# Observed Choice and Optimism in Estimating the Effects of Government Policies

August 4, 1994

Eric Rasmusen

*Abstract*

A policy will be used more heavily in a particular time and placewhere its marginal cost is lower. The analyst who treats times and places as identical will overestimate the policy's net benefit, especially for policy intensities greater than exist in his sample. In regression analysis, the problem can be solved by instrumentalvariables and a correction for heteroskedasticity. In an example usingstate-level data, the technique substantially increases theestimated responsiveness of the illegitimacy rate to transferpayments.

Indiana University School of Business, Rm. 456, 10th Street and Fee Lane, Bloomington, Indiana, 47405-1701. Office: (812) 855-9219. Fax: 812-855-3354. Internet:Erasmuse@indiana.edu.
JEL numbers and Keywords: C1, C3, C5, H3, I3. Estimation bias.Poverty. Political economy. Instrumental variables.

I would like to thank Robert Barsky, Trudy Cameron, John Garen,Hashem Pesaran, Simon Potter, Sunil Sharma, Hal Varian, and seminarparticipants at Indiana University, the University of Michigan, theUniversity of Rochester, and the Wharton School for comments.Much of this work was completed while the author was an Olin Fellowat the Center for the Study of the Economy and the State, Universityof Chicago, and on the faculty of UCLA's Anderson Graduate School ofManagement.

## 1. Introduction.

An important problem is how to judge the effect of a governmentpolicy by looking at data on its use and impact in various times and places. The task might be to estimate the effect of government transfers on poverty, of unemployment insurance on unemployment, or of the tax rate on tax revenue. Let thehypothesized relationship be *Impact* = $\beta$*Policy*, or

$$y = \beta x. \quad (1)$$

The observed-choice problem, the subject of this article, is that very commonly $x = x(\beta)$,because the observed policies are not random. They are chosen in recognition of their costs and benefits in particular times and places, so $x$ depends on $\beta$, which differs across observations. If policies are used more where they are more effective on the margin, then both casual empiricism and estimates usingordinary least squares, are biased towards optimism about the effect of the policies. This is not like typicalsources of bias such as omission of relevant variables, whichcan cause bias in either direction; rather, it is like measurement error with one regressor, which generates bias in a predictable direction.

The mathematics of the observed-choice problem are relativelysimple, relying on the theories of instrumental variables and randomcoefficients that are by now well-established in the econometricsliterature, though perhaps not in exactly this combination. Nor isthe idea that individuals make decisions based on costs and benefitsnew; this is the heart of economics. What this paper will contributeis the observation that if decisions are made by rational actors, then cross-sectionestimation of the effects of government policies will be biased, andbiased systematically in favor of government activism.

Section 2 will set up the estimation problem and the bias thatresults (subsection 2.1), show the sign of the bias (2.2), devise aconsistent estimator (2.3), and discuss a different approachsuggested by Garen (2.4). Section 3 will explain the problem moreintuitively (3.1), distinguish it from other econometric problems(3.2), discuss related examples with discrete variables ornonlinearities (3.3), and compare the policymaking problem with theprediction problem (3.4). Section 4 will apply the analysis in aparticular context, the effect of government transfer payments onillegitimacy. Section 5 concludes.

## 2. The Observed-Choice Problem

### 2.1. The Model

The analyst is trying to estimate relationship (2):

$$y = \beta x \,. \tag{2}$$

Each of his $n$ observations consists of an impact level $y$ and a policy level $x$ for a particular time and place, subscripted $i$. The standard approach is to regress $y$ on $x$ in the belief that the true specification is

$$y_i = \beta x_i + \epsilon_i, \tag{3}$$

where $\epsilon \sim (0, \sigma_\epsilon^2)$. As always inestimation, the analyst does not believe equation (3) to be more than an approximation. The true relationship is unlikely to be precisely linear, for example, but linearity is a good approximation when one does not know whether a convex, concave, or wavy function would be appropriate. Similarly, each time and place does not have exactly the same true coefficient, and a more accurate specification would be equation (4), in which the effect of the policy is different for each observation:

$$y_i = \beta_i x_i + \epsilon_i \,. \tag{4}$$

Equation (4), however, is impossible to estimate, since it has $n$ parameters and there are only $n$ observations. Moreover, using approximation (3) might not be misleading, since, inthe absence of other considerations, the regression of $y$ on $x$ does give an unbiased estimate of the average $\beta$. To see this, suppose that the true specification for $\beta_i$ in equation (4) is

$$\beta_i = \overline{\beta} + v_i, \tag{5}$$

where $v \sim (0, \sigma_v^2)$and is independent of $\epsilon$.Using (5), equation (4) becomes

$$y_i = \overline{\beta} x_i + x_i v_i + \epsilon_i \,. \tag{6}$$

The ordinary least squares (OLS) estimate of $\overline{\beta}$ is

$$\widehat{\beta}_{OLS} = \frac{\sum x_i y_i}{\sum x_i^2}, \tag{7}$$

where $\sum$ will denote $\sum_{i=1}^{n}$ throughout the paper. If $v_i$ and $x_i$ are independent, the OLS estimate of $\overline{\beta}$ is unbiased, because the expected value of expression (7) is

$$E\left(\frac{\sum x_i(\overline{\beta}x_i + v_i x_i + \epsilon_i)}{\sum x_i^2}\right) , \tag{8}$$

which equals

$$E\left(\overline{\beta}\frac{\sum x_i^2}{\sum x_i^2}\right) + E\left(\frac{\sum x_i^2 v_i}{\sum x_i^2}\right) + E\left(\frac{\sum x_i \epsilon_i}{\sum x_i^2}\right) . \tag{9}$$

The first and last terms of (9) equal $\overline{\beta}$ and 0, and the middle term equals 0 if $E(x_i^2 v_i) = 0$. Thus, if$x_i$ and $v_i$ are independent, OLS is unbiased.

Despite the unbiasedness of $\widehat{\beta}_{OLS}$,heteroskedasticity does make OLS inefficient and biases the estimated standard errors. The variance of the error termfor observation $i$ is $x_i^2\sigma_u^2 + \sigma_\epsilon^2$, fromequation (6), which varies depending on $x_i$. Although $E(x_i v_i) = 0$, observation $i$'s disturbancedepends on the size of $x_i$. When $x_i$ is large, so is the disturbance, and observation $i$ ought to be weighted less heavily in the estimate. This "varying-parameters" heteroskedasticity is a well-known problem, and the estimate can be improved by weighted least squares as described below in Section 3.2.[1]

A greater difficulty is that $v_i$ and $x_i$ are unlikely to be independent. After all, why is $x_i$ different from $x_j$? Policies are chosen for many different reasons, but benefits are always weighed against costs, and the variable $y$ that the econometrician is examining is very likely to be part of either the benefit or the cost. Suppose, for example, that $x$ is the level of cigarette taxation and $y$ is the amount of deadweight loss. Deadweight loss is a cost, and states where taxes create more deadweight loss will choose lower levels of taxation. Or suppose that $x$ is the level of cigarette taxation, and $y$, the amount of revenue raised, which depends on the potential for smuggling. Revenue is a benefit, and states where cigarette taxes raise more revenue may choose higher levels of taxation.

The relevance of costs and benefits is robust to the details of why the policies are chosen. If the legislators aim to maximize social welfare, it is obvious that they will weigh costs and benefits. But even if their primary concern is to please special

[1]For textbook discussions of varying-parameter models, see pp. 75-89 of Kennedy (1985) and pp. 390-393 of Maddala (1977).

interest groups such as cigarette companies or the beneficiaries of state spending, the legislators will still consider the public costs and benefits if the general public has any political influence whatsoever (as Peltzman [1976] points out). It may well be that lobbying by cigarette companies makes every state set the tax too low from the viewpoint of social welfare, but states where the cost of the tax is low and the benefit is high will nonetheless have the highest taxes, because lobbyists would have to spend more there to obtain a given tax reduction.

This logic says that $x_i$ depends on $\beta_i$, and on otherfactors

which will be incorporated as an exogenous variable $w$, so a third equation, equation (12), is required to describe the complete system:

$$y_i = \beta_i x_i + \epsilon_i \ , \tag{10}$$

$$\beta_i = \overline{\beta} + v_i \ , \tag{11}$$

and

$$x_i = \gamma_1 + \gamma_2 \beta_i + \gamma_3 w_i + u_i \ , \tag{12}$$

where it will be assumed that: (i) $\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 \sum w_i/N > 0$, (ii)$\overline{\beta} > 0$, (iii) $w$ and $\overline{\beta}$ are nonstochastic, (iv) $\epsilon, u$ and $v$ are independent stochastic disturbances with mean zero and finite variance, and (v) $v$ has a symmetric distribution.

