TECHNICAL WORKING PAPER SERIES

OBSERVATIONAL AGENCY AND SUPPLY-SIDE ECONOMETRICS

Tomas Philipson

Technical Working Paper 210

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1997

I thank John Cawley, David Grabowski, and Charlie Mullin for their research assistance, and Gary Becker, Michael Boozer, Norman Bradburn, Phil DePoy, William Dow, James Heckman, Joseph Hotz, Edward Kaplan, Michael Kremer, and Casey Mulligan for comments. Seminar participants at The University of Chicago, Harvard University, Yale University, The 1995 NSF/CEME Micro Econometrics Conference at the University of Wisconsin, and Santa Fe Institute also provided useful input. I am especially thankful to many members of The National Opinion Research Center (NORC) for many discussions, especially Craig Hill and Krishna Winfrey at the Health Section of NORC for helping me produce the data used. Financial support through The Alfred P. Sloan Foundation's Faculty Research Fellowship is gratefully acknowledged. This paper is part of NBER's research program in Health Care. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Observational Agency and Supply-Side
Econometrics
Tomas Philipson
NBER Technical Working Paper No. 210
February 1997
Health Care

ABSTRACT

A central problem in applied empirical work is to separate out the patterns in the data that are due to poor production of the data, such as e.g. non-response and measurement errors, from the patterns attributable to the economic phenomena studied. This paper interprets this inference problem as being an agency problem in the market for observations and suggests ways in which using incentives may be useful to overcome it. The paper discusses how wage discrimination may be used for identification of economic parameters of interest taking into account the responses in survey supply by sample members to that discrimination. Random wage discrimination alters the supply behavior of sample members across the same types of populations in terms of outcomes and thereby allows for separating out poor supply from the population parameters of economic interest. Empirical evidence for a survey of US physicians suggests that survey supply even for this wealthy group is affected by the types of wage discrimination schemes discussed in a manner that makes the schemes useful for identification purposes. Using such schemes to correct mean estimates of physician earnings increases those earnings by about one third.

Tomas Philipson
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
toma1@cicero.spc.uchicago.edu

1 Introduction

Much of economic data is obtained through surveys which record the behavior or experiences of its sample members or through social experiments in which treatments are assigned by investigators. Whether the data of surveys or experiments are collected through mail, phone, interview, or direct observation, the difficulties introduced of not being able to produce data for all sample members, so called non-response, or production of erroneous data, so called measurement errors, are obviously well-known in applied work. More precisely, a central problem in applied work is to separate out the patterns in the data that are due to such problems with the data itself from the patterns that are of economic interest to be studied. Therefore, in desiring to learn about the economic phenomena of interest, investigators often would prefer to be able to attribute the impact of poorly produced data on the coefficient estimates of ultimate interest.

This paper interprets this inference problem as being an agency problem in the market for observations and suggests ways in which using incentives may be useful to overcome it. In this agency problem, the ultimate goal is to learn about the labor force made up by the sample. The relationship is one in which there is asymmetric information between the privately informed sample members and their principal investigator. The information asymmetry is essential because the production of the data would be unnecessary if the principal knew the information of his agents. However, the agency problem arises because the sample member has little stake in the principal's product and effort is unobservable; observations with measurement error cannot be separated out from sampling variance on the individual level.

It follows that two dimensions affect the data produced. The first is the type of population sampled and the second is the observational supply behavior of that population. The main point of this paper is that a solution to the problem of separating out these two dimensions, supply behavior and population parameters, involves using random price discrimination across the sample. This generates differences in supply behavior across sub-samples for which both population parameters and supply schedules are the same on average. This in turn allows one to separate out the effects of imperfect supply behavior from the population parameters of economic interest.

This observational agency problem studied, and the random incentives we propose to solve it, involve several margins on the supply-side. The first is the extent of measurement errors which may be interpreted to correspond to the intensive quality margin of supplying the truth conditional on supplying anything. The second is the participation rate of the sample, the so called response rate, which may interpreted to correspond to the extensive margin of supplying observations at all. Lastly, there is the margin which concerns the size of the entire sample as opposed to the two quality margins within it. All these margins of supply together with the true distribution of outcomes contribute to the produced data and the paper proceeds by dealing with each margin on a subsequently more general level.

Section 2 first discusses the use of incentives to asses the impact of measurement errors assuming that the sample is fixed in size and everyone participates. The standard, but implicitly *behavioral*, assumptions made on the supply-side in this context is that it is only of concern to micro-data because it cancels out in the aggregate (zero mean) and is completely inelastic with respect to all observable or unobservable factors (independently distributed with covariates). This section shows how to test such common behavioral assumptions on observational supply by the use of so called validation incentives. The paper also discusses how to use such incentives to identify population parameters of substantive interest regardless of whether such behavioral assumptions hold.

Section 3 goes on to consider how incentives may be used to asses the impact of measurement errors when they interact with incomplete participation of sample members. Of key importance here are the multi-task agency tradeoffs that arise because participation is observable to the principal but measurement errors are not. The tradeoff between observable and unobservable activities, in particular the reduction in unobservable effort to increased incentives on observable effort, has been stressed repeatedly in other agency problems.[2] Here the tradeoff arises because randomly compensating for observable performance in terms of participation may result in larger unobservable measurement errors when the two are substitutes. The paper suggests how to use data available in the frame of the survey, i.e. the listing from which the sample of the survey is initially drawn, to estimate the cross- elasticities between the two margins and show how to correct data for supply-side effects using such incentives. The type of incentives includes both pecuniary ones and those in-kind in nature such as when paying sample members in an

[2]See e.g. Lazear (1986), Milgrom and Holmstrom (1991), and Baker (1992)).

income survey with confidentiality instead of dollars.

Section 4 considers how incentives affect the tradeoff between the two quality margins discussed and the total size of the sample employed. This quality-quantity tradeoff in survey production is important for optimal sample size determinations. The elasticity of supply greatly affects such optimal sample sizes because it affects how many observations must be foregone to finance an increase in performance within the survey, e.g. through a higher participation rate or lower measurement errors.

Section 5 provides empirical evidence using data from an experiment specifically designed to investigate our type of incentives; *The Survey Supply Experiment* conducted at The National Opinion Research Center (NORC) at The University of Chicago. This experiment was added to a national survey of wealthy US physicians; a group with median income of about 200 thousand dollars and therefore a group that a priori would seem inelastic to our types of survey incentives. However, the empirical results support that survey supply is highly elastic to the types of incentives discussed even among these physicians. Interestingly, we find that the two quality margins, errorless supply and participation, are complements rather than substitutes as stressed in other multi-task agency problems. Using our incentives to estimate mean earnings levels of these physicians, the paper finds quite dramatic production biases in these earnings; mean incomes are about a third higher than would be estimated without our random incentives. Our findings suggest that separating out poor survey supply from population parameters of economic interest using our randomized incentives may be feasible and important.

The paper relates to several strands of previous work. There is, of course, extensive literature on the designs of surveys and experiments, indeed large enough to make a review of it meaningless in the space provided here.[3] However, the orthodox theory of data production is a single-person theory about the *demand-side* in the market for observations[4] where the single person is the investigator who makes his production choices in absence of consideration of

[3]Statistical classics on experimental design include Cox(1958), Cochrane (1979), Fleiss (1982) and on survey design Hansen, Hurwitz, and Madow (1953), Cochrane (1979), Kish (1986), Bradburn and Sudman (1988), Beimer et al. (1992), and Lessler and Kalsbeek (1992).

[4]This single person focus is implicit in the well-known treatments of Savage (1977) and Berger (1987) in general or Cochrane (1979) for survey techniques.

the interactions or exchanges with other factors of production, in particular the supply side of observations. Indeed, econometrics seems almost exclusively concerned with the consumption of data on the demand side, rather than the supply-side aspects of its production as discussed here.[5] Furthermore, there is a considerable distance between theoretical work on survey design, which ignores incentives on the supply side, and survey practice, which has long used incentives to get individuals to participate in surveys. The novel proposal here is to use such incentives to identify population parameters of economic interest by separating out survey supply from those parameters.

2 Measurement Errors and Random Incentives

This section considers using randomized incentives to assess the bias introduced by measurement errors. It is first assumed that the sample is fixed in size and that everyone participates, but this is generalized in later sections to include aspects of non-participation. Following standard notation for measurement errors, let Y denote the supplied outcome and Y^* the true outcome. It is assumed that the problem of measurement errors occurs because there is asymmetric information between sample members and the investigator; sample members are assumed to know their value of Y^* while investigators only observe the supplied outcome Y. The inference problem lies in designing incentives so that supply behavior can be separated out from the distribution of Y^*.[6] A key type of incentive considered will be *validation incentives*, which are defined by monitoring π percent of the sample and paying them w in case of a correct response $y = y^*$. The incentives of sample members to report correctly are captured in the utility function $U(y, w|y^*)$, where y and y^* are values of Y and Y^* corresponding to the supplied and true status of the sample member. The truth is supplied at wage w if the expected utility

[5] However, see Griliches (1986) who discusses aspects of data quality.

