# Statistical Analysis of Compliance Using the NRP Data: Detection Controlled Models

Jonathan S. Feinstein
Yale School of Management

Presentation at IRS Research Conference
June 2, 2004
Washington D.C.

# Overview

Development of statistical methodology – based on detection controlled modeling -- for using the National Research Program (NRP) database to estimate individual income tax reporting noncompliance and the associated "tax gap." In this brief presentation:

- The NRP.
- Idea of Detection Controlled Estimation.
- Simplest estimation of DCE model using calibration sample from NRP.
- Base model using full sample: estimation of classification process jointly with noncompliance and detection.
- More complex models based on classification of issues on returns.

# The NRP

Historically the IRS measured income tax reporting compliance through the Taxpayer Compliance Measurement Program (TCMP). Thorough audits of stratified random sample of taxpayers; potentially high taxpayer burden. Last TCMP for 1988, thus not current. (Evidence on degradation over time – no change rate rise.) Evidence of substantial non-detection in TCMP from both multiplier studies and detection controlled modeling.

IRS initiated the National Research Program (NRP) to collect current information - current NRP for tax year 2001. (IRS personnel expert on NRP.)

- NRP is again a stratified random sample – just under 50,000 taxpayers.
- Not all taxpayers in sample subject to full audit; rather, some only limited contact or accepted. For those subject to audit issues classified – extensive audit coverage but not exhaustive. Reduces burden; but raises issues for statistical analysis of the data, including non-detection. Supplemental data.
- The calibration sample – as a control. All subject to audit, classification of issues aggressive. As currently configured about 1100 taxpayers (?).

# Idea of Detection Controlled Estimation

• Incorporate a model of the detection process explicitly into the statistical analysis.

• Model factors that influence likelihood of detection, and fraction of noncompliance detected, on return or case. Most importantly, identity (kept anonymous) of examiner conducting the audit, as well as examiner grade. Other factors include district, type of audit.

• Full model has two equations: (i) model of noncompliance (traditional in tax evasion and other applications); (ii) model of detection process. The two are estimated jointly (via maximum likelihood). Variation in the data (returns with high or low detected noncompliance) is factored out to individual factors influencing degree of noncompliance and factors influencing detection rate.

• Model of fractional detection used to estimate tax gap using TCMP data (RAND Journal): evidence of considerable variation in detection rates across examiners (controlling for kinds of returns assigned); overall multiplier of 2 for estimating tax gap.

# Simple Estimation of Noncompliance: DCE Using Calibration Sample

The calibration sample is a stratified random sample, with weights (can be constructed) aggregating to U.S. population. Can directly estimate basic fractional detection model using this dataset:

- Equation 1: Noncompliance. Tobit model: evasion can be 0 or some positive value (extension to over-reporting).
- Equation 2: Fractional Detection. Conditional on evasion detection can be: (i) zero – no evasion detected; (ii) 1 – complete detection; (iii) a fraction r between 0 and 1 (r can be any value) – partial detection.
- Error terms between the two equations are correlated. Dummy variable in detection equation for examiner (if enough cases – questionable).
- Most directly comparable to TCMP sample and estimation.
- But – (i) a small sample and (ii) may not be enough cases per examiner to identify detection equation.

# Detection Controlled Model Using Full NRP Sample

In the full sample returns are classified into three groups: (i) accepted (no change or contact with taxpayer); (ii) correspondence audit – a few issues identified that are examined; (iii) full audit – with a set of issues identified for examination (other issues may arise during audit). Clearly noncompliance on a return accepted or an issue on a correspondence return not identified escapes detection. On a full audit return, detection is likely to be higher for issues identified than for those not identified.

For the calibration sample, we *know* which group a return was classified into before being selected into the calibration sample. Extremely valuable. (If return had been subject to full sample procedure (with exam or contact), then calibration audit, could do even more; but burdensome, not done.)

# Model with Classification and DCE

• First stage: estimate a model of classification of returns into the three groups: accepted; correspondence; full. Use an ordered probit or equivalent. All returns used.

• Second stage: estimate noncompliance and detection equations for each of the three groups:

- For the accepted group, estimate noncompliance and detection using returns in the calibration sample that were initially placed in the accepted group.

- For correspondence group: estimate noncompliance and detection on full sample, with dummy in detection equation for being in calibration sample; or divide into *identified issues* and *remainder*, estimate noncompliance and detection separately (correlated), with detection zero for *remainder* for non-calibration.

- For full audit group: estimate noncompliance and detection on full sample, with dummy variable in detection equation for being in the calibration sample.

The classification process may be correlated with the detection and noncompliance processes. So estimate as a system, allowing errors to be correlated.

Advantage of this model: allows detection processes to be different for the different groups (and controls for selection of which group in correlated with this). If there are significant differences in detected noncompliance between the calibration and non-calibration samples this model can help us understand whether this difference is due to returns being classified as accepted or correspondence that in fact have significant noncompliance (for correspondence, on non-identified issues), or alternatively if there is a significant difference in detection between calibration and non-calibration returns within groups – eg, within the full sample group and within identified issues on correspondence audits.

Even simpler model: pool all returns; one noncompliance equation; one detection equation, with dummies for which group placed in, and for whether calibration sample. (Learn less about classification process.)

# More Complex Models: Issue Selection

Above model does not address the selection of issues for examination, especially in full audit group (the vast majority of returns). There are many issues (items) on the tax return, hence these models can be quite complicated. Suggested structure:

• Break up return into blocks: for example 1040 and A; C; remainder; with a few clusters of items in each block (say in C, income and deductions; perhaps for each business).

• Error components structure: an overall individual effect $\delta$; then within each block errors are allowed to be correlated; but independent across blocks. So for block 1 $\varepsilon_{11}$,$\varepsilon_{12,}$, for block 2 $\varepsilon_{21}$,$\varepsilon_{22}$, ….

• Detection error structure mirrors compliance structure, with correlations between corresponding pairs – eg, overall detection error, $\eta$, correlated with $\delta$, and within blocks detection error terms allowed to be correlated with compliance errors.

• Classification process error $\tau$; may be correlated with $\delta$ and $\eta$.

- Estimation: integrate out over $\delta$; estimate classification, giving restriction on $\tau$; integrate over $\eta$ and estimate compliance and detection models for each block, as (eg) bivariate or quadrivariate normals.

# Benefits of More Complex Model

The more complex model allows the detection and noncompliance equations to be estimated separately for different parts of the return. This may allow improved projection of noncompliance for purposes of estimating the tax gap.

# Conclusion

I have proposed a series of models, of increasing complexity, to model statistically, using detection controlled approach, noncompliance, detection, and classification processes.

Results are then used to forecast undetected noncompliance on each return, then weighted up to U.S. population. Overall tax gap is then:

*detected noncompliance + estimate of undetected noncompliance*

Standard errors of estimates can be computed.

Models can also be used to explore factors associated with noncompliance; and to study classification and detection.