Assumptions (i) and (ii) are normalizations. Assumption (i) says that the average value of $x$ is positive.Assumption (ii) says that the policy has a positive effect on the impact, whether the impact be desirable ornot. Assumptions (iii) and (iv) establish what is exogenous. Assumption (v) says that the true coefficients are symmetrically distributed around their average of $\overline{\beta}$.[2]

The system of equations (10) to (12) violates theassumptions of the OLS model in two ways, each harmless by themselves: random parameters and stochastic regressors. The simpler system consisting of (10) and (11) has random parameters, but OLS is still unbiased as an estimate of the expected value of the parameter. The simpler system consisting of (10) and (12) (in which case $\beta_i = \overline{\beta}$) has stochastic regressors, but OLS is also unbiased in that system.Like binary nerve gas, the two problems are harmless individually, but dangerous in combination.

[2]Thisassumption is used just following equation (17) below. Thebias will exist regardless of whether there is skewness or not, butif $Ev_i^3 \neq 0$, analysis of the sign of the bias becomes morecomplicated.

To see that the OLS estimate of $\overline{\beta}$ is biased,combine equations (11) and (12) to obtain

$$x_i = \gamma_1 + \gamma_2\overline{\beta} + \gamma_2 v_i + \gamma_3 w_i + u_i \ . \tag{13}$$

The critical middle term in the $\widehat{\beta}_{OLS}$ equation, (9), which for unbiasedness must equal zero in expectation, is

$$\frac{\sum x_i^2 v_i}{\sum x_i^2} \tag{14}$$

or, using (13),

$$\frac{\sum(\gamma_1 + \gamma_2\overline{\beta} + \gamma_2 v_i + \gamma_3 w_i + u_i)^2 v_i}{\sum x_i^2}. \tag{15}$$

The summed quantity in the numerator can be written as

$$([\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 w_i + u_i] + \gamma_2 v_i)^2 v_i \ , \tag{16}$$

which equals

$$[\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 w_i + u_i]^2 v_i + 2[\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 w_i + u_i]\gamma_2 v_i^2 + \gamma_2^2 v_i^3, \tag{17}$$

the expectation of which equals

$$2\gamma_2[\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 w_i]\sigma_v^2, \tag{18}$$

since $(E(v^3) = 0$ by assumption (v), and $u$ and $v$ are independent.

Expression (18) has the same sign as $\gamma_2[\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 w_i]$. Summed across the $n$ observations, this takes the same sign as $\gamma_2$, since the term in square brackets is positive by assumption (i).

The parameter $\gamma_2$ represents how the marginal impact of the policy affects the policy level chosen. If the policy is used more where it is more effective, then $\gamma_2 > 0$ if $y$ is a desirable impact and $\gamma_2 < 0$ if $y$ is undesirable. Expression (18) takes the same sign as $\gamma_2$, so the conclusion would be that $\beta$ is overestimated if $y$ is desirable and underestimated if $y$ is undesirable. Whether $\gamma_2$ takes those signs is not obvious, however, and Section 2.2 is devoted toinvestigating it.

### 2.2. The Sign of $\gamma_2$: Is a Policy Used More Where it is More Effective?

Section 2.1 showed that the sign of the bias depends on the sign of $\gamma_2$ in equation (12), which is repeated here:

$$x_i = \gamma_1 + \gamma_2\beta_i + \gamma_3 w_i + u_i \ .$$

What can be said about $\gamma_2$ in general, without knowing the particular application? Is the policy used more where it is more effective, so that $\gamma_2$ is positive where the impact is desirable and negative where it is undesirable?

Let us use a general optimization problem to address the question. Consider one time and place $i$ (so we can drop the subscript $i$) where the policy $x$ has an impact $\beta_b x$ which produces a utility benefit of $B(\beta_b x)$, with $B' > 0, B'' \leq 0$; and an impact $\beta_c x$ which produces a utility cost of $C(\beta_c x)$, with $C' > 0, C'' \geq 0$ (and either $C'' > 0$ or $B'' > 0$, to givethe problem an interior solution). Assume the benefit and the costto be separable, so the policymaker's problem is

$$\overset{Max}{x}\ M(x) = B(\beta_b x) - C(\beta_c x). \tag{19}$$

The first order condition is

$$\frac{\partial M}{\partial x} = \beta_b B' - \beta_c C' = 0, \tag{20}$$

and the second order condition is

$$\frac{\partial^2 M}{\partial x^2} = \beta_b^2 B'' - \beta_c^2 C'' < 0. \tag{21}$$

The cross-partials are

$$\frac{\partial^2 M}{\partial x \partial \beta_b} = B' + \beta_b x B'' \tag{22}$$

and

$$\frac{\partial^2 M}{\partial x \partial \beta_c} = -C' - \beta_c x C'' < 0. \tag{23}$$

Because

$$\frac{dx}{d\beta_b} = -\frac{\frac{\partial^2 M}{\partial x \partial \beta_b}}{\frac{\partial^2 M}{\partial x^2}} \qquad \frac{dx}{d\beta_b} = (-)\frac{(?)}{(-)} \tag{24}$$

and

$$\frac{dx}{d\beta_c} = -\frac{\frac{\partial^2 M}{\partial x \partial \beta_c}}{\frac{\partial^2 M}{\partial x^2}} \qquad \frac{dx}{d\beta_c} = (-)\frac{(-)}{(-)} \tag{25}$$

we can conclude that $\frac{dx}{d\beta_c}$ is always negative, but $\frac{dx}{d\beta_b}$ might be positive. A less intense value of the policy is chosenwhen the cost parameter is big, but not necessarily when the benefitparameter is small. There are two implications for the bias of theOLS estimates in Section 2.1:[3]

(a) If $y$ is undesirable, a cost of the policy, then $\gamma_2 < 0$in equation (12). A bigger $\beta_c$ leads to a smaller $x$. Hence, in the original estimation problem, OLS underestimates $\overline{\beta}$ when the impact is undesirable.

(b) If $y$ is desirable, a benefit of the policy, then $\gamma_2$might be either positive or negative. If $B(\cdot)$ is close to linear, then $B''$ is small, expression(22) is positive, and $\gamma_2 > 0$: a bigger $\beta_b$ leads to a bigger $x$. If $B(\cdot)$ is heavily concave (i.e., the benefit $y$ has sharply diminishing marginal utility), then $B''$ is large and $\gamma_2 < 0$.The more intuitive sign is $\gamma_2 > 0$, which says that the policyis used more intensively where it is more effective, in which caseOLS overestimates $\overline{\beta}$, the positive marginal impact. It isalso possible, however, that the policy is used more intensivelywhere it is less effective (the policymaker may wish to attain athreshold benefit, for example, which requires greater use of thepolicy if it is less effective).

It may be helpful to think of the policy $x$ as an expenditure, $PQ^d$, and the impact $\beta_b x$ as the quantity demanded, $Q^d$. Then $\frac{x}{\beta_b x} = \frac{1}{\beta_b}$ is like the price of the good—it is the expenditure divided by the quantity. When $P$ falls, $Q^d$ always rises. But for some goods, demand is elastic, and when $P$ falls, $PQ^d$ rises. For other goods, demand is inelastic, and $PQ$ falls. For goods with elastic demand, $\gamma_2 > 0$, and for goods with inelastic demand, $\gamma_2 < 0$. The direction of the bias of OLS thus depends on the elasticity of demand for the policy's benefits. In the original estimation problem, OLS will overestimate $\overline{\beta}$ if demand for the impact is elastic, and underestimate it if demand is inelastic.

Yet another way to understand this is by realizing that the sameproblem comes up in trying to predict how factor choice changes withtechnical change. If the cost

[3]It is interesting to notethat the result on costs, and sometimes an benefits, leads to thesame conclusion as the folk wisdom that estimation problems usuallylead to coefficients that are too small.

of labor goes up, one can confidentlypredict that a factory's use of labor will fall. If theeffectiveness of labor goes up, one cannot predict whether thefactory will use more labor or less. Ordinarily, we think it will usemore, but that need not be the case.