[6] Manski (1990) and Philipson (1997) discuss other sources of measurement errors such as incomplete knowledge about the future or low effort on the part of agents but are not concerned with incentives in general and identification problems discussed here in particular. Furthermore, as long as the supply of truthful observations is still elastic to the compensation schemes discussed, our results are generalizable.

of doing so is larger

$$\pi U(y^*, w|y^*) + (1-\pi)U(y^*, 0|y^*) \geq U(y(y^*), 0|y^*)$$

where $y(y^*)$ is the most preferred outcome of type y^* in absence of any incentives. This implies a reservation wage property of supplying the truth above which sample members sell the truth. In other words, there exists a wage level z such that $y = y^*$ if $w \geq z$ and $y = y(y^*)$ if $w \leq z$. This follows directly from that the benefit of selling the truth, but not its cost, rises with the wage.

A given validation incentive spends πw on sample members plus the costs of the validation itself; often extremely high. The underlying rationale adapted here is that incentives may be kept close to the same by lowering the validation π but increasing the wage w, thereby cutting capital outlays spent on the monitoring itself.[7] Therefore, the whole point of a validation incentive is to only monitor a very small fraction of the sample, possibly only a single sample member, but which implies that the information generated by the validation is valueless for estimation purposes. The point is that such incentives are much cheaper than full-scale validation, and can therefore be routinely included in surveys. They alter the supply of errors in a way that allows for identification of population parameters among a large group that receives the incentive. Full-scale validation is almost always prohibitively costly and therefore rarely used. Therefore, the point of a validation incentive is to limit monitoring.

2.1 Incentive-Estimators for Continuous Variables

In the case when Y and Y^* are continuous, denote by $s(z)$ the reservation wage distribution and hence $s(w)$ the fraction supplying the truth under a wage w of the validation incentive. Note that this supply curve concerns supplying the truth conditional on participating as opposed to supplying observations *per se* as incorporated later. Without loss of generality, one can write the outcome as the sum of the truth plus the error

$$Y = Y^* + \epsilon(Y^*)$$

[7] Becker (1969) discusses this value of limited monitoring in the context of punishment for criminal behavior. Our arguments generalize to when the monitoring is imperfect although the tradeoff between w and π that makes agents indifferent may change.

where $\epsilon(Y^*)$ is the measurement error, possibly dependent on the true type. The mean of supplied outcomes under a given level of compensation is

$$E[Y|w] = E[Y^*] + (1 - s(w))E[\epsilon(Y^*)]$$

This is the true mean corrected for the fraction not selling the truth and how big the error is for them.

In the case of estimating the *level* of a continuous outcome variable Y, one standard assumption on measurement errors used almost universally in econometrics is that they have zero means, that is, that they are unsystematic. If measurement errors are claimed to vanish in the aggregate, this amounts to a behavioral assumption on uncompensated supply of the truth

$$E[Y|w_0] = E[Y^*] \Rightarrow (1 - s(w_0))E[\epsilon(Y^*)] = 0$$

where w_0 denotes the lack of wage compensation. This says that either everyone sells the truth or that some do not but for those erroneous supply cancels out. The assumption of zero measurement error means is untestable in standard practice but using validation incentives allows us to test it. Consider when a single incentive w is randomized out, so that some randomly get it and some remain uncompensated at a zero wage. This random price assignment neutralizes both population parameters and supply schedules across the two groups. The difference in supplied outcomes is thereby

$$E[Y|w] - E[Y|w_0] = [s(w_0) - s(w)]E[\epsilon(Y^*)]$$

If the standard assumption of a zero uncompensated mean holds, then this implies that the effect of the wage incentive on the mean outcome must be zero. If the uncompensated supply is fully truthful, $s(w_0) = 1$, then the first factor on the right hand-side is zero because increased compensation cannot improve upon perfect performance. If the uncompensated error when not supplying truthfully cancels out, $E[\epsilon(Y^*)] = 0$, then the second factor on the right hand side is zero because better performance has a zero marginal effect. Therefore, if uncompensated measurement errors vanish in the aggregate, the effect of the wage on the mean outcome must be zero

$$E[Y|w_0] = E[Y^*] \Rightarrow E[Y|w] - E[Y|w_0] = 0$$

Moreover, if supply is strictly elastic then the sign of the first factor is known, so that the *sign* and not only the existence of the error can be identified.

This occurs because of the monotonic convergence of the measured mean to the true mean as compensation rises; the effect of the wage is negative (positive) whenever the observed outcome is above (below) the true one; $E[Y|w_0] > E[Y^*]$ if and only if $E[Y|w] < E[Y|w_0]$.

In estimating the *effects*, as opposed to levels, of a covariate on the outcome consider the effect made up of the difference in outcomes Y_A and Y_B for two covariate groups, A and B. A second common assumption on the supply of measurement errors is that they are independently distributed with covariates. Here this amounts to the behavioral assumption on uncompensated supply

$$\frac{[1-s(w_0)]E[\epsilon(Y_A^*)]}{[1-s(w_0)]E[\epsilon(Y_B^*)]} = \frac{E[\epsilon(Y_A^*)]}{E[\epsilon(Y_B^*)]} = 1$$

Before, the zero-mean assumption implied that the *level* of the true mean was identified from the level of the supplied mean. Here, this assumption ensures that the *difference* in true means is identified by the difference in supplied means

$$E[Y_A|w_0] - E[Y_B|w_0] = E[Y_A^*|w_0] - E[Y_B^*|w_0]$$

In estimating effect, this often motivates the differencing of aggregate sample means because measurement errors are claimed to cancel out even if they were not equal to zero, as under the first assumption. Using incentives, this second assumption of independence between measurement error and covariates can also be tested. Consider randoming out two wage levels, w_0 and w, repeating the type of incentive above on each covariate group. The effects of the wage on the mean outcomes for the two covariate groups are then

$$\Delta_A \equiv E[Y_A|w] - E[Y_A|w_0] = [s(w_0) - s(w)]E[\epsilon(Y_A^*)]$$

$$\Delta_B \equiv E[Y_B|w] - E[Y_B|w_0] = [s(w_0) - s(w)]E[\epsilon(Y_B^*)]$$

If the mean for a given covariate group and compensation level is estimated by its sample mean , denoted $\bar{Y}(w)$, then the probability limit of the ratio of the incentive effects is

$$plim \frac{\bar{Y}_A(w) - \bar{Y}_A(w_0)}{\bar{Y}_B(w) - \bar{Y}_B(w_0)} = \frac{\Delta_A}{\Delta_B} = \frac{E[\epsilon(Y_A^*)]}{E[\epsilon(Y_B^*)]}$$

This allows us to identify whether this ratio is unity or not as corresponds to the independence assumption. As was the case when testing the zero mean assumption, the *sign* of the bias of the differences of the supplied outcome means is identified

$$E[\bar{Y}_A(w_0) - \bar{Y}_B(w_0)] \geq E[Y_A^*] - E[Y_B^*] \Leftrightarrow \frac{\Delta_A}{\Delta_B} \geq 1$$

Further identifying assumptions naturally means that more can be learned about the effect through incentives and in the next section we consider point-identification of parameters through incentives.

2.2 Incentive-Estimators for Binary Variables

This section considers the case of supplying the outcomes of a binary variable such as one indicating unemployment, uninsurance, or educational status. In particular, the section shows that in this case, incentives often allow for point-identification of parameters of interest, as opposed to only determining their signs. In the binary case, Y and Y^* equal 0 or 1. Let $s_0(z)$ and $s_1(z)$ denote the reservation wage distribution of the two types which yields the increasing supply functions $s_1(w)$ and $s_0(w)$ for a given wage. Let $\theta \equiv P(Y^* = 1)$ denote the true fraction of the population that has the condition. If this fraction is estimated by the sample mean of the supplied outcomes, this estimator has a mean $m(w)$ determined by those who have the condition who truthfully report having it and the proportion of those without the condition who falsely report having the condition

$$m(w) \equiv \theta s_1(w) + (1 - \theta)(1 - s_0(w))$$

To study the use of incentives to estimate the *levels* of a binary variable, consider the parametric supply functions that satisfy the linear index specifications

$$s_0(w) = G(\alpha_0 + \beta_0 w) \quad \& \quad s_1(w) = G(\alpha_1 + \beta_1 w)$$

where (α, β) are unknown parameters and G is a known index function. Since the parameters (α_0, α_1) are unknown then so are the levels of uncompensated supply. A severe consequence of this is that *nothing* can be learned about the true proportion, θ, from the uncompensated proportion, $m(w_0)$, without additional assumptions. More precisely, it is straightforward to show that for any supplied proportion $m(w_0)$, for all true proportions θ there exists a pair (α_0, α_1) that generates the supplied proportion from the true one.

As opposed to the continuous case, given the levels of uncompensated supply it is therefore not always feasible to have measurement errors vanish in the aggregate, that is, to be of mean zero for *all* true proportions θ. Indeed, the only case when this can occur is when supply is errorless as then $m(w_0) = \theta$ for all θ. Even more troublesome is that the *sign* of the bias cannot be determined without knowing the true proportion θ. Indeed, it can be shown that whenever there is some erroneous supply, the supplied proportion has downward bias for large values of the true proportion and upward bias for small values of the true proportion. More precisely, if $s_1(w_0), s_0(w_0) \in (0,1)$ then there exists a cut-off value θ^* of the true proportion above which there is under-reporting and below which there is over- reporting

$$m(w_0) < \theta \Leftrightarrow \theta > \theta^*$$

This occurs because erroneous supply involves under-reporting by those who have the condition, and so when that fraction is high, under-reporting dominates. Similarly, erroneous supply of those who do not have the condition involves over-reporting so that when they dominate in size such over-reporting occurs. This inability to sign the bias is naturally troublesome when attempting to estimate the true proportion from the supplied proportion of the sample.

