fractional detection in which all, none or some fraction of the total noncompliance is detected. In addition, the model as a whole can be generalised in two ways: to allow for a correlation between ϵ_{1i} and ϵ_{2i}, which might arise for example if monitors are assigned to potential offenders based in part on features of potential offenders that are not observable; and to allow for the possibility of false detection ($L_{1i} = 0$ but $L_{2i} = 1$).

Interestingly, L_{1i} and L_{2i} are separately unobservable. It is only the product $L_{1i}L_{2i}$, which refers to a detected violation, that is observable. Nonetheless, expressions (1) and (2) can be estimated jointly by means of maximum likelihood. In particular, the observations fall into two disjoint sets. One set, labelled A, consists of those cases for which a violation has been detected; for a case i in this set, the likelihood is $F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1)\,G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2)$, referring to the probability that a violation has occurred multiplied times the probability that it has been detected, conditional on its having occurred. The other set, A^c, consists of the remaining cases, for which no detected violation is recorded. Cases in this set fall into two groups, which cannot be distinguished in the data; the first group consists of cases for which no violation has occurred and the second cases for which a violation has occurred but has escaped detection. Summing over these two possibilities, the likelihood for a case i in this set is $1 - F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1) + F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1)[1 - G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2)] = 1 - F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1)\,G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2)$. The log likelihood of all the observations is then

$$\sum_{i\in A} \log[F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1)\,G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2)] + \sum_{i\in A^c} \log[1 - F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1)\,G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2)]. \qquad (3)$$

Denoting the estimates of the parameters obtained through the the maximisation of expression (3) $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_2^*$, the proportion of violations remaining undetected can be computed through a straightforward application of Bayes' Law.[2]

I and several other researchers have used the detected controlled methodology to analyse noncompliance and enforcement patterns for a variety of activities, including tax collection, safety regulation, environmental regulation, and health care screening and utilisation review programmes. In my analysis of tax noncompliance (Feinstein, 1991) I used data drawn from the 1982 and 1985 TCMP datasets, described above, and estimated a model that allows for fractional detection. The results of the analysis indicated that detection of evasion during TCMP examinations is quite imperfect. In particular for each year I computed a mean detection rate for each of more than 40 examiners as well as a grand mean detection rate for the sample as a whole. The results

[2] The formal expression is:

$$(1/T)\sum_{i\in A^c} \frac{F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1^*)[1 - G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2^*)]}{1 - F(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1^*)\,G(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2^*)}.$$

For purposes of comparision, note that, if the the total sample size (the sum of the number of cases in A and A^c) is T, the proportion of potential offenders for whom violations are detected simply equals the number of cases in A divided by T and that the proportion of violations that are detected equals the number of cases in A divided by the number of cases in A plus the estimate of this expression. It is straightforward to work out the standard error of the estimate generated by this expression.

implied significant variation in mean detection rates among examiners and that the grand mean was approximately 50%. I then used the results of the analysis to develop an estimate of the tax gap. Interestingly my estimates were nearly identical to the IRS estimates previously mentioned. This is somewhat surprising because although both sets of estimates are based on TCMP data, they were developed independently through quite distinct methods of analysis. In particular, the IRS estimates rely on multipliers that allow for varying rates of detection across different line items, but implicitly assume that all examiners are equally competent at detecting noncompliance on any particular line item, whereas the detection controlled method allows for differing detection rates across examiners but does not, at least in the version of the model on which I based my estimates, allow for variation in detection rates across line items. Developing a model that encompasses both approaches would be a useful task for future researchers.