### 2.3. A Consistent Estimator for the Observed-Choice Problem

The observed-choice problem can be solved by using instrumentalvariables, even though it is not a conventional simultaneity problem.Begin with the system above: equations (10), (11), and (12). Equations (10) and (11) were combined togive (13),

$$x_i = \gamma_1 + \gamma_2\overline{\beta} + \gamma_2 v_i + \gamma_3 w_i + u_i \ ,$$

which can itself be rewritten as

$$x_i = (\gamma_1 + \gamma_3\overline{w} + \gamma_2\overline{\beta}) + \gamma_2 v_i + \gamma_3(w_i - \overline{w}) + u_i \ , \tag{26}$$

where $\overline{w}$ is the sample mean of $w$. Using $(w_i - \overline{w})$ as an instrument for $x_i$, the instrumental variables estimator is

$$\widehat{\beta}_{IV} = \frac{\sum(w_i - \overline{w})y_i}{\sum(w_i - \overline{w})x_i}. \tag{27}$$

Combining equations (10) and (11) yields $y_i = \overline{\beta}x_i + v_i x_i + \epsilon_i$, which can besubstituted into (27) to yield

$$\begin{aligned} plim\ (\widehat{\beta}_{IV}) &= plim\ \left(\frac{\sum(w_i-\overline{w})(\overline{\beta}x_i+v_i x_i+\epsilon_i)}{\sum(w_i-\overline{w})x_i}\right) \\ &= \overline{\beta} + plim\left(\frac{\sum(w_i-\overline{w})v_i x_i}{\sum(w_i-\overline{w})x_i}\right) + plim\left(\frac{\sum(w_i-\overline{w})\epsilon_i)}{\sum(w_i-\overline{w})x_i}\right). \end{aligned} \tag{28}$$

Substituting for $x_i$ from equation (26) gives

$$\begin{aligned} plim\ (\widehat{\beta}_{IV}) &= \overline{\beta} + plim\left(\frac{\sum(w_i-\overline{w})v_i(\gamma_1+\gamma_3\overline{w}+\gamma_2\overline{\beta})}{\sum(w_i-\overline{w})x_i}\right) + plim\left(\frac{\sum(w_i-\overline{w})v_i^2\gamma_2}{\sum(w_i-\overline{w})x_i}\right) + plim\left(\frac{\sum(w_i-\overline{w})^2 v_i\gamma_3}{\sum(w_i-\overline{w})x_i}\right) + \\ &\quad plim\left(\frac{\sum(w_i-\overline{w})v_i u_i}{\sum(w_i-\overline{w})x_i}\right) + plim\left(\frac{\sum(w_i-\overline{w})\epsilon_i)}{\sum(w_i-\overline{w})x_i}\right). \\ &= \overline{\beta}. \end{aligned} \tag{29}$$

Thus, a consistent estimator can be obtained for $\overline{\beta}$ if an instrument, $(w-\overline{w})$, is available for$x$.[4]

Heteroskedasticity is also a problem, because the error in equation (**??**) is $v_i x_i + \epsilon_i$, the variance of which, $x_i^2\sigma_v^2+\sigma_\epsilon^2$, is different for each observation. Weighted instrumental variables is appropriate, with weights $1/\sqrt{(x_i^2\sigma_v^2+\sigma_\epsilon^2)}$, which requires estimates of $\sigma_v^2$ and $\sigma_\epsilon^2$. One procedure to generate estimates of $\sigma_v^2$ and $\sigma_\epsilon^2$ is:

(a) Regress $x$ on $w$ and a constant to get fitted values$\widehat{x}$.

(b) Regress $y$ on $\widehat{x}$ to get the estimated coefficient $\widehat{\beta}$.

(c) Construct estimated errors $e_i = y_i - \widehat{\beta}x_i$.

(d) Regress $e^2$ on $x^2$ and a constant. Let$\widehat{\sigma_\epsilon^2}$ be the estimate of the constant and $\widehat{\sigma_v^2}$ be the estimate of the coefficient.

### 2.4. The Garen Technique

Garen (1984) solves a problem similar to the present one without using instrumental variables, though his procedure isequivalent to 2SLS in some examples (see Garen [1987]). Let us assume that $w$ is not a determinant of $x$, so no instrument is available. The system to be estimated is then:

$$y_i = \overline{\beta}x_i + v_i x_i + \epsilon_i \ , \tag{30}$$

and

$$x_i = \gamma_1 + \gamma_2\overline{\beta} + \gamma_2 v_i + u_i \ , \tag{31}$$

Let us also assume that $u \equiv 0$, which will replaceidentification-by-instrument.

The reason that OLS is biased in equation (30) is that if $y$ is regressed on $x$, the regressor $x$ is correlated with the error term $vx$. This can be viewed as an

[4]The constant is another suitable instrument for $x$here, since $v$ has mean zero. If a constant is used as aninstrument, then $w$ itself can be used, instead of $w-\overline{w}$. This problem differs from the standard instrumentalvariables problem, in which the difficulty is that $x$ iscorrelated with the disturbance $\epsilon$, so, since $\epsilon$has mean zero, the instrument does not itself need to have mean zero.The special difficulty here is the $wv^2\gamma_2$ term. Since $Ev^2 \neq 0$, the instrument must have mean zero or the set ofinstruments must include a constant.

omitted-variable problem, and including a consistent estimate of $vx$ as a separate regressorwould eliminate the bias asymptotically. The analyst can estimate $v_i$ by $\hat{v}_i = x_i - \overline{x} = \gamma_2 v_i$. This is biased unless $\gamma = 1$, but that is unimportant, since the coefficient on $v_i x_i$ in equation (30) is known to be unity and its regression estimate will be ignored anyway. Theanalyst can therefore regress $y$ on $x$ and $\hat{v}x$ to obtain aconsistent estimate of $\overline{\beta}$.

This procedure cannot be used when $u$ does not equal zero—that is, when the policy is partly determined by factors unobserved by the analyst. In that case, $\hat{v}_i = x_i - \overline{x} = \gamma_2 v_i + u_i$, which is correlated with $x_i$ because $x_i$ and $u_i$ are correlated. Because of the correlation with $x_i$, $\hat{v}_i x_i$ is not a consistent estimator even of $\gamma_2 v_i x_i$, and a regression of $y$ on $x$ and $\hat{v}_i x_i$ would not produce a consistent estimate of $\overline{\beta}$. Equation (30) can be rewritten as

$$\begin{aligned} y_i &= \overline{\beta} x_i + (\gamma_2 v_i x_i + u_i x_i) + ([1 - \gamma_2] v_i x_i - u_i x_i) + \epsilon_i \\ &= \overline{\beta} x_i + \hat{v}_i x_i + ([1 - \gamma_2] v_i x_i - u_i x_i) + \epsilon_i \ . \end{aligned} \tag{32}$$

Thus, if $y$ were regressed on $x$ and $\hat{v}_i x_i$, the regressor $x$ would be correlated with $u_i x_i$ in the error term, and the estimate of $\overline{\beta}$ would be biased. The bias disappears only if $u \equiv 0$. Hence the Garen technique, although it does not require an instrument for $x$, does require the analyst to have precise knowledge of the variables that determine $x$.

## 3. Explanation, Examples, and Prediction

### 3.1 An Intuitive Explanation of the Observed-Choice Problem

The algebraic development of Section 2 makes it clear that OLS isbiased when the observed-choice problem is present, but yields verylittle intuition as to why. Diagrams can make it considerablyclearer, and can show why the sign of the bias is unambiguous whenthe impact is a cost but ambiguous when it is a benefit.

Each of the diagrams in Figures 1 and 2 shows two localities, eachwith its own relationship between $x$ and $y$. These relationships, and the average of the two, are shown as rays through the origin. Localities 1 and 2 have slopes $\beta_1$ and $\beta_2$, and the average has slope $\overline{\beta} = (\beta_1 + \beta_2)/2$. Policymakers 1 and 2 choose points on their respective rays. If they choose $x$ ignoring local conditions, $x_1$ and $x_2$ have the same expected value, and the expected average of the two observations is on the middle ray. This corresponds to OLS being unbiased.

In Figure 1, $y$ is a benefit of $x$. In Figure 1a, the moreeffective a policy is in a locality, the *more* intensely it isused. $\gamma_2$ is positive, and a steeper slope makes a policymaker choose a higher level of $x$. Indiana, with a

greater marginal benefit, chooses a higher policy level than Michigan, and $x_1 > x_2$. If the econometrician draws a line through the origin to lie between the two observations and minimize the squared deviations, that line will have a slope *greater* than $\overline{\beta}$. OLS overestimates the marginal benefit.