Using randomized incentives allows one to overcome this identification problem through separating out poor supply from the true outcome distribution. In particular, in the Appendix it is shown that there exists a set of randomized wages Υ for which θ is identified from the supplied outcome means

$$\{m(w); w \in \Upsilon\}$$

The intuition of this is as before; the incentives are randomized out across sub- populations with the same θ enabling one to identify the supply schedules after which the true proportion can be backed out again.

To investigate the use of random incentives to estimate binary *effects*, as opposed to levels, consider comparing self-reported data for two generic covariate groups A and B. For this comparison there exists a simple incentive scheme that allows point-identification of their differential effect. This scheme randomly assigns on each covariate group two wages, a smaller one $\underline{w}_1$ and a larger one $\bar{w}_1$, differentially awarding *only* the type that has the condition for supplying the truth. As before, the important aspect of randomization of the wages across sample members is that both the supply schedules as well as other determinants of outcomes are neutralized across groups with different wages. If $\bar{Y}_A(w)$ and $\bar{Y}_B(w)$ denote the sample averages of supplied outcomes at a given wage for the two covariate groups, the effects of the wage on the mean outcome *within* each group are

$$E[\bar{Y}_A(\underline{w}_1) - \bar{Y}_A(\bar{w}_1)] = \theta_A[s_1(\underline{w}_1) - s_1(\bar{w}_1)]$$

$$E[\bar{Y}_B(\underline{w}_1) - \bar{Y}_B(\bar{w}_1)] = \theta_B[s_1(\underline{w}_1) - s_1(\bar{w}_1)]$$

Therefore, the relative difference *across* the two covariate groups has the probability limit

$$plim \frac{\bar{Y}_A(\underline{w}_1) - \bar{Y}_A(\bar{w}_1)}{\bar{Y}_B(\underline{w}_1) - \bar{Y}_B(\bar{w}_1)} = \frac{\theta_A}{\theta_B}$$

In other words, the differential effect of the incentive across the two covariate groups represents the relative fraction of the groups having the condition in A and B. The intuition behind this is that those having the condition respond to the differential incentive in the same way across the two treatments so that only the relative size of the groups responding, i.e. the relative effect, affects the ratio. Note that this type of incentive scheme identifies the relative effect θ_A/θ_B for any non- parametric form of the supply functions s_0 and s_1.

Many times the difference in outcomes between the two covariate groups reflects other things than the covariate values themselves, such as time trends when doing pre- and post comparisons. Therefore, one estimator that is often considered is the differences in the pre-post effects for two groups, so called difference-in-difference estimators. If sample members produce outcomes in the two groups A and B at two times t and $t+1$, randomizing incentives in a parallel fashion on the two groups allows for unbiased estimation of this difference-in-difference effect. The incentive scheme that randomly assigns wages $\underline{w}_1$ and $\bar{w}_1$ to group A at both times will allow for asymptotically unbiased estimation of $\theta_A^{t+1}/\theta_A^t$ in a similar manner as discussed before. Likewise, randomly assigning wages[8] $\underline{w}_1$ and $\bar{w}_1$ on group B in both periods allows for estimation of $\theta_B^{t+1}/\theta_B^t$. The probability limit of the difference of the two is

[8]Note that these wages do not necessarily have to be the same on group A as on B as long as they are the same across the periods t and $t+1$.

therefore the desired effect

$$plim[\frac{\bar{Y}_A^{t+1}(\underline{w}_1) - \bar{Y}_A^{t+1}(\bar{w}_1)}{\bar{Y}_A^t(\underline{w}_1) - \bar{Y}_A^t(\bar{w}_1)} - \frac{\bar{Y}_B^{t+1}(\underline{w}_1) - \bar{Y}_B^{t+1}(\bar{w}_1)}{\bar{Y}_B^t(\underline{w}_1) - \bar{Y}_B^t(\bar{w}_1)}] == \frac{\theta_A^{t+1}}{\theta_A^t} - \frac{\theta_B^{t+1}}{\theta_B^t}$$

It says that the relative effect of the wage incentive across periods within group A compared to the same relative effect for group B identified the parameter of interest. The point is that the fraction of suppliers who have the condition drives the effects of wages on mean outcomes and therefore the relative fraction across time for two groups can be identified in the same ways as before.

3 Multiple Quality Margins and Random Incentives

This section expands on the previous one by considering multiple margins of supply so that now both measurement errors and non-participation may occur. Let the capital letter wage W denote the wage for participation (the extensive margin) and let w denote the validation incentive for errorless supply (the intensive margin). Given a contract (W, w) for quality on both margins, let the capital letter $S(W, w)$ denote the fraction of the sample participating and the lowercase letter $s(W, w)$ the fraction of truthful supply as discussed before. Denote by U the non-produced and unobservable[9] factors affecting the outcome vector, that is, factors that outcomes are not produced for and assume that the supplied mean depends on the two margins of supply according to

$$E[Y|W, w] = f(S(W, w), s(W, w)) + E[U|W, w]$$

As before, the key point in estimating the production bias introduced by participation is that the randomized wages allow for something that effects participation but not other non-produced outcome determinants. That is, the wages serve as an instrumental variable which is explicitly, as opposed to implicitly, randomized.

[9] We avoid discussing observable determinants because it simply adds notation without affecting any of our conclusions.

Consider randomizing the levels of the participation wage across the sample. The impact of such participation incentives depends on how the two margins interact. When there are no cross-elasticities between the two margins, $dS/dw = ds/dW = 0$, the effect on the conditional mean function of raising participation incentives is

$$\frac{dE[Y|W,w]}{dW} = f_S \frac{dS}{dW} + \frac{dE[U|W,w]}{dW}$$

The random wage implies that the second term is zero. As before when contracts are randomized out across suppliers then they are independently distributed with unobservable outcome determinants; $E[U|W,w]$ does not vary with assigned wages. This implies that the production bias is identified through unitizing the total outcome effect by how much participation affects outcomes

$$f_S = \frac{dE[Y|W,w]}{dW} / \frac{dS}{dW}$$

This is simply the population conditions justifying the instrumental variable estimator of non-response bias made up of the participation incentive; a variable that drives participation but does not affect supply schedules on either margin nor unobserved outcomes.

Now consider the case when there are cross-elasticities across the two margins which implies that it is more difficult to separate out participation effects from other, unobservable, effects occurring through measurement errors. The tradeoff between monitored and non-monitored activities, in particular the reduction in unobservable quality to increased incentives on observables, has been stressed repeatedly in other agency problems.[10] Given that the effects of contracts on unobservable factors are still neutralized, $dE[U|W,w]/dW = 0$, the total effect of a participation incentive is now

$$\frac{dE[Y|W,w]}{dW} = f_S \frac{dS}{dW} + f_s \frac{ds}{dW}$$

This simply says that the outcome effects of raising the incentive to participate is determined by the direct effect through increased participation together with the indirect effect due to any cross-elasticity with the intensive margin. The substitutability or complementarity between the two margins,

[10] See e.g. Lazear (1986), Milgrom and Holmstrom (1991), and Baker (1992)).

i.e. the sign of ds/dW, is important because it implies that even though the participation rate may rise with compensation, measurement errors may be larger for those that were made to produce observations.

If one wants to identify the impact of participation incentives on measurement errors then monitoring must take place. If the substitution matrix is *constant across items*, that is, the elasticity of measurement errors with respect to the participation incentive, ds/dW, is constant across all items on the survey, then there is the ability to cheaply identify this elasticity through the inclusion of *frame-items*. These items are here defined as questions which one knows the answers to through their inclusion on the survey frame. The purpose of having sample members supply these frame-items although they are already known would be to estimate the response in measurement errors to incentives to participate, ds/dW, by being able to monitor the measurement error without any cost. If one knows the answer to the question asked, through asking sample members about items on the frame, then the measurement error can be produced directly without any costly validation undertaken by the survey producer. In other words, the measurement error effect ds/dW can be identified and separated out from the effects of participation dS/dW in the total effects of participation incentives on observed outcomes. In particular, one can identify whether cross-elasticities are zero so that the simple correction above can be performed.

3.1 In-Kind Compensation: The Case of Confidentiality

The random variation in participation incentives does not have to be pecuniary but may be in-kind. This section considers how the discussion generalizes to the in-kind benefit of confidentiality, which is a major concern in producing one of the most important types of economic data; income.

The preference for confidentiality by sample members may be interpreted as preferring to report their outcome as an interval rather than as a point. Interval reports, or 'bracketing', has been used extensively in practice.[11] Interval reports enable the sample member to avoid fully revealing his income in place of doubly censoring it. Consider when sample members are char-

[11]See e.g. Manski (1995) and Dominitz and Manski (1996), and Juster and Suzman (1994).

acterized by the pair (Y^*, C). As before, Y^* represents true outcome, here assumed to have a finite support normalized without loss of generality to the unit interval $[0,1]$. The variable C is the reservation level of confidentiality interpreted as the smallest sized interval the sample member agrees to be identified with. The subject is assumed to not supply an observation if the confidentiality is smaller than C and supply it if it is larger (this abstracts from erroneous supply although the discussion could be generalized to cover this margin as well). The subject population is characterized by a distribution $S(c)$ of reservation values of confidentiality with the special case of $S(0) = 1$ representing the standard case of no preference for confidential supply. A survey is said to provide confidentiality of size W if it allows sample members to supply which bracket among a set of $1/W$ equally sized brackets their true outcome lies. The distribution of reservation levels of confidentiality induces a participation rate $S(W)$ for a survey with confidentiality W. The in-kind benefit of confidentiality therefore acts like the pecuniary wage discussed before; the supply of observations rise with the confidentiality of the survey $dS/dW \geq 0$.