There are both strengths and weaknesses associated with use of the detection controlled methodology to estimate noncompliance. I believe there are two principle advantages to this approach. First, the methodology focuses attention on the detection process, incorporating it into the analysis; often discussion and estimation of noncompliance ignore detection, which is unfortunate, because it is the failure to detect and record all instances of noncompliance that is in some sense responsible for the difficulty in measuring noncompliance. Second, the detection controlled model generates a precise mathematical formula for estimating noncompliance. The most important weaknesses of the method relate to the fact that it is statistical in nature and not based on detailed information about nondetection in specific cases. The most serious statistical issue is identification of the parameters and distributions of the model. Intuitively, the problem of identification arises because the detection controlled model decomposes a single observable variable, detected violations, into two disjoint causal expressions, violation and detection. As an illustration of the problem consider the following example. Suppose that the probability of potential offender i committing a violation is $F(X_{1i}\boldsymbol{\beta}_1) = p_0 e^{X_{1i}\boldsymbol{\beta}_1}$, where p_0 is the average level of noncompliance in the population and $e^{X_{1i}\boldsymbol{\beta}_1}$ fluctuates around one. Similarly, suppose that $G(X_{2i}\boldsymbol{\beta}_2) = q_0 e^{X_{2i}\boldsymbol{\beta}_2}$. Data are available only for detected violations, the product FG on which the likelihood in expression (3) depends. In this example that product is $e^{X_{1i}\boldsymbol{\beta}_1} p_0 q_0 e^{X_{2i}\boldsymbol{\beta}_2}$, and it follows that p_0 and q_0 cannot be separately identified, only their product $p_0 q_0$. Put differently, a given average level of detected violations (the product $p_0 q_0$) might refer to a high average level of violation and low average detection rate (high p_0 and low q_0) or to the converse, and we cannot distinguish between these two possibilities when we possess data only about detected violations. It can be shown that the double-exponential is the only function form for the distributions F and G for which identification formally fails, so that the detection controlled model is semi-parameterically identified for all other distributions. Nonetheless, the example serves as a warning that identification of average levels of noncompliance is difficult in these models.

3. Comparison of Estimates Generated by Distinct Datasets and Models

In many circumstances noncompliance can be estimated indirectly, either through a comparison of compliance data with other data that measure a closely related activity, or through estimation of a 'residual' effect in which observable behaviours are taken into account and the residual is assumed to refer to concealed or unreported noncompliant behaviours. Many applications of this approach use macroeconomic data about cash holdings or monetary transactions to measure illicit economic activity (see Cagan (1958), Feige (1989); see Schneider and Enste (1998) for a comprehensive review, critique and some recent applications). Although these studies make strong assumptions that are open to criticism, the general conceptual approach of comparing estimates from alternative data sources is likely to be a fruitful one for estimating noncompliance in a wide variety of contexts. In the remainder of this section I shall outline my own use of this method for the estimation of noncompliance with estate and gift tax filing laws in the United States (Feinstein, 1997); then I shall discuss the need for formal modelling in applications of this approach.

Consider the estimation of the incidence of nonfiling of estate tax returns and the associated estate tax gap due to nonfilers. In the United States an estate tax return must be filed for any decedent whose estate satisfies certain conditions, of which the most important is that the gross value of the estate exceeds a threshold value; the threshold has been $600,000 and will rise gradually over the next decade.[3] The IRS tabulates the number of estate return filings for each year; for decedents who died during 1992, the year I focused on in my analysis, IRS statistics indicate that approximately 60,000 returns were filed.

My strategy for estimating the estate tax gap due to nonfilers involved using several non-IRS datasets to generate an estimate of the number of decedents for whom estate tax returns would have been expected to have been filed, and then comparing this estimate with the IRS tabulation of actual filings. My calculation of the number of estates for which a return would be expected to be filed proceeded in three main steps. First, I used asset data from two datasets, the Health and Retirement Survey (HRS) and the Assets and Health Dynamics Among the Oldest Old (AHEAD), to construct a measure of the distribution of assets in the population.[4] I computed asset distributions for each of 80 separate population subgroups or cells by dividing the U.S. population over the age of 50 into 5 age groups, the two sexes, two racial groups – white and non-white, and four marital status groups; an example of a cell is married white women between the ages of 70 and 74. For individuals

[3] Tax is not owed in all cases in which a return must be filed, mainly for two reasons: certain fees can be deducted from the gross value prior to the tax calculation; and a decedent who was married at the time of death could will any part or all of his estate to his spouse as a marital bequest, in which case tax is deferred until the death of the spouse.