In Figure 1b, the more effective a policy is in a locality, the *less* intensely it is used. $\gamma_2$ is negative, and a steeperslope makes a policymaker choose a lower level of $x$. Ohio, with a greater marginal benefit, chooses a lower policy level than Nevada, and $x_1 > x_2$. (Note, however, that $y_1 > y_2$; Ohio stillends up with a greater benefit than Nevada.) If the econometriciandraws a line through the origin to lie between the two observations and minimize the squared deviations, that line will have a *negative* slope,contradicting theory. OLS underestimates the marginal benefit, andin fact gives an impossible result.

In Figure 2, *y* is a cost of *x*, and a steeper slope makes a policymaker choose a *lower* level of *x*: $\gamma_2$ is negative. Iowa, with a greater marginal cost, chooses a lower level than Wisconsin: $x_1 < x_2$. If the econometrician draws a line through the origin to lie between the two observations and minimize the squared deviations, that line will have a slope *less* than $\overline{\beta}$. OLS underestimates the marginal cost.

### 3.2 Other Problems, to be Distinguished from theObserved-Choice Problem

The observed-choice problem is easily confused with other problems in estimation such as the mutual-cause problem, simultaneity, and the Lucas critique. It may be useful to distinguish it from them atthis point.

The *mutual cause problem* is present when variables $x$ and $y$ do not really have a causal relationship but are both caused by a third variable $w$ such that $x = x(w)$ and $y = y(w)$. If richer citieshave better roads and fewer high-school dropouts, the correlation between good roads ($x$) and fewer dropouts ($y$) of is positive because of income ($w$). The quality of roads may be a good predictor of the dropout rate in equilibrium, but if the quality were changed arbitrarily the relationship would disappear. The result is an overestimate of the impact, whether it be a benefit or a cost.

*Simultaneity* is present when not only does $y$ depend on $x$, but $x$ depends on $y$: $y = y(x)$ and $x = x(y)$. Adding hospitals to a city reduces mortality, but a city with less mortality needs fewer hospitals. Simultaneity is not special to policy, and the bias can be either overestimation or underestimation, depending on therelationships between $x$ and $y$.

The *Lucas critique* applies when the relation between $x$ and $y$ only lasts until the government tries to take advantage of it, because if $x$ changes, so does $\beta$: $\beta = \beta(x)$. Aggregate output only rises with the money supply if money supply growth is low, so any attempt to increase output by increasing the money supply fails. This problem, which is equivalent to nonlinearity in the relationship between $x$ and $y$, is special to policy, and it can cause either overestimation or underestimation, depending on how $\beta$ changes in response to $x$.

The observed-choice problem is not the mutual cause problem, because $y$ does indeed depend directly on $x$. It is not simultaneity,because $x$ does not depend on $y$. And it is not the Lucas critique, because $\beta$ does not depend on $x$. It is most closely related to the "selection bias" or "self-selection" problems found in binary-choice models.

The most obvious form of self-selection occurs when some individuals take actions that prevent them from appearing in the observed sample, but even if the individuals in the data set are chosen randomly and selection *per se* is not a problem, the values of independent variables that result from individual decisions might depend on unobserved heterogeneity (see Mundlak [1961], Heckman [1976, 1979], and Lee [1978]). The name "selection bias" is, infact, misleading, since the problem exists even if the sample is theentire population. The observed-choice problem can be considered a form of the selectionbias problem, because in both problems the level of the policydepends on other variables or disturbances in the model and OLS isbiased. The standard selection bias model, however, does not involvevarying coefficients or policy choices based on the effectiveness ofthe policy being analyzed. Example 4 in the next subsection may helpto show the similarities and differences between the twoproblems.[5]

### 3.3 Examples with Discrete Choice, Nonlinearities, andSelection Bias.

[5] Garen (1984) has extended the standard selection-bias techniques to a context in which observations are sampled randomly from the population but the values of the regressors depend on the heterogeneity. He specifies the value of the regressors as being chosen from a large set of discrete choices, and the bias arises from an error term with a nonzero expectation that interacts with individual characteristics and the particular choice made by the individual. The solution he proposes is to estimate the nonzero expectation consistently and include it in the regression. The present paper approaches the problem more simply, using a regression model in which coefficients vary across observations and the policies depend on the coefficients, but a version of the Garen technique was discussed in Section 2.4.

A selection of verbal examples may help to further reveal theintuition behind the observed-choice problem, to extend theimplications to nonlinear estimation, and to distinguish it from thestandard selection bias problem. In the following four examples, the policy takes just two levels, adoption or rejection.

*Example 1: Hotel tax revenue, a desirable impact.* A state either has a low or a high hotel tax, trading off the increase in revenue against the harm to the hotel industry. In 25 states, the high hotel tax would raise $100 in revenue per capita more than the low tax, and those states adopt the tax. In the other 25 states, the higher tax would so discourage business that the change in tax revenue per capita would be $0. The analyst notices that the 25 states with the high tax have $100 higher revenue per capita, a difference that is statistically significant. He therefore advises all states to impose high taxes, even though, in truth, the added benefit iszero. He has overestimated the benefit of increasing a policy's intensity.

*Example 2: Welfare mothers, an undesirableimpact.*[6] Transfer payments to unwed mothers can be set at amount 2 or amount 3. In 25 states, the illegitimacy rate will be 200 or 300, depending on the transfer level, and those states set transfers equal to 2 (see Table 1). In 25 other states, the illegitimacy rate will be 200 regardless of the transfer level, and those states set transfers equal to 3. The analyst sees 25 states with transfers of 2 and illegitimacy of 200 and 25 with transfers of 3 and illegitimacy of 200. He concludes that transfers have no effect on illegitimacy, and he suggests that the low-transfer states can increase their transfers to 3 without any adverse effects. But doing so would in fact increase illegitimacy considerably, and the true average effectis an increase of 50 (= [25(100) + 25(0)]/50) in illegitimacy goingfrom transfers of 2 to 3. He has underestimated the cost of increasing the intensity of a policy.

[6] Section 4 contains an empirical version of Example2.

TABLE 1

EXAMPLES 2 and 3

| HIGH RESPONSE STATE | | LOW RESPONSE STATE | |
|---|---|---|---|
| Transfer | Illegitimacy | Transfer | Illegitimacy |
| **2** | **200** | 2 | 200 |
| 3 | 300 | **3** | **200** |
| 4 | 600 | 4 | 600 |

*Example 3: The potential for bias is especially strong forpolicy intensities outside the sample range.* Add another transfer level to Example 2: amount 4, which would result in illegitimacy of 600. Thelow-transfer states keep their transfers at 2, and the high-transfer states stay at 3. The naive analyst advises that transfer levels can be increased to 4 in every state without any effect on illegitimacy. He is wrong; illegitimacy will rise everywhere. The value of policy is especially overestimated for intensities greater than exist in the sample.

This last effect is not just the usual hazard of forecasting out of the observed sample range. The naive analyst may well admit that his predictions for transfers of 4 are outside of his sample range and less trustworthy because of possible nonlinearity in the effect of transfers on illegitimacy. But he will add that although possible nonlinearity reduces the reliability of the prediction, it could result with equal likelihood in either an overestimate or an underestimate of the effect. That is wrong. The very reason why the transfer level of 4 is not in his sample is that the effect is nonlinear in the particular direction unfavorable to the active policy.

Nonlinearities outside the observed sample range could lead to either overestimation or underestimation. It could be that the policy is much *more* effective than we estimate in the range *lower* than we observe. Table 1 and Figure 3 illustrates the problems with extrapolation in either direction. Although the data in Figure 3 mayrepresent the entire population of policy choices, it is not random;it is purposively chosen to be on the middle part of the benefitcurve.

Example 3 has some similarity to the Lucas Critique problem, because the marginal effectiveness of the policy depends on the policy level chosen. This dependence, however, would exist even if the policy levels were chosen randomly. What the observed-choice problem adds is the idea that the policies will be chosen so as to make the Lucas critique especially applicable. The Lucas critique says that *if* the variation in the data is too small,nonlinearities in the function being estimated are a big problem, where "too small" depends on the context. The observed-choice problem explains *why* the variation will be too small.