The important aspect of such a confidential survey is that even though the sample members prefer not to be point-identified on the individual level, the distribution characterizing them can be point-identified on the aggregate level. Consider when the outcome distribution is characterized by $G(y^*|\alpha)$ where α is an M-dimensional parameter to be estimated. It is argued that these parameters can be estimated consistently using a survey with confidentiality of at least $W \geq 1/M$. This estimation is possible because even though individual outcomes are not point-identified in a confidential survey, the cumulative outcome distribution G is. More generally, for any partition represented by the bracket cut-offs $[y_1, .., y_M]$ the corresponding values

$$(G(y_1|\alpha), ..., G(y_M|\alpha))$$

of the cumulative income distribution can be consistently estimated by computing the values of the empirical distribution function $(\hat{G}_1, ..., \hat{G}_M)$ where $\hat{G}_m$ is the sample fraction supplying outcomes in the lowest m number of intervals, $[y_1, y_m]$. It follows immediately that if there are a total of M intervals then an income distribution with up to M parameters is identified. This is true when the in-kind benefit of confidentiality is low enough to generate the required number of intervals; $W \geq 1/M$.

Just as was the case for pecuniary incentives, participation rises with confidentiality, $dS/dW \geq 0$, but it may affect the outcomes produced, $df/dS \neq 0$, if those preferring confidentiality more have different outcomes. Therefore, assessing the impact of confidentiality is completely analogous to assessing participation bias through randomization of pecuniary incentives. If α_W is a parameter identified under confidentiality W, the outcome elasticity of the outcome mean $E[Y|W] \equiv \int y dG(y|\alpha_W)$ can be used in a similar manner to estimate the ultimate parameter of interest; the mean when full confidentiality was guaranteed

$$E[Y|W=1] = E[Y^*]$$

For example, this would be the mean income among those willing to reveal they have positive income, which presumably would be everyone. Unless there is a level of confidentiality at which there is a full supply of observations, this will involve extrapolation. If each sample member supplies the entire support of the distribution, nothing is identified except the support itself.

4 The Quantity-Quality Tradeoff and Optimal Design

The previous discussion showed how randomized incentives may be used to identify and correct for production bias when survey supply is elastic on both quality margins. This section discusses how the elasticity of survey supply also determines the optimal choice of quantity vs. quality, that is, how large a labor force is required to make up the sample size relative to how much to award performance of it.

Consider the quantity-quality tradeoff between sample size and participation, although the issues generalize to other forms of quality. For a survey with sample size N, let $U(N, S)$ denote the utility function of the investigator which is assumed to be increasing in both arguments so that larger surveys with higher participation rates are preferred. Perhaps the most common type of *revealed* preferences in practice are those that do not distinguish between sample size and participation, being indifferent as long as the number of observations produced to be analyzed is the same; $U(N, S) = NS$. It seems that most applied econometrics is carried out under such revealed preferences; attention is only paid to the size of the available sample to be consumed as opposed to the sample drawn from the frame.[12]

The expenditures of the survey are made up of fixed costs F, such as preparing the survey (e.g. expenditures on the survey frame from which the sample is selected), and variable sampling costs C incurred for each sample member regardless of whether they participate (e.g. search costs). Consider unconditional and conditional compensation for quality defined as being paid regardless of participating or not versus conditional on participation only. Although few economic theories predict that unconditional pay will be optimal, our evidence turns out to lend support to the idea that it may. Let $W = (W_u, W_c)$ be the unconditional and conditional wage and let $S(W)$ be the participation rate given the levels of both those wages. As we will allow people to supply without requiring a wage, let (δ_c, δ_u) be the fraction requiring pay for supplying observations, under conditional respective to unconditional pay. These fractions are not necessarily unity and are important because unconditional pay may be taken advantage of by sample members. The survey designed and analyzed in the empirical section allowed these quantities to depart from unity which is the quid-pro-quo value that is obtained when all sample members get paid for supplying. The budget constraint of the survey is then

$$F + N(C + S(W)[W_c\delta_c + W_u\delta_u]) = B$$

where B is the total budget. The wage expenditures are made up of the conditional pay as well as the unconditional pay corrected for any free supply, with the standard case of $\delta_u = \delta_c = 1$ prevailing if each sample member that supplies requires the wage. The sample size is simply the maximum one that can be offered once the wages have been chosen

$$N(W) = \frac{B - F}{V(W)}$$

where $V(W) \equiv A + S(W)[W_c\delta_c + W_u\delta_u]$ is defined as the total variable cost per observation. Since the sample size is decreasing in both wages there is a negative tradeoff between quality and quantity determined by how elastic

[12]Another example of how quality and quantity enters into the production is through the mean-squared error (MSE) of a single outcome in which quantity affects the part of the MSE made up of the variance and the response rate part made up of bias. However, in this case the parameter to be estimated affects the objective function.

survey supply is; the foregone sample size that needs to be sacrificed for a one percent increase in the participation rate is larger the less elastic survey supply is.

The optimal quantity and quality is the one that maximizes $U(N, S(W))$ over the feasible combinations defined through this budget set. Substituting in the induced sample size into the necessary first order conditions one gets

$$\frac{dU}{dN}\frac{dN}{dW_c} + \frac{dU}{dS}\frac{dS}{dW_c} = 0$$

$$\frac{dU}{dN}\frac{dN}{dW_u} + \frac{dU}{dS}\frac{dS}{dW_u} = 0$$

An interior solution $W = (W_c, W_u)$ satisfies these two conditions directly implies that the foregone sample sizes for the expenditures devoted to the two methods of increasing participation must be equalized

$$\frac{dN}{dW_c}/\frac{dS}{dW_c} = \frac{dN}{dW_u}/\frac{dS}{dW_u}$$

The expressions for the foregone sample sizes are

$$\frac{dN}{dW_c} = \frac{B-F}{V^2}[\frac{dS}{dW_c}(W_c\delta_c + W_u\delta_u) + S\delta_c]$$

$$\frac{dN}{dW_u} = \frac{B-F}{V^2}[\frac{dS}{dW_u}(W_c\delta_c + W_u\delta_u) + S\delta_u]$$

Substituting these in and rearranging one gets

$$\frac{\frac{dS}{dW_c}(W_c\delta_c + W_u\delta_u) + S\delta_c}{\frac{dS}{dW_c}} = \frac{\frac{dS}{dW_u}(W_c\delta_c + W_u\delta_u) + S\delta_c}{\frac{dS}{dW_u}}$$

Multiplying both sides by the product $W_c W_u$ and rearranging one gets

$$\frac{W_c}{W_u} = \frac{\eta_c}{\eta_u}\frac{\delta_c}{\delta_u}$$

where η_c and η_u denote elasticities of conditional and unconditional supply. This says that relative wages relate to relative elasticities in a standard way, corrected for the difference in how much pay is required per supplied observation. The correction is important since unconditional pay may take place without supply and conditional pay may not require pay. It is therefore not only relevant how elastic the supply is to the two methods of pay but also how close to quid-pro-quo the two methods are. The standard case of completely elasticity determined relative wages arises when supply is quid-per-quo even though it does not have to be.

5 Empirical Analysis

This section provides empirical evidence on survey supply using data from an experiment we produced to estimate the elasticities of survey supply with respect to the incentives discussed and the effects of the participation incentives discussed in section 3 on estimates of mean levels of income among physicians.

5.1 The Survey Supply Experiment

The supply effects of survey incentives were investigated in a module called the *Survey Supply Experiment* of the physician survey *The Index of Hospital Quality.* The main survey was produced by The National Opinion Research Center (NORC) at The University of Chicago during October - December 1995. The main objective of the survey was to obtain quality rankings of hospitals by physicians. The experiment was added to the main survey on a subsample of the survey.[13] The survey was privately financed by the magazine *US News & World Report.* The private funding is important for the types of incentives used to identify production bias because both the levels and variation in survey wages are regulated in publicly financed surveys in the US through The Office and Management and Budget.

The sample was a random sample from the US population of physicians and consisted of 2550 physicians with 150 individuals from each of 17 specialties. The survey was administered by mail and quite short in length, it contained a total of 34 items and took about 5 minutes to complete. The survey frame of the sample was the physician directory of the American Medical

[13]See Philipson and Grabowski (1996) for a more complete description of the experiment. The survey budget without the experiment was $ 170,000 with an average cost per sample member of $ 170.000/2,550= $ 67 (throughout the paper 1996 dollars are used).