[4] At the time of my analysis HRS contained information about households for which the head was (approximately) between ages 51 and 63, while AHEAD provided information about households for which the head was age 70 or above.

who were not married, I defined estate value to be equal to a standard measure of household net worth plus life insurance proceeds, while for married persons I defined it to be equal to household net worth divided by two plus life insurance proceeds. I used these estimated distributions to estimate, for each cell, the proportion of individuals whose assets exceed the $600,000 filing threshold. Second, I obtained data about 1992 deaths in the U.S. population from the National Center for Health Statistics, again for each of the 80 cells defined above. Finally, for each cell I multiplied the number of deaths times the proportion of individuals in the cell with assets above $600,000 to obtain an estimate of the number of estates in the cell for which an estate tax form should have been filed for 1992. I estimated that the actual number of deaths for which a filing should have been made was approximately 90,000, or significantly more than the 60,000 forms actually filed. Further, I found that the predicted number of filings was approximately equal to actual filings for decedents over the age of 80, but significantly greater for decedents under 80.

Although my analysis was suggestive, it failed to take into account a number of relevant factors. Most importantly, in my simple calculation I did not control for the relationship between socioeconomic status and mortality risk, specifically that individuals who are wealthier face a significantly lower mortality risk than those who are poorer. Additional calculations I performed suggested that, if such a correction were made, the number of predicted filings would not be significantly larger than the actual number of filings reported by the IRS. Thus my overall conclusion was that it is unlikely that there is a significant noncompliance problem with filing of the estate tax form in the United States. The fact that my conclusions changed substantially when I incorporated the effect of socioeconomic status on mortality risk illustrates one of the major challenges in applying this third approach, which is that it may be difficult to think of and include in the analysis all the relevant factors that might contribute to a disparity between two different estimates that are meant to refer to the same underlying phenomenom.

As part of my analysis I also investigated noncompliance with the gift tax, using an even simpler approach. In the United States a household must pay a gift tax if during a year it makes a gift or gifts to an individual the total monetary value of which exceeds the household's gift tax threshold.[5] Both HRS and AHEAD ask about gifts so I was able to compare the responses in these surveys to IRS tabulations, again for 1992. My analysis suggested that there is a substantial noncompliance problem: the IRS collected approximately one billion in gift taxes in 1992, but extrapolations from the HRS and AHEAD data indicate that households most likely owed more than three billion in taxes.

It is my sense that the general approach of measuring noncompliance by developing and comparing multiple estimates or through calculation of a residual has been widely used in an informal manner, but as far as I am aware there has been little work on the development of a formal methodology for

[5] This threshold is $20,000 for married couples and $10,000 for others.

applying this approach. I believe the development of a formal methodology, especially an econometric methodology, would be extremely useful for several reasons. First, such a methodology would provide a framework which practitioners could draw upon in the development of specific models. Practitioners might directly apply the general framework; alternatively, the framework might clarify how a particular application is related to previous work. Second, a framework would be of assistance in the identification of likely sources of error and the quantification of confidence bounds for estimates. Third, once many applications, all based upon a common methodology, had been developed, a meta-analysis encompassing the various studies might be performed; such an exercise would surely be useful, since information about illegal activity is scarce and individual estimates are extremely uncertain.

4. Conclusion

As long as criminal activity in all its myriad forms remains an important social problem there will be a need for some means of assessing the nature, incidence and magnitude of this activity, both for general education and for policy formulation. In my opinion the appropriate approach for gathering information about these activities should include an acknowledgment of the need to develop statistics that are as accurate as possible, work to develop systematic methods for measurement, and ongoing critical debate about the methods that have been developed and implemented.

Yale School of Management

References

Cagan, Phillip. (1958), 'The demand for currency relative to the total money supply.' *Journal of Political Economy*, vol. 66, pp. 302–28.

Feige, Edward L., ed. (1989), *The Underground Economies: Tax Evasion and Information Distortion*. New York: Cambridge University Press.

Feinstein, Jonathan S. (1990), 'Detection controlled estimation.' *Journal of Law and Economics*, vol. 33, pp. 233–76.

Feinstein, Jonathan S. (1991), 'An econometric analysis of income tax evasion and its detection.' *RAND Journal of Economics*, vol. 22, no. 1 (Spring), pages 14–35.

Feinstein, Jonathan S. (1997), 'Strategies for estimating the estate and gift tax gap.' Mimeo (September).

Schneider, Friedrich and Enste, Dominik (1998), 'Increasing shadow economies all over the world – fiction or reality?' Manuscript.

United States Internal Revenue Service. (1996), 'Federal tax compliance research: individual income tax gap estimates for 1985, 1988, and 1992.' Publication 1415 (Revised April), Washington D.C.