*Example 4. Job Training and Selection Bias.* The effect of job training programs is the paradigmatic problem for which economists haveworried about selection bias (see Heckman & Robb [1985], Heckman, Hotz & Dabos [1987], orLalonde [1986]). Suppose half of a group of unemployed people had wages of 100 in theirprevious jobs and half had wages of 120. They are offered training, but only the people with past wages of 120 accept the training, forsome exogenous reason. The training makes no difference in productivity for either group. Afterwards, however, the trained people get jobs with wages of 120, and the untrained get wages of 100. If the naive analyst ignores the previous wages, he concludes that the training raised wages by 20percent. Just as easily, the problem could have been that only the 100-wage people accepted training, in which case the bias would have been pessimistic rather than optimistic. In either case, techniques are available for correcting the problem.[7]

The observed-choice problem is different, because it arises out of heterogeneous effects of the training rather than heterogeneous initial wages. Suppose that all the unemployed had previous wages of 100, but half of them would get a benefit of 0 from the training, and half would get a benefit of 20. Those that would benefit from the training accept it. Afterwards, the trained workers have wages of 120and the untrained workers have wages of 100. The inference that the training raised wages by 20 is correct, but the inference that the average effect of training across the entire population is 20 is incorrect; it is 10. In the observed-choice problem, unlike in the problem of heterogeneous initial wages, economics provides prior informationon the direction of the bias.[8]

### 3.4 Prediction without Policymaking

The most important implication of the observed-choice problem is that OLS or the equivalent informal reasoning will lead the analyst to be too optimistic in recommending changes in policy because he will overestimate benefits and underestimate costs. Making predictions for policy recommendations, however, is different from making predictions in general, as has long been known.[9] Policy recommendations implicitly contain a kind of prediction answering the question: "What will happen to $y_i$ if$x_i$ is changed by forces outside the model?" A purer form of prediction asks: "What willhappen to $y_i$ if $x_i$ changes?" These are two different things:"What will happen after I change the policy" might be different from "What will happen after thepolicy changes?"

[7]An early article on this problem is Mundlak (1961), which notes that if good farm management, which is unobserved, has a positive additive effect onoutput and is correlated with use of some input, then the analyst will overestimate the effect of the input on output. For a simpleexposition of this story, see pp. 204-207 of Varian (1992).

[8]More generally, in a continuous-variable version ofthis story, it could be that the workers with *lower* marginalbenefit decide to get more training, because they need more hours oftraining to get the same improvement. Then the estimation biasgoes in the opposite direction. But the direction of bias isunambiguous in a 0/1 model of training/no training.

[9]See Haavelmo (1943), p. 278 of Hurwicz (1950) and p. 56 of Mundlak (1961).

Recall the mutual-cause example in section 3.2 in which high-school dropouts and road quality are inversely correlated across cities. An OLS regression would mislead in making the policy recommendation that the roads be improved to reduce the dropout rate. But the OLS regression would correctly predict that a city with good roads is likely to have a low dropout rate. Likewise, simultaneity is a less dangerous problem for prediction than for policymaking. If a city has a large police force, then usingthe correlation between police and crime to predict a large amount of crime may be correct even though the causal link is that more police reduces crime. If the analyst wants reliable policy implications, he needs a theory of causation; if he just wants to predict, he can use correlation.

Prediction given the observed-choice problem is more tortuous. OLS will underestimate the average impact on $y_i$ of a recommended increase in $x_i$ if $y$ is an undesirable impact, and instrumental variables estimates that impact correctly. But what if $x_i$ takes a large value for reasons internal to the model?

If the analyst is asked to predict $y_i$ for a new observation $i$ that has a policy level of $x_i$, his answer should not be $\hat{y} = \hat{\beta}_{IV} x_i$, even though $\hat{\beta}_{IV}$ is a consistent estimator of $\overline{\beta}$ and the true specification is $y_i = \overline{\beta} x_i + x_i v_i + \epsilon_i$. A large value of $x_i$ is produced by a small value of $\beta_i = \overline{\beta} + v_i$ and therefore by a negative value of $v_i$. The IV estimator will overpredict $y_i$, because $E(y|x) \neq \overline{\beta} x$; instead, $E(y|x) = \overline{\beta} x + E(xv|x)$. The bias in prediction is the *opposite* of the bias in policy recommendation. But whether the bias for observation $i$ is positive or negative depends on the value of $x_i$. Although the bias is downwards when $x$ is large, it is *upwards* when $x$ is small. When $x_i$ is small, the marginal effect of policy is great, and $y_i$ is greater than predicted by the IV estimate. One could use Bayes Rule to estimate $E(\beta_i|x_i) = \int \frac{f(x|\beta) f(\beta)}{f(x)} d\beta$, but this requires knowledge of the functional form of the distribution of $v$, since $\beta_i = \overline{\beta} + v_i$.

TABLE 2

PREDICTION: HOTEL TAX REDUCTION

| Tax of new state | True effect of a high tax | True revenue | Naive Prediction | Sophisticated Prediction |
|---|---|---|---|---|
| High | 100 | 100 | 100 | 50 |
| Low | 0 | 0 | 0 | 0 |

Return to Example 1, the hotel tax. The naive analyst predicts that a state with a high hotel tax will have $100 more in revenue, whereas the analyst who corrects for the observed-choice problem predicts $50. The sophisticated analyst will do better in predicting the effect of increasing the tax in a state that currently has a low tax; he will predict $50, the naive analyst will predict $100, and the true increase will be $0. For high-tax states, the sophisticated analyst predicts a $50 from lowering the tax, the naive analyst $100, and the truth is $100, but over both kinds of states the sophisticated analyst will have lower mean squared error, as well as an unbiased estimate.

In pure prediction, however, the naive analyst does better. Suppose that the problem is to predict the hotel tax revenue in a state outside the original sample, knowing only that the state has a high hotel tax. The naive prediction is that the new state's revenue will be $100 higher than in low-tax states, and the "sophisticated" prediction is $50. Since the reason the new state imposed a high tax was because it would raise revenue there, the true value is $100, and the naive analysis yields the correct answer. The same would be true of a new state with a low hotel tax; the naive prediction that its revenue is $100 below that of states with high taxes is correct, and the sophisticated prediction of $50 is incorrect.

The analyst must decide which kind of question he is answering. Instrumental variables is appropriate for answering questions about exogenous changes in policies, but not for answering questions about endogenous changes or for out-of-sample predictions.

## 4. An Empirical Example: Illegitimacy and Aid to Families withDependent

## Children

As an empirical example, let us consider the problem of estimating the effect of welfare on illegitimacy. Simple economic rationality suggests that if transfer payments are made to individuals contingent on their being single mothers, the number of single mothers will increase. The only question is how much. A survey by Elwood & Crane (1990) on the state of the black family suggests that the answer is very little. As Table 3 shows, thelevels of transfer payments do not show any clear relation to the percentage of black children living with only a single parent, and we have no reason to believe that black women are less sensitive to monetaryincentives than white women.[10] Since Aid For DependentChildren (AFDC) levels vary across states, cross-section estimates have also been made, both reduced-form and structural, but Elwood & Bane tell us that “In general, both methods reveal only weak to moderate effects of welfare” (Elwood & Bane, 1990, p. 74). A 1990 study by Darity & Myers, for example, finds, using CPS data on individuals in different states, that the elasticity of female headship of black families with respect to welfare levels was just .075. This is ageneral finding from time-series and cross-sectional studies; in his*Journal of Economic Literature* survey, Moffit (1992, p. 31)says, “The failure to find strong benefit effects is the mostnotable characteristic of this literature.” At the same time, onelongitudinal study, that of Kneisner, McElroy & Wilcox (1988b),does find a significant effect of monetary incentives onillegitimacy: greater AFDC payments increases the number of womenwho become single mothers, especially for black women, although thesize of the payment does not seem to affect how long they stay onAFDC. Thus, the general conclusion seems to be that the AFDC level ina state does not much affect the number of illegitimate births inthat state, but apparently at the level of the individual, the AFDClevel does affect the decision to become a welfare mother.

[10]In fact, some evidence existsthat black women are more sensitive to monetary incentives, not less.Kneisner, McElroy & Wilcox (1988a) find that a greater correlationbetween poverty and illegitimacy for blacks than for whites,suggesting that a given monetary incentive might be more powerfulfor blacks simply because it is a larger proportion of total income.See also Kneisner, McElroy & Wilcox (1988b), discussed below.

TABLE 3

TRANSFER PAYMENTS OVER TIME [11]

| | 1960 | 1970 | 1980 | 1988 |
|---|---|---|---|---|
| AFDC and food stamp payment level (family of 4 with no income– 1988 dollars CPI-U adjusted) | $7,324 | $9,900 | $8,325 | $7,741 |
| Percent of black children not living with two parents | 33.0 | 41.5 | 57.8 | 61.4 |
| Estimated percent of black children collecting AFDC | 10.4 | 33.6 | 34.9 | 30.1 |

Source: Table 3 of Elwood & Bane (1990).