Association (AMA).[14] This frame is typical in that it contains information on the frame members prior to sampling that can be used to test for the substitutability between quality margins as discussed earlier. The allocation of incentives in the experiment was through a randomized block design. There were a larger set of specialties in the survey than those assigned incentives. However, within those six specialties used to assign the incentives, each incentive was assigned in equal proportions so that each specialty had all the four treatment groups discussed here.[15] More precisely, each of the 3 incentives was allocated to 120 individuals, 20 from each of the six specialties, with a larger control group from these same six specialties.[16]

The first treatment assigned was an unconditional participation incentive in which the sample members received $W_u = 50$ dollars with the questionnaire which could be cashed regardless of whether the questionnaire was returned or not. The second treatment assigned was a conditional participation incentive in which the sample members received $W_c = 50$ dollars if the questionnaire was returned. The third treatment was a validation incentive for which $\pi = 10$ percent of the group got monitored and received $w = 500$ dollars if they had no measurement error ($y = y^*$). The last treatment group was a control group made up of a balanced set of sample members within the same specialties as those in the three treatment groups. For each treatment group, the sample members received a different set of instructions on how their behavior would translate into pay. Each sample member was instructed about his particular incentives through an added form explaining the conditions under which pay would take place, including the absence of conditions when unconditional payment was used (Appendix 1 contains one such form for illustration). The sample members were not informed about the fact that they were participating in an experiment nor that their incentives may differ from others.

[14] This frame, the primary frame used for physician surveys in the US, had been produced by NORC for previous physician surveys and includes members as well as non-members of the AMA.

[15] As selected by NORC, the six specialties were the ones with lowest participation rates in the previous years of the survey; Neurology, Urology, Cancer, Cardiology, Ophthalmology, Orthopedics.

[16] The sample sizes were chosen for a power of 10 percent to detect 10 percent effects under a one-sided alternative (positive supply elasticities) given historical levels of uncompensated supply.

The variables measured in the experiment are described in Table 1 below in terms of their overall means for the entire sample.

The table contains mainly three sets of variables. The first is the four treatment groups considered. The second set of variables is a set of demographic outcomes such as gender and region of residence.[17] The third set is a set of supply variables that includes survey participation (S), item-response, and measurement errors (s). The most important item of the survey is the variable *HospSupp* indicating the number of hospitals ranked on quality by the sample member as this was the main purpose of the survey and it had involved substantial item-non response in the past . The variable *NHospSupp* is the number of items supplied on the rest of the survey excluding the hospital item. The measurement errors were observed for two frame-items represented by the variables *MtchGrad* and *MtchCode*, which are dummies indicating that supplied outcomes matched the frame ($y = y^*$). These items concerned the exact date of graduation from medical school as well as the *number* representing the AMA Code for the physician=s medical specialty.[18] Since these items were part of the frame, we could observe measurement errors for the entire sample and not only those validated.[19] Due to the limitations of this particular survey, income could not be monitored so that validation incentives could not be used on this measure. Furthermore, the choice of the frame variables for which the validation incentives were tested were obviously made not because of their substantive importance but because of their availability in this particular survey.

[17]The randomization of incentives implied that when we tested for differences across treatments we could not reject equality in the frame variables, e.g. age and gender.

[18]The exact wording of the two questions were (italic not added); 1.What is the *exact date* when you received your medical degree ? 2. What is the (3 digit) AMA code of your primary specialty ?

[19]No differences in compensation were made for different true types, mainly because types were not of substantive interest in themselves in this particular experiment.

5.2 Measurement Error Elasticities and Validation Incentives

This section provides estimates of how elastic measurement errors are with respect to validation incentives. Table 2A-B below brakes up these error effects by sub-groups to investigate which groups are relatively more elastic and shows that there is substantial difference between who responds as well as the type of question validated.

Each row corresponds to a given subgroup, part of which received the validation incentive and part which were in the control group. The four columns correspond to the fraction of observations produced without error among those on the validation incentive, the controls, their difference, and the p-value at which they are significantly different. The tables indicate that measurement errors are highly elastic for the specialty code item but inelastic for the graduation date item. Note that even though the levels of correct supply are relatively low for the specialty code, the elasticities are large. This is important since it is elasticities, and not the levels, that allow for the identification through our discussed incentive schemes.

Table 3 reports the regression results estimating more efficiently the error elasticities for the specialty code item[20] and allowing for tests of the conditional independence assumption. The table shows that for successively larger specifications taking into account differences in other determinants of errorless supply where the randomization may not perfectly balance, the effect of the incentive remains fairly robust and highly significant. Indeed, this effect is the most significant determinant of errorless production throughout the specifications considered.

The evidence above suggests that among a group of individuals who a priori would seem unlikely to respond to survey incentives, a rich subset of specialties of US physicians, there is substantial evidence that modest

[20]We found no significant effects of validation incentives for the graduation date item. The corresponding table is available on request.

validation incentives can have strong effects on measurement errors. An open question concerns why the responses differ across the two items as it is important to have positive error elasticities to utilize our identification strategies.

When measurement errors are elastic, tests like the ones above may be useful to assess their patterns to subsequently identify population parameters aside from errors. As discussed, the standard behavioral assumptions on production process is that it cancels out in the aggregate (mean zero) and cannot be systematically explained by a set of observable determinants (independently distributed with covariates). A special case of the second behavioral assumption is that error-less production is wage-inelastic. In the table, we cannot assess the first assumption because the questions themselves do not lead to variable values that cancel out. However, were one to regress such data in the format of the table, the prediction would correspond to a zero value of the intercept. However, the second prediction can be tested because it corresponds to all slope coefficients being zero. It seems to be rejected, in particular by the fact that production of errorless observations is wage-elastic. Of particular interest is that the elasticities are present in a very rich group with median incomes around 200 thousand which suggests that groups with lower opportunity costs of time may respond in similar ways.

5.3 Multiple Quality Margins and Participation Incentives

This section discusses the evidence concerning the impact of participation incentives on both quality margins of supply (S and s). The two main findings reported are surprising and interesting; we find that participation responds more to *unconditional*, as opposed to conditional, pay and that the two margins of quality are found to be gross *complements* rather than substitutes.

The summary statistics by the two participation treatments and the control group are displayed in Table 4 below which shows how the supply behavior varies across the three groups, that is the patterns of $(S, s)(W_u, W_c)$.

The columns of the table represent unconditional pay, conditional pay, as well as the control group; a balanced subset of the six specialties that were treated. The table displays the unconditional evidence (which is more informative than unconditional observational data due to the random assignment of treatments). The last column indicates the results of an F-test testing for homogeneous means across the three treatment groups (columns) for a given variable (row). The table indicates that the randomization of the incentives induced for the non-behavioral variables are not affected by the incentives, there are no significant differences due to the randomization of the incentives. However, there does exist significant differences in survey supply behavior across the treatments. The first is the elasticities with respect to the two types of incentives relative to the controls, that is, dS/dW_c and dS/dW_u. The table shows that supply is much more elastic with respect to unconditional pay than to conditional pay. This is interesting because economic theory predicts that the unconditional elasticity should be zero.[21]

The second aspect of the table to note is the multi-task tradeoffs it displays. More precisely, the direct effect of incentives on observable participation are coupled with their indirect effects on unobservable activity such as the intensive margin, ds/dW_c and ds/dW_u. Here, the intensive margin is represented by both measurement errors and item-response. Interestingly, using the frame-items to estimate cross-elasticities, the table reveals that unconditionally, the two margins are gross complements or there is no cross-substitution (as opposed to substitutes as generally argued in multi-task agency). Both the number of hospitals supplied, measured by *HospSupp*, as well as the fraction of errorless outcomes in *MtchCode* and *MtchGrad*, are significantly higher or not significantly different among those under a participation incentive relative to controls. The fact that the measurement error effect is produced virtually without cost, through frame-validation as discussed above, suggests that testing for the cross-elasticities in this simple manner may be a useful method to assess whether they are zero so that the total production bias is simply due to the participation margin.

The third aspect of the table to note is that the degree of incorrect compensation is extremely low as witnessed by the estimates of the paid sample

[21] Traditional labor-supply models may predict positive levels of supply under unconditional incentives but it is less straight forward to predict positive *elasticities* to such pay.

members per supplied unit-response, δ_u and δ_c, being close to unity. In other words, the estimated fraction of sample members paid per supplying member is close to one. For sample members receiving unconditional pay the experiment measured whether the included checks were cashed. The variable δ_u is therefore the number of those cashing the checks per questionnaire received, regardless of whom supplied the questionnaire or cashed the cheque. For all sample members receiving conditional pay in which payment was sent after they returned the questionnaire, the experiment allowed them to choose to not receive a payment even though they supplied the questionnaire. The variable δ_c is therefore the number of those requesting pay per questionnaire supplied. This evidence suggests that survey producers pay for what they get in the sense that few unconditionally compensated take advantage of it and few conditionally compensated do it for free. Our discussion concerning optimal compensation shows that relative wages were determined by both supply elasticities and these variables as in

$$\frac{W_c}{W_u} = \frac{\eta_c}{\eta_u}\frac{\delta_c}{\delta_u}$$

Therefore, this finding suggests that in contrast to what is suggested by standard labor supply models, unconditional pay should be used over conditional pay since supply is more elastic relative to the former and pay is close to equalized when pay is quid-pro-quo.[22]

5.4 Production Bias in Mean Earnings Estimates

The previous section suggested an absence of cross-substitution between participation and measurement errors as evidenced by our test for zero cross-elasticities using frame items that are monitored without cost. Therefore, correcting outcomes using the participation margin alone can be performed when such cross-substitution is absent. This section shows how this can be done when estimating the mean earnings of these physicians. Note that this does not involve the use of validation incentives on earnings as this was not available through this particular survey. Rather, the participation incentive

[22]Naturally, understanding why unconditional pay is elastic in spot markets represented by cross-sections is a useful question for future research. In long term contracts such as panel surveys, current supply may be elastic to unconditional incentives due to future benefits.

of the experiment is here used to correct the production bias of estimated mean earnings and it is found that they rise with about one third, or 90 thousand dollars.