The observed-choice problem applies to this situation and may helpexplain the discrepancy between the aggregate and the individualestimates. The observed-choice problem applies if the explanatoryvariable is a policy and the dependent variable is a cost. Although economists, with their occupational interests, tend tothink of the disincentive AFDC provides to supplying labor, in theminds of the public, illegitimacy is viewed as one of the chiefcosts of AFDC, and it is reasonable to suppose that the marginaleffect of AFDC differs across states for a variety of cultural andeconomic reasons that are difficult to pick up in aggregateregressions. One explanation for the time series evidence is thatthe social breakdown occurring in the 1960s and 1970s increased themarginal impact of AFDC on illegitimacy for any level of AFDC,shifting up the entire curve, so the government reduced the size ofAFDC payments. Theory cannot predict whether the final effect of anincrease in the marginal impact would be an increase or decrease inillegitimacy; here, it seems to have increased despite the cuts inAFDC. Similarly, the cross-sectional evidence might be the resultof states in which AFDC would have a bigger effect on illegitimacychoosing lower levels of AFDC. It might be, for example, that thenumber of women of each age in a state is important to the effect ofAFDC, and this is difficult to put into a state-by-state regression,with its limited degrees of freedom In longitudinal studies, on theother hand, more individual variables can be taken into account, andthe observed-choice problem is diminished, which might explain thegreater size and significance of the estimatedcoefficients.[12]

[12]Longitudinal studies are not immune from theobserved-choice problem, but it is less likely to

In this section I will use state-level cross-sectional data toillustrate how instrumental variables with a heteroskedasticycorrection might be used to improve our estimates of the effect oftransfer payments on illegitimacy by researchers more familiar withwelfare policy than myself.[13] Table 4 shows the complete dataset. The 1989 *Annual Statistical Supplement* of the *SocialSecurity Bulletin* provides data on average monthly payments perrecipient from the Aid to Families with Dependent Children (AFDC)program for each state plus the District of Columbia (a sample sizeof 51).[14] This varies from state tostate because the federal government does not pay for the entireamount, and gives states some flexibility in eligibilityrequirements, or even in whether they wish to participate atall.[15] The 1990 *StatisticalAbstract of the United States* provides data on the illegitimacyrate, as well as on the average disposable personal income percapita in the state, the percentage of urbanization, and thepercentage of the population that is black. It

be severe.Suppose that individual Vermont women of given race, age, income,etc. respond more to AFDC than do Maine women. The Vermontlegislature will choose a lower level of AFDC, other things equal,and the observed-choice problem is present. The advantage ofindividual data is that the analyst can at least adjust for race,age, and income, so if there exists a missing variable causing theproblem, it must be something special to Vermonters *qua* Vermonters, not to Vermonters *qua* white, young, poor people.

[13]A more thorough analysis woulduse data on counties or individuals, assemble price indices foreach location, try nonlinear specifications, use more instruments,test overidentifying restrictions, test for whether the model shouldbe fully simultaneous, etc. Most importantly, it would aggregateall the benefits of poverty, including medical benefits, housingbenefits, and illegal income. In fact, Orr (1992) suggests analternative explanation for the small effects of AFDC that have beendiscovered: that overall transfer payments show much less varianceacross states than do AFDC payments. Since the objective here isjust to illustrate the observed-choice problem, such refinements areignored. Certain of the model's simplifications would generallytend towards obtaining insignificant results and a small coefficientfor AFDC. Adjusting for local prices, for example, would increase thereal AFDC levels in the Southern states, which have highillegitimacy rates. Also, Nelson & Startz (1990) find thatwhen one variable is being instrumented using one instrument, the IV estimator has a central tendency insmall samples that is biased in the direction of the OLSestimator—towards too small a coefficient, here. Nonetheless, giventhe small number of observations, the main substantive contributionof this analysis is simply to cast doubt on existing estimates thatignore the observed-choice problem.

[14]"AFDC" is "Aid to Families withDependent Children, Amount of Payments, Monthly average perRecipient," for 1987, p. 342, p. xiii, 1990 *StatisticalAbstract of the United States.*

[15]For details of the state and federal responsibilitiesin funding and eligibility criteria, see the *1993 Green Book*,the annual report on entitlement programs of the Committee on Waysand Means, U.S. House of Representatives, which contains additionaldata on maximum possible benefits per family, state shares of thepayments, payments over time, etc.

will be assumed thatthese are the relevant exogenous variables.[16]

A simple regression of illegitimacy on AFDC and a constantyields the following relationship:

$$\begin{aligned} \textit{Illegitimacy} &= \underset{(3.05)}{26.91} \; \underset{(0.026)}{-0.034} * \textit{AFDC}, \end{aligned} \tag{33}$$

(standard errors in parentheses) with $R^2 = .03$. Equation(33) implies that high AFDC payments reduce the illegitimacyrate, but this is, of course, misleading because the simple regression leaves out importantvariables. Regression (34) more appropriately controls for avariety of things which might affect the illegitimacy rate:

$$\begin{aligned} \textit{Illegitimacy} &= \underset{(3.65)}{15.74} \; \underset{(0.021)}{+0.016} * \textit{AFDC} \; \underset{(0.00042)}{-0.00011} * \textit{Income} \; \underset{(0.033)}{+0.024} * \textit{Urbanization} \\ &\quad \underset{(1.71)}{-1.60} * \textit{South} \; \underset{(0.06)}{+0.56} * \textit{Black}, \end{aligned} \tag{34}$$

with $R^2 = 0.79$. Equation (34) would leave us with theconclusion that AFDC payments have almost no effect on theillegitimacy rate. Nor, surprisingly, do any of the other variablesexcept race have large or significant coefficients. Thecoefficients are small enough, in fact, that one might doubt whether increasing the size of the dataset wouldchange the conclusions: the variables are insignificant not becauseof large standard errors, but because of small coefficients.

If the theory of this paper is correct, the problem with equation (34) isnot lack of data, but that the coefficient on AFDC, $\beta_{AFDC}$,is properly a cause of the level of AFDC. For purposes ofestimation, some identifying instrument is needed to replace AFDC,although fortunately a complete model of political decisionmakingis not required. The instrument used here is the percentage of thestate's vote in the

[16] "Illegitimacy" is "1987 births to unmarried women, percent,"p. xiii, 1990 *Statistical Abstract of the United States*."Income" is "Disposable personal income per capita, 1988," p.xviii. "Urbanization" is "Resident population in metro areas,1988, percent," p. xii. "Dukakis vote" is calculated from "1988percent for leading party," p. 246. "South" takes the value of 1if the state is southern under the *Statistical Abstract's*definition, and 0 otherwise. "Black" is the 1990 percentage, p. 26.Estimation was done using the matrix operations in *Mathematica*(Champaign, Illinois: Wolfram Research).

1988 presidential election that went to MichaelDukakis, which is correlated with a state's liberalism and hencewith its tendency to prefer higher levels of AFDC. This is asuitable instrument if (i) liberals tend to value the net benefitsof AFDC more highly than do conservatives, (ii) the presence ofDukakis voters, conditioning on the other variables in the model, isnot a direct cause of illegitimacy, and (iii) the presence of Dukakisvoters is not a direct result of the current rate of illegitimacy. The decisionmaking model does need to be separable in$\beta_{AFDC}$ and the instrument:

$$AFDC = \gamma_1 f(\beta_{AFDC}) + \gamma_2 g(Dukakis\ vote) + u. \tag{35}$$

Equation (35) is the equivalent of equation (12)in the theoretical part of the article. Even if the functions $f$and $g$ were known, equation (35) could not be estimated,since $\beta_{AFDC}$ is unknown. But equation (35) does nothave to be estimated to use instrumental variables. Instead, theother exogenous variables in (36) plus the vote for Dukakiscan be used as instruments for *AFDC*. If $Z$ is the 51-by-6 matrix

$$Z = (Constant, Dukakis\ Vote, Income, Urbanization, South, Black),$$

and

$$X = (Constant, AFDC, Income, Urbanization, South, Black),$$

then using the instrumental variables estimator $(Z'X)^{-1}Z'y$,the estimates become

$$\begin{array}{llllll} Illegitimacy & = 18.43 & +0.19 * AFDC & -0.0023 * Income & +0.091 * Urbanization \\ & (6.03) & (0.096) & (0.0013) & (0.064) \\ & & & & \\ & +3.32 * South & +0.65 * Black. & & \\ & (3.72) & (.11) & & \end{array} \tag{36}$$

In regression (36), the signs on the variables matchintuition and theory. AFDC causes more illegitimacy, and higherincomes reduce it. Not all variables are statistically significant,but the standard errors are at least smaller than the coefficients.From this regression, one might hope that a larger sample size wouldbring all the variables into significance.[17]

[17]The biggestoutlier for three variables— the illegitimacy rate, percentage ofblacks, and vote for Dukakis— is the District of Columbia. WhenD.C. is excluded, the coefficient on AFDC in equation (36)is .18 (with standard error .11) instead of .096.

The theory of this paper instructs us to take an additional step:heteroskedasticity is still present, so weighted least squares shouldbe used. Following the procedure suggested in Section 2.3 generatesthe following equation, where $\widehat{s^2}$ is the variance of theresiduals from regression (39):[18]

$$\widehat{s^2} = \underset{(11.27)}{20.43} \quad \underset{(.00065)}{+.00065 * AFDC^2} \tag{37}$$

From equation (37), $\widehat{\sigma_\epsilon^2} = 20.43$and $\widehat{\sigma_v^2} = .00065$. This suggests that if the$\beta_{AFDC}$ coefficients are normally distributed, abouttwo-thirds of the states' coefficients lie within an interval oflength $.51 (= 2 \cdot \sqrt{.00065})$.[19]

Having estimated an equation determining the size of the error foran individual observation, it is now possible to use weighted leastsquares or GLS to re-estimate the main equation. If we define$\Omega$ to be a diagonal matrix with $20.43 + .00065 * AFDC_i^2$ ondiagonal $i$, then the GLS-IV estimator is $\widehat{\beta} = (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}y$, with standarderrors being the square roots of the diagonals of $[(y - X\widehat{\beta})'\Omega^{-1}(y - X\widehat{\beta})/(51-6)][X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}$, whichyields

$$\begin{aligned} Illegitimacy = \; & \underset{(6.44)}{18.62} \quad \underset{(0.10)}{+0.21 * AFDC} \quad \underset{(0.0014)}{-0.0024 * Income} \quad \underset{(0.070)}{+0.094 * Urbanization} \\ & \underset{(3.68)}{+3.19 * South} \quad \underset{(0.11)}{+0.64 * Black}, \end{aligned} \tag{38}$$

The estimates stay roughly the same as with unweighted instrumentalvariables (though note that AFDC does pass the boundary ofsignificance at the 5 percent level, since$t(45) = 2.014$).[20]

[18]The standard errorsare not presented here to test whether the regression coefficientsare different from zero. The theory says that $\sigma_\epsilon^2$and $\sigma_v^2$ are positive; the only question is how to bestestimate the magnitudes.

[19]The fact that therounded standard error equals the rounded coefficient is coincidence.Since the estimated average coefficient is on the order of .21, thesize of the interval implies that assuming normality and ignoringour prior beliefs that the coefficient on $AFDC^2$ is positive isprobably a mistake. The need to incorporate prior information iseven clearer if the regression is run without the outliers of DC andUtah, in which case the estimated coefficient on$AFDC^2$ isnegative: $-.00011$, with a standard error of .00030.

[20]The small size of the heteroskedasticity maymake one wonder whether the observed choice

Equation (38) says that if the average monthly AFDC payment in a randomly chosen state rises by 10 dollars, our best estimate of the increase in theillegitimacy rate is 2.1 percent. For the average state, this would be an increase in the AFDC payment of 8.1 percentproducing an 8.6 percent increase in the illegitimacy percentage, anelasticity of 1.06. The coefficient on AFDC is both economically andstatistically significant.[21]Regarding the other variables: if the state's per-capita income rises1000 dollars, illegitimacy falls 2.4 percent; if urbanizationincreases 10 percent, illegitimacy rises .94 percent; if the state isin the South, the illegitimacy rate is 3.19 percent higher, and ifthe black percentage rises 10 percent, the illegitimacy rate rises6.4 percent.

Notice the contrast with the initial multiple regression using OLS,equation (34). The sign has changed on *South*, allcoefficients except the constant and *Black* have at least doubled,those on *Urbanization* and *South* have more than doubled, andthose on *AFDC* and *Income* are more than ten times their initialsize. The estimated elasticity of illegitimacy with respect to AFDCfor the average state rises from .08 to 1.06. Adjusting for theobserved-choice problem clearly has a large effect.

As a final, perhaps tangential, point, the regression of *AFDC* on the other variables may be of interest.Since the theory assumes that $\beta_{AFDC}$ is one of the variablesthat explains *AFDC*, and since illegitimacy is endogenous, thisregression

---

problem is the trueproblem in this example. The observed choice problem grows worsewith increasing heterogeneity in the $\beta_i$, and so does theamount of heteroskedasticity. It is not follow, however, that theobserved-choice problem is trivial if heteroskedasticity is small,because the observed-choice problem also depends on how the statesreact to the differences in the $\beta_i$. If $\beta_i$ is almost thesame in every state, but states react very strongly to $\beta_i$ inchoosing their levels of $x_i$, then the observed-choice problem canstill be severe.

[21]Do recall the caveat earlier:this analysis ignores other welfare benefits such as food stamps,medicaid, and housing subsidies. If they are correlated state bystate with AFDC, then what looks like the impact of a 10-dollar, 8.6percent increase in AFDC is actually the effect of amore-than-ten-dollars, 8.6 percent increase in total welfarebenefits. Thus, increasing AFDC by itself might not have such a largeimpact, though welfare policy as a whole would. If, on the otherhand, AFDC and other benefits are negatively correlated, the methodhere underestimates the effect of additional welfare income.

is misspecified and biased, but it gives some idea of theimportant correlations:

$$\begin{array}{llllll} AFDC & = -63.08 & -0.50 * Illegitimacy & +0.012 * Income & -0.37 * Urbanization \\ & (32.40) & (1.13) & (0.002) & (0.22) \\ \\ & -19.21 * South & -0.71 * Black, & +\mathbf{1}.51(Dukakis\ Vote) & \\ & (11.37) & (0.69) & (0.63) & \end{array} \quad (39)$$

with $R^2 = 0.69$. The high $R^2$ indicates that the fit of theinstrumental variables is quite good (and it only falls to .68 if*Illegitimacy* is dropped). The coefficient on *Dukakis Vote* islarge and positive with a small standard error, indicating thatthere is a strong correlation between AFDC and the vote for Dukakiseven after conditioning on the other variables. The negative sign and high standard error on *Illegitimacy* gives some comfort that the instrumental variables regression results are not due to instrumenting for an endogenous AFDC, instead of througha negative correlation with $\beta_{AFDC}$. One might wonder whether the political strength of parents would cause a positive link between illegitimacy and AFDC, but, in fact, the conditional correlation is negative. If it had been positive, and the most important effect is that illegitimacy causes AFDC, then the instrumental variables estimator here would still be consistent, butone would expect that instrumental variables would produce *smaller* coefficients than OLS, not larger, because under OLS some of the apparent effect of AFDC on illegitimacy would really be due to the positive correlation between the political power of the parents and AFDC.