The earnings data of the experiment was in the confidential bracketed form which is the standard of NORC and many other survey producers producing earnings data.[23] Table 3 below displays the produced income data by the two treatment groups of participation pay (whether conditional or unconditional) and the control group with no incentive.

The table displays the estimated levels of the distribution function across the confidential income intervals with cut-offs (in thousands) $(y_1, y_2, .., y_M) = (50, 100, 150, 175, 200)$. The table shows that the income distribution for those with the participation incentive stochastically dominates that of the controls. It also reports the mean incomes across the two treatments which differ in 16.5 thousand dollars. These mean values were estimated as discussed in the section of estimating confidentiality effects on estimates.[24] For all, we used the GMM-estimator $\hat{\alpha}_W$, defined as the parameter that made the empirical distribution function come closest to the theoretical one

$$\min_{\alpha} \sum_{m=1}^{M} [\hat{G}_m - G(y_m|\alpha, W)]^2$$

where the parametric distribution function used was of the form

$$G(y|\alpha) = 1 - \frac{1}{(1 + \alpha_1 x^{\alpha_2})^{\alpha_3}}$$

This distribution has been argued to fit the US income distribution better than the lognormal or Pareto distribution which fit poorly at high respective to low levels of the distributions (see Singh and Maddala (1976)).

[23]Most studies estimating physician earnings involve such bracketing, see e.g. Schowalter and Thurston (1996) who uses the *Physician Practice Cost and Income Survey* (PPCIS) produced by NORC using the same AMA Master File as used here. Also, see Burstein and Cromwell (1985) who uses AMA *Profile of Medical Practice* 1969-80.

[24]However, the survey did not allow randomization of different confidentiality levels to carry out such estimates here.

To illustrate the production biases in as simple manner as possible, we use a linear conditional mean function , f, describing the relationship between mean outcomes and the participation rate

$$E[Y|W] = \beta_0 + \beta_1 S(W) + E[U|W]$$

As discussed previously, the effect of the incentives on the outcomes is through the direct effect of incentives on survey supply coupled up with the indirect effect of survey supply on produced outcomes with unobservable effects canceling due to the randomization. For our linear illustration, the overall effect of the wage incentive is therefore

$$\frac{dE[Y|W]}{dW} = \beta_1 \frac{dS}{dW}$$

From the table we see that the effect of the participation rate on the mean earnings is given by

$$\beta_1 = \frac{dE[Y|W]/dW}{dS/dW} = \frac{206.4 - 189.9}{59.3 - 50.3} = 1.8$$

Once this effect of supply behavior on outcomes is estimated, the unconditional mean is[25]

$$E[Y] = \beta_0 + \beta_1 = E[Y|W_0] + \beta_1[1 - S(W_0)] = 189.9 + 1.8 \times [100 - 50.3] = 281.0$$

The production bias is then the difference between the true and estimated mean as in

$$E[Y] - E[Y|W_0] = \beta_1[1 - S(W_0)] = 91.1$$

This simple example of the production bias in earnings illustrates that for physicians or other income groups in which low participation rates are the norm, drastically different levels of income would be estimated with or without correcting for production bias. Indeed it may even be argued that this

[25] Naturally, these results rely heavily on the functional form of how the supply behavior affects the outcome measured. The specification discussed here is a simple version of Heckman (1979) with a randomly assigned covariate (the wage) which drives participation but not outcomes. See also Heckman and Robb (1986). However, the same general idea applies to the case when more than two different levels of the incentives are used so that non-linear production biases can be identified from randomized price schemes.

is an under-estimate of the production bias given the large earnings at the upper tail which may possibly make the participation rate earnings relationship convex as opposed to linear. Naturally, a larger number of survey wages would be necessary to estimate such a non-linear effect of production bias.

6 Concluding Discussion and General Remarks

This paper stressed the agency problem inherent in separating out poorly supplied data from parameters of economic interest to be estimated. The key point was that random wage discrimination alters the supply behavior of sample members across the same types of populations and thereby allows for separating out the effects of supply and population parameters. Empirical evidence for a survey of US physicians supported the notion that survey supply is affected by the types of wage discrimination schemes discussed in a manner that makes them useful for identification purposes. We conclude by discussing some general aspects of the approach taken as well as the future directions of inquiry it suggests.

Statistics in general may be interpreted as a normative theory of guessing the features of larger populations while only having access to a possibly imperfectly observed subset of them. This definition involves estimating levels or treatment effects, whether factual or counter-factual. However, statistics is more contingent on economics than one often realizes because if there were no economic constraints then statistics, in the sense defined above, would not be necessary; perfect data for the whole population would always be produced. However, economics seems to have not been fully utilized in statistics in that greater emphasis in systematic analysis has been paid by economists to the *consumption* of data, rather than on its production. This focus seems unfortunate, because the sampling errors routinely reported in journals often seem small relative to other production errors not due to finite survey supply. There is. of course, a whole research agenda, mainly in sociology and psychology, concerned with making survey instruments better and more user-friendly to increase both quantity and quality of the observations produced. However, in previous discussions not incorporating incentives, in many cases it even becomes infeasible to ask questions that are of interest because they are ruled out as unlikely to be producible by sample members that are not willing to undertake much time and effort to get things right. However, this *is* the principal-agent problem of survey production; sample members need user-friendly instruments only in absence of incentives. Few other employers than survey producers expect work from their employees at zero wages but somehow a free lunch is expected in the market for observations. Answers to better questions, as well as higher quality for existing questions, is only producible at wages that compensate the agents for internalizing the objectives of the demand side. A part of the principal-agent aspect of the problem, that the quality of produced observations is unobservable, also has an impact on survey *methodology* since it is hard to evaluate what is a good methodology when the quality of the product produced cannot be determined. Often this results in investigators establishing that a particular survey production technology, e.g. interviewing technique, has an *effect* on outcomes, rather than being able to determine which is the best technology. Without principals monitoring outcomes, it seems that such evaluations do not provide increased understanding of the preferred methods to use.[26]

Survey producers for publicly funded surveys in the US are affected by wage regulations set by federal agencies, such as the Office of Management and Budget. In particular, maximum wage policies are often invoked. These policies are many times justified, by economists as well as other survey producers, by the argument that compensating sample members would unjustifiably inflate already tight survey budgets. However, such restrictions on the feasible set of survey production inputs must obviously increase, rather than decrease, the cost of production. A more sensible argument against compensation may be that when compensation is used, income effects arise which imply that behavior is observed that would not be otherwise. This argument does not apply to most surveys since they are *retrospective*, and past behavior is presumably inelastic to unexpected current compensation. Such concerns may be relevant for panel surveys, however, if back-loaded compensation is used. Indeed, panels raise a large set of issues abstracted from here. They correspond to long-term contracts between the principal and agent as opposed to cross-sections which involve spot markets. Of particular interest is the fact that the production of panels may often be inconsistent with perfect recall on the part of sample members, a model of knowledge which dominates current economic models of dynamic choice. There is a tension therefore between models of the behavior of agents sampled and the survey design since if sample members behaved according to current dynamic models, few panels should be produced. The interaction between memory and compensation in

[26]This is particularly troublesome if the outcomes produced are subjective or attitudinal since it seems that no one has managed to operationalize the measurement of errors of such attitudinal data.

panel production are non-trivial, since with perfect recall the production of repeated measurement is not optimal relative to a single retrospective study asking sample members to recall long histories.

The discussion suggests several ways in which surveys used routinely by economists could account for the presence of production biases. An allowance for price-randomization in surveys and the reporting of compensation variables to users of the survey would enable the assessment of production biases to be addressed more fully in routine work. This paper suggested that part of the reason why the same level of sophistication that is brought to bear on data consumption by economists has not been brought to bear on production may be due to the fact that economists have not treated the market for observations as rational. Under such rational survey supply, this paper suggested that one may view the empirical analysis of economic behavior as being comprised of two parts: the first involves the rational motivations that generate the behavior sampled, and the second the rational motivations that generate observations on that behavior. For example, in producing income data, the income of a sample member may be adequately generated by economic models studied, although the observations obtained on that income fail to reflect it, due to an inadequate understanding of how rational sample members act within a particular survey. Given this rationality on the supply-side of the market for observations, this paper has stressed that it may be taken advantage of in survey production to better learn about the population of interest through random incentives that separates out the population sampled from supply behavior. The incentives discussed here can easily be replicated on larger scale surveys produced and routinely analyzed by including in produced data sets variables relating to the incentives used. In such a case, incorporating production errors into reported regression results would be as routine as the current practice of only reporting sampling errors due to a finite labor force. It seems that only tradition can justify the continuation of the routine detailed attention to consumption bias by economists, with virtually no serious attention to biases caused by production.

References

[1] Anderson, K., and R. Burkhauser, (1984), 'The Importance of The Measure of Health in Empirical Estimates of the Labor Supply of Older Men', *Economic Letters*, v 16, 375-380.

[2] Baker, G., (1992), 'Incentive Contracts and Performance Measurement=, *Journal of Political Economy*, v. 100, 598-614.

[3] Becker, G., (1968), 'Crime and Punishment: An Economic Analysis', *Journal of Political Economy*, v 76, p 169-217.

[4] Bond, J., (1993), 'Self-Reported vs. Objective Measures of health in Retirement Models', *Journal of Human Resources*, v 26, 106-20

[5] Bradburn, N., and S. Sudman, (1988), *Polls and Surveys*, Jossey-Bass Publishers, San Francisco, CA.

[6] Burstein, P., and J. Cromwell, (1985), 'Relative Incomes and Rates of Return for US Physicians', *Journal of Health Economics*, v 4, 63-78.

[7] Butler, J., R. Burkhauser, J. Mitchell, and T. Pincus, (1987), 'Measurement Error in Self-Reported Health Variables', *Review of Economics and Statistics*, 644-650.

[8] Cochrane, W., (1979), *Survey Sampling*, Wiley & Sons , New York.

[9] Cox, D., (1958), *The Design of Experiments*, New York: John Wiley & Sons.

[10] Griliches, Z., (1986), 'Economic Data Issues', chapter 25 in *Handbook of Econometrics*, edited by Griliches, Z., and M., Intriligator, Volume II, New York: North-Holland.

[11] Heckman, J., (1979), 'Sample Selection Bias as a Specification Error', *Econometrica*, v 47, 153-161.

[12] Heckman, J., and R. Robb, (1986), 'Alternative Identifying Assumptions in Econometric Models of Selection Bias', *Advances in Econometrics*, v 5, 243-287.

[13] Holmstrom, B. and P. Milgrom, (1991), 'Multitask Principal-Agent Analysis : Incentive Contracts, Asset Ownership, and Job Design=, *The Journal of Law, Economics, and Organization*, v 7, 25-52.

[14] Juster, T., and R. Suzman, (1994), 'The Health and Retirement Study: An Overview', *The Health and Retirement Study Working Paper Series*, # 94-1001, Institute for Social Research, University of Michigan.

[15] Lazear, E., (1986), 'Salaries and Piece Rates', *Journal of Business*, v 59, 405-31.

[16] Manski, C. F., and J. Dominitz, (1996), 'Eliciting Student Expectations of the Returns to Schooling', *Journal of Human Resources*, v31 n 1, pp. 1-26.

[17] Manski, C. F., (1995), *Identification problems in the social sciences* , Cambridge and London: Harvard University Press.

[18] Manski, C. F., (1990), 'The Use of Intentions Data to Predict Behavior: A Best-Case Analysis', *Journal of the American Statistical Association*, v 85 n 412, pp. 934-40.

[19] Margquis, M., S. Margquis, and M. Polich , (1986), 'Response Bias and Reliability in Sensitive Topic Surveys', *Journal of The American Statistical Association*, v 81, 381-91.

[20] Philipson, T., (1997), 'Data Markets and The Production of Surveys', forthcoming, *The Review of Economic Studies.*

[21] Philipson, T., and D. Grabowski, (1996), *The Survey Supply Experiment*, Population Research Center, NORC, University of Chicago.

[22] Schowalter, M., and N. Thurston, (1996), 'Taxes and Labor Supply of High Income Physicians', forthcoming, *Journal of Public Economics.*

[23] Singh, S., and G. Maddala, (1976), 'A Function for the Size Distribution of Incomes', *Econometrica*, v 44, No 5, p 963- 69.

[24] Steffey, D., and N. Bradburn, (1994), *Counting People in The Information Age*, National Academy Press, Washington, DC.

A GESTURE OF APPRECIATION

Due to the importance of this survey in evaluating the indicators of outstanding hospital care, the National Opinion Research Center at the University of Chicago will compensate you for the short time involved in completing this questionnaire.

As a small expression of our gratitude for your participation, we have included a two dollar bill. We will provide additional compensation of **500 dollars** to those among a randomly selected ten percent of respondents who have completed the entire questionnaire and provided correct answers to two factual questions. The first of these factual questions, #4, asks for your 3-digit AMA primary specialty code. The second factual question, #5, requests the exact date that you received your Medical Degree.

We will confidentially check 10 percent of the returned surveys against the records of the AMA. If you are among those selected and your responses are consistent with AMA records, we will pay you 500 dollars.

Your views on care given to patients in United States hospitals are of immeasurable benefit to our research. Thank you in advance for your contribution.

[ ] No Reimbursement Requested

If you wish to decline our offer of additional compensation, please check this box and include this form with your returned questionnaire.

Appendix 2

We want to show that there exists a set of randomized wages Υ for which θ is identified from the supplied outcomes

$$\{m(w); w \in \Upsilon\}$$

We first prove this for the case when those who do not have the condition have errorless supply; $s_0 = 1$. In this case, two randomized strictly positive wage levels (w, w') and a control group with no pay yields the supplied proportions

$$m(w) = \theta G(\alpha_1 + \beta_1 w)$$

$$m(w') = \theta G(\alpha_1 + \beta_1 w')$$

$$m(w_0) = \theta G(\alpha_1)$$

This has three unknown parameters, $(\theta, \beta_0, \beta_1)$, to be identified from three supplied proportions and as long as supply is elastic, $\beta_1 > 0$, this is feasible with a known index-function G.

If those who do not have the condition have erroneous supply, then more wages are needed but the same idea applies. Now compensation for errorless supply takes place only for those who have the condition. This allows for differencing out the ones who do not have the condition. More precisely, if (w, w, w'') are wages for individuals having the condition supplying without error the differences in supplied proportions are

$$m(w) - m(w_0) = \theta[G(\alpha_1 + \beta_1 w) - G(\alpha_1)]$$

$$m(w') - m(w_0) = \theta[G(\alpha_1 + \beta_1 w') - G(\alpha_1)]$$

$$m(w'') - m(w_0) = \theta[G(\alpha_1 + \beta_1 w'') - G(\alpha_1)]$$

which again involves three parameters to be estimated from three observable differences $\Box$.

Table 1: Summary Statistics for Variables

Variable	Coding Algorithm	Obs	Mean	Std Dev	Min	Max
TREATMENT GROUPS						
UQuantity (W_U)	**Dummy=1 if individual received an unconditional incentive of W_U=$50,**	**601**	**0.20**	**0.40**	**0**	**1**
CQuantity (W_C)	**Dummy=1 if individual received a conditional incentive of W_C=$50 on response**	**601**	**0.20**	**0.40**	**0**	**1**
Val_Incent	**Dummy=1 if individual received validation incentive of Π=10% and w=$500**	**601**	**0.20**	**0.40**	**0**	**1**
Control	**Dummy=1 if individual received no incentive**	**601**	**0.40**	**0.49**	**0**	**1**
SURVEY SUPPLY VARIABLES						
Particpn (S)	**Dummy=1 if individual supplied survey**	**601**	**0.50**	**0.50**	**0**	**1**
HospSupp (s)	**Assuming response, number of hospitals supplied on survey.**	**319**	**4.21**	**1.50**	**0**	**5**
NhospSup (s)	**Assuming response, number of non-hospital related questions answered.**	**319**	**25.77**	**1.13**	**12**	**26**
MtchGrad (s)	**Dummy=1 if individual supplied correct graduation date**	**319**	**0.87**	**0.34**	**0**	**1**
MtchCode (s)	**Dummy=1 if individual supplied correct specialty code**	**319**	**0.13**	**0.34**	**0**	**1**
DEMOGRAPHIC VARIABLES						
Age2539	**Dummy=1 if individual aged 25-39**	**601**	**0.17**	**0.38**	**0**	**1**
Age4054	**Dummy=1 if individual aged 40-54**	**601**	**0.51**	**0.50**	**0**	**1**
Age55P	**Dummy=1 if individuals age 55 and above**	**601**	**0.32**	**0.47**	**0**	**1**
Male	**Dummy=1 if Male**	**601**	**0.91**	**0.28**	**0**	**1**
West	**Dummy=1 if in West Region of U.S.**	**601**	**0.23**	**0.42**	**0**	**1**
North	**Dummy=1 if in North Region of U.S.**	**601**	**0.21**	**0.41**	**0**	**1**
South	**Dummy=1 if in South Region of U.S.**	**601**	**0.36**	**0.48**	**0**	**1**
Central	**Dummy=1 if in Central Region of U.S.**	**601**	**0.19**	**0.40**	**0**	**1**
Sp_Cancr	**Dummy=1 if Medical Specialty is Cancer**	**601**	**0.16**	**0.37**	**0**	**1**
Sp_Cardi	**Dummy=1 if Medical Specialty is Cardiology**	**601**	**0.16**	**0.37**	**0**	**1**
Sp_Neuro	**Dummy=1 if Medical Specialty is Neurology**	**601**	**0.17**	**0.38**	**0**	**1**
Sp_Optha	**Dummy=1 if Medical Specialty is Ophthalmology**	**601**	**0.17**	**0.37**	**0**	**1**
Sp_Ortho	**Dummy=1 if Medical Specialty is Orthopedics**	**601**	**0.17**	**0.38**	**0**	**1**
Sp_Uro	**Dummy=1 if Medical Specialty is Urology**	**601**	**0.16**	**0.37**	**0**	**1**
InB50	**Dummy=1 if income below $50,000**	**271**	**0.04**	**0.20**	**0**	**1**
In50100	**Dummy=1 if income between $50,000 and $100,000**	**271**	**0.06**	**0.24**	**0**	**1**
In100150	**Dummy=1 if income between $100,000 and $150,000**	**271**	**0.20**	**0.40**	**0**	**1**
In150175	**Dummy=1 if income between $150,000 and $175,000**	**271**	**0.12**	**0.32**	**0**	**1**
In175200	**Dummy=1 if income between $175,000 and $200,000**	**271**	**0.13**	**0.34**	**0**	**1**
In200A	**Dummy=1 if income above $200,000**	**271**	**0.45**	**0.50**	**0**	**1**

Table 2A: Proportion of sample members without measurement error for Specialty Code across Incentive and No Incentive Groups (Standard errors in parentheses)[1]

	Val_Incent	Control	Difference	F-Statistic (p-value)
Full Population	0.27 (0.06)	0.11 (0.02)	0.16	11.29 (0.01)
Male	0.28 (0.07)	0.12 (0.02)	0.16	9.43 (0.00)
Inc<200K	0.30 (0.09)	0.09 (0.04)	0.21	6.41 (0.01)
In200A	0.25 (0.10)	0.12 (0.03)	0.13	2.64 (0.11)
West	0.33 (0.14)	0.03 (0.02)	0.30	17.82 (0.00)
South	0.42 (0.15)	0.14 (0.03)	0.28	6.32 (0.01)
Central	0.36 (0.13)	0.15 (0.04)	0.21	3.40 (0.07)
Sp_Cancr	0.25 (0.16)	0.02 (0.02)	0.23	9.62 (0.00)
Sp_Neuro	0.08 (0.08)	0.05 (0.03)	0.03	0.29 (0.66)
Sp_Optha	0.50 (0.22)	0.22 (0.05)	0.28	2.35 (0.13)
Sp_Ortho	0.64 (0.15)	0.32 (0.06)	0.32	4.06 (0.05)
Age2539	0.44 (0.18)	0.03 (0.02)	0.41	21.74 (0.00)
Age4054	0.13 (0.06)	0.10 (0.02)	0.03	0.36 (0.55)
Age55P	0.46 (0.14)	0.16 (0.03)	0.30	7.47 (0.01)

[1] Certain categories which were non-comparable were excluded.

Table 2B: Proportion of sample members without measurement error for graduation date across Incentive and No Incentive Groups. (Standard errors in parentheses)[1]

	Val_Incent	Control	Difference	F-statistic (p-value)
Full Population	0.83 (0.05)	0.87 (0.02)	-0.04	0.83 (0.36)
Male	0.83 (0.06)	0.87 (0.02)	-0.04	1.03 (0.31)
Inc<200K	0.81 (0.08)	0.93 (0.03)	-0.08	2.55 (0.11)
In200A	0.85 (0.08)	0.87 (0.03)	-0.02	0.05 (0.83)
West	0.92 (0.08)	0.84 (0.04)	0.08	0.49 (0.48)
North	0.71 (0.13)	0.89 (0.03)	-0.18	3.29 (0.07)
South	0.83 (0.11)	0.85 (0.03)	-0.02	0.03 (0.86)
Central	0.86 (0.10)	0.92 (0.03)	-0.06	0.67 (0.42)
Sp_Cancr	0.88 (0.12)	0.88 (0.04)	0.00	0.00 (0.99)
Sp_Cardi	0.83 (0.17)	0.85 (0.05)	-0.02	0.01 (0.92)
Sp_Neuro	0.83 (0.11)	0.86 (0.04)	-0.03	0.08 (0.78)
Sp_Optha	0.67 (0.21)	0.88 (0.04)	-0.21	2.20 (0.14)
Sp_Ortho	1.00 (0.00)	0.92 (0.03)	0.08	0.94 (0.34)
Sp_Uro	0.67 (0.17)	0.84 (0.04)	-0.17	1.56 (0.22)
Age2539	0.67 (0.17)	0.89 (0.04)	-0.22	3.11 (0.08)
Age4054	0.83 (0.07)	0.87 (0.02)	-0.04	0.33 (0.57)
Age55P	0.92 (0.08)	0.86 (0.03)	0.06	0.32 (0.57)

[1] Certain categories which were non-comparable were excluded.

Table 3: Coefficient Estimates for Measurement Error Effects.
Dependent Variable: MtchCode (Supplied Errorless Specialty Code)

	Equation			
Independent Variable	(1)	(2)	(3)	(4)
Val_Incent	1.13** (0.35)	1.15** (0.38)	1.35** (0.41)	1.97** (0.53)
InB50	---	0.43 (0.68)	0.31 (0.77)	1.62* (0.96)
In50100	---	-0.76 (0.77)	-0.70 (0.83)	-0.50 (0.94)
In100150	---	-0.07 (0.42)	-0.15 (0.44)	0.39 (0.52)
In150175	---	0.19 (0.51)	0.09 (0.54)	1.08 (0.67)
In175200	---	-0.21 (0.53)	-0.02 (0.54)	0.23 (0.60)
West	---	---	-1.22** (0.51)	-1.22** (0.58)
North	---	---	-1.03** (0.48)	-0.96* (0.55)
Central	---	---	-0.14 (0.40)	-0.08 (0.49)
Age2539	---	---	0.00 (0.55)	0.01 (0.61)
Age55P	---	---	0.84** (0.37)	0.61 (0.44)
Male	---	---	1.53 (1.07)	1.07 (1.15)
Sp_Cancr	---	---	---	1.28 (1.05)
Sp_Cardi	---	---	---	---[a]
Sp_Neuro	---	---	---	0.39 (1.11)
Sp_Optha	---	---	---	3.35** (0.92)
Sp_Ortho	---	---	---	3.63** (0.90)
Constant	-2.13** (0.17)	-2.06** (0.25)	-3.39** (1.11)	-5.52** (1.53)
# of Obs	429	368	368	311
Log Likelihood	-157.82	-137.26	-127.51	-92.37
Pseudo R-Squared	0.03	0.04	0.11	0.31

[a]Sp_Cardi (≠0) was dropped because it predicts failure perfectly.
*Indicates that variable is statistically different from zero at 10 percent level.
**Indicates that variable is statistically different from zero at 5 percent level.

Table 4: Means by Conditional Pay, Unconditional Pay and Controls (Standard Errors in Parentheses)

	UQuantity	CQuantity	Control	F-Statistic (p-value)
Age2539	0.16 (0.03)	0.16 (0.03)	0.18 (0.02)	0.22 (0.80)
Age4054	0.53 (0.05)	0.56 (0.05)	0.47 (0.03)	1.68 (0.19)
Age55P	0.31 (0.04)	0.28 (0.04)	0.35 (0.03)	1.03 (0.36)
Male	0.91 (0.03)	0.88 (0.03)	0.93 (0.02)	1 (0.37)
West	0.29 (0.04)	0.27 (0.04)	0.17 (0.02)	4.22 (0.02)
North	0.18 (0.04)	0.21 (0.04)	0.24 (0.03)	0.67 (0.51)
South	0.32 (0.04)	0.30 (0.04)	0.48 (0.03)	8.04 (0.01)
Central	0.21 (0.04)	0.22 (0.04)	0.11 (0.02)	5.11 (0.01)
Particpn (S)	0.67 (0.04)	0.42 (0.05)	0.52 (0.03)	7.63 (0.01)
HospSupp (s)	4.67 (0.11)	4.37 (0.19)	3.88 (0.15)	7.19 (0.01)
NHospSup (s)	25.68 (0.16)	25.78 (0.09)	25.77 (0.12)	0.18 (0.83)
MtchGrad (s)	0.92 (0.03)	0.78 (0.05)	0.89 (0.03)	3.18 (0.04)
MtchCode (s)	0.13 (0.04)	0.15 (0.05)	0.06 (0.02)	2.10 (0.12)
Sp_Cancr	0.17 (0.03)	0.17 (0.03)	0.16 (0.02)	0.05 (0.95)
Sp_Cardi	0.17 (0.03)	0.17 (0.03)	0.16 (0.02)	0.02 (0.98)
Sp_Neuro	0.17 (0.03)	0.17 (0.03)	0.18 (0.03)	0.15 (0.86)
Sp_Optha	0.17 (0.03)	0.17 (0.03)	0.17 (0.02)	0 (0.99)
Sp_Ortho	0.17 (0.03)	0.17 (0.03)	0.17 (0.02)	0.01 (0.99)
Sp_Uro	0.17 (0.03)	0.17 (0.03)	0.16 (0.02)	0 (0.99)
δ_U	1.00	N/A	N/A	N/A
δ_C	N/A	0.93	N/A	N/A

Table 5: Income Distributions by Incentives

	Incentives	Controls
G(50)	**3.2**	**4.9**
G(100)	**8.9**	**11.4**
G(150)	**26.0**	**31.7**
G(175)	**39.8**	**43.9**
G(200)	**52.8**	**59.4**
S(W)	**59.3**	**50.3**
E(Y|W)	**206.4**	**189.9**
Sample Size	**240**	**298**