| State | Illegitimacy (%) | AFDC ($/month) | Income ($/year) | Urban-ization (%) | Black (%) | Dukakis vote (%) | Unexplained Illeg. (from(38)) (%) |
|---|---|---|---|---|---|---|---|
| Maine | 19.8 | 125 | 12,955 | 36.1 | 0.3 | 44.7 | 2.8 |
| New Hampshire | 14.7 | 140 | 17,049 | 56.3 | 0.6 | 37.6 | 2.3 |
| Vermont | 18.0 | 159 | 12,941 | 23.2 | 0.4 | 48.9 | -4.9 |
| Massachusetts | 20.9 | 187 | 17,456 | 90.6 | 4.8 | 53.2 | -6.2 |
| Rhode Island | 21.8 | 156 | 14,636 | 92.6 | 3.8 | 55.6 | -5.2 |
| Connecticut | 23.5 | 166 | 19,096 | 92.6 | 8.2 | 48.0 | 2.3 |
| New York | 29.7 | 166 | 16,036 | 91.2 | 16.1 | 51.6 | -3.8 |
| New Jersey | 23.5 | 119 | 18,615 | 100 | 14.4 | 43.8 | 6.2 |
| Pennsylvania | 25.3 | 111 | 14,072 | 84.8 | 9.4 | 50.7 | 3.4 |
| Ohio | 24.9 | 102 | 13,326 | 78.9 | 11.0 | 45.0 | 2.6 |
| Indiana | 22.0 | 84 | 12,834 | 68.1 | 8.4 | 40.2 | 4.9 |
| Illinois | 28.1 | 101 | 15,150 | 82.5 | 16.1 | 49.3 | 6.7 |
| Michigan | 20.4 | 156 | 14,094 | 79.9 | 14.6 | 46.4 | -14.0 |
| Wisconsin | 20.7 | 160 | 13,296 | 66.5 | 4.8 | 51.4 | -8.5 |
| Minnesota | 17.1 | 171 | 14,037 | 66.6 | 1.6 | 52.9 | -11.0 |
| Iowa | 16.2 | 124 | 12,475 | 43.4 | 1.9 | 54.7 | -3.5 |
| Missouri | 23.7 | 87 | 13,340 | 66.0 | 10.8 | 48.2 | 5.9 |
| North Dakota | 13.9 | 125 | 11,388 | 38.4 | 0.5 | 44.0 | -7.2 |
| South Dakota | 19.4 | 94 | 11,611 | 29.1 | 0.3 | 47.2 | 6.2 |
| Nebraska | 16.8 | 108 | 12,773 | 47.6 | 3.4 | 39.8 | -0.2 |
| Kansas | 17.2 | 110 | 13,235 | 53.4 | 5.8 | 44.2 | -1.2 |
| Delaware | 27.7 | 99 | 14,654 | 65.9 | 18.9 | 44.1 | 2.1 |
| Maryland | 31.5 | 115 | 16,397 | 92.9 | 26.1 | 48.9 | -0.4 |
| DC | 59.7 | 124 | 17,464 | 100 | 68.6 | 82.6 | 0.5 |
| Virginia | 22.8 | 97 | 15,050 | 72.2 | 19.0 | 40.3 | -2.1 |
| West Virginia | 21.1 | 80 | 10,306 | 36.5 | 2.9 | 52.2 | 2.1 |
| North Carolina | 24.9 | 92 | 12,259 | 55.4 | 22.1 | 42.0 | -6.0 |
| South Carolina | 29.0 | 66 | 11,102 | 60.5 | 30.1 | 38.5 | -5.0 |
| Georgia | 28.0 | 83 | 12,886 | 64.8 | 26.9 | 40.2 | -3.5 |
| Florida | 27.5 | 84 | 14,338 | 90.8 | 14.2 | 39.1 | 5.0 |
| Kentucky | 20.7 | 72 | 11,081 | 46.1 | 7.5 | 44.5 | 1.4 |
| Tennessee | 26.3 | 54 | 12,212 | 67.1 | 16.3 | 42.1 | 5.7 |
| Alabama | 26.8 | 39 | 11,040 | 67.5 | 25.6 | 40.8 | 0.5 |
| Mississippi | 35.1 | 39 | 9612 | 30.5 | 35.6 | 40.1 | 2.4 |
| Arkansas | 24.6 | 63 | 10,670 | 39.7 | 15.9 | 43.6 | 1.3 |
| Louisiana | 31.9 | 55 | 10,890 | 69.2 | 30.6 | 45.7 | -1.4 |
| Oklahoma | 20.7 | 96 | 10,875 | 58.8 | 6.8 | 42.1 | -4.8 |
| Texas | 19.0 | 56 | 12,777 | 81.3 | 11.9 | 44.0 | 0.9 |
| Montana | 19.4 | 120 | 11,264 | 24.2 | 0.2 | 47.9 | 0.5 |
| Idaho | 13.0 | 95 | 11,190 | 20.0 | 0.4 | 37.9 | -0.6 |
| Wyoming | 15.8 | 117 | 11,667 | 29.2 | 0.8 | 39.5 | -2.3 |
| Colorado | 18.9 | 109 | 14,110 | 81.7 | 3.9 | 46.9 | 1.3 |
| New Mexico | 29.6 | 82 | 10,752 | 48.9 | 1.7 | 48.1 | 14.0 |
| Arizona | 27.2 | 92 | 13,017 | 76.4 | 2.7 | 40.0 | 12.0 |
| Utah | 11.1 | 116 | 10,564 | 77.4 | 0.7 | 33.8 | -14.0 |
| Nevada | 16.4 | 86 | 14,799 | 82.6 | 6.9 | 41.1 | 3.2 |
| Washington | 20.8 | 157 | 14,508 | 81.6 | 2.4 | 50.0 | -4.8 |
| Oregon | 22.4 | 123 | 12,776 | 67.7 | 1.6 | 51.3 | 1.5 |
| California | 27.2 | 191 | 16,035 | 95.7 | 8.2 | 48.9 | -6.8 |
| Alaska | 22.0 | 226 | 16,357 | 41.7 | 3.4 | 40.4 | -10.0 |
| Hawaii | 21.3 | 134 | 14,374 | 76.3 | 1.8 | 54.3 | 1.1 |
| United States | 24.5 | 124 | 14107 | 77.1 | 12.4 | 46.6 | 0 |

## 5. Concluding Remarks

When the independent variable in an econometric problem is the result of a policy decision and the dependent variable is a cost or benefit of that decision, the OLS estimate will have a tendency to overestimate the net benefit of the policy. This will happen if the decisionmakers are rational, even if the dependent variable is not their main concern, and the coefficients vary across observations, two conditions which are harmless separately but dangerous when present in combination.

The observed-choice problem applies to a variety of policies. Whether the analyst wishes to estimate the effects of unemployment insurance, transfer payments, police protection, or speed limits, he should worry about the source of the variation in policies across space and time. If the variation arises from factors unrelated to the main effect being analyzed, OLS is unbiased, but if it arises from differences in the marginal cost or benefit of the policy, bias is introduced. If every decisionmaker is optimizing, then in equilibrium there is no net benefit from changing any policy, but an outside observer, seeing differences in policies correlated with differences in total benefits, may be fooled into thinking that change would help.

Even if the variation in policies does not arise from differences in the coefficient, there may still be an observed-choice problem for any extrapolation beyond the observed data. If the coefficient changes with the level of policy—that is, if the policy has a nonlinear effect—then policymakers will avoid policy ranges for which the marginal costs are high or the marginal benefits low. The absence of a policy from the data provides information about its effect.

The observed-choice problem provides a reason why social experiments are useful. In one experiment, described by Woodbury &Spiegelman (1987), unemployed people in Illinois were selected randomly and offered a $500 bonus if they accepted a job within 11 weeks and held it for at least 4 months. The most obvious reason for such an experiment is that existing variation in policies was insufficient: no state offered such a policy, so its effect could not be measured. A second reason is that the experiment controlled for state-specific effects. A third reason is the observed-choice problem: if Illinois adopted such bonuses as a general policy, instead of being chosen for an experiment, one might conclude that Illinois adopted the policy because it was especially effective there. Experiments that assign policies randomly eliminate this problem. They are, on the hand, costly and full of practical difficulties, as Heckman et al. (1987) point out, so the clever econometrician may, in the end, still be more cost-effective than the cleverexperimentalist.

When policies differ, one should ask why. For the economist, as forthe Freudian, nothing happens by accident. If policies depend on their potential impacts, then naive estimates of those impacts are biased. This willordinarily be the case, since costs and benefits, not random whims, are the motivations behind policy. Therefore, not only must one construct a model of how $x$determines $y$; one must think about whether $\beta_i$ determines $x_i$. If it does, then the uncorrected estimates should only be used as upper bounds on policy effectiveness, or instrumental variables should be usedto correct the estimates. This can make an important differencein problems such as estimating the effect of AFDC on illegitimacy.

## References

Committee on Ways and Means, U.S. House of Representatives (1993)*Overview of Entitlement Programs: 1993 Green Book.* Washington:U.S. Government Printing Office, 1993.

Darity, William and Samuel Myers (1990) "Impacts of Violent Crime on Black Family Structure," *Contemporary Policy Issues* (October 1990), 8: 15-29.

Department of Commerce (1990) *Statistical Abstract of theUnited States*, Washington: Superintendent of Documents, U.S.Government Printing Office.

Ellwood, David & Jonathan Crane (1990) "Family Change Among Black-Americans: What Do We Know?" *Journal of Economic Perspectives*,4: 65-84 (Fall 1990).

Garen, John (1984) "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52 (September 1984) pp. 1199-1218.

Garen, John (1987) "Relationships among Estimators of Triangular Econometric Models," *Economics Letters*, 25, 39-41.

Haavelmo, Trygve (1943) "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 2:1-12. January 1943.

Heckman, James (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5: 475-492.

Heckman, James (1979) "Sample Selection Bias as a SpecificationError," *Econometrica*, 47: 153-161, January 1979.

Heckman, James, V. Joseph Hotz & Marcelo Dabos (1987), "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings," *Evaluation Review*, 11: 395-427.

Heckman, James, & Richard Robb (1985), "Alternative Methods forEvaluating the Impact of Interventions: An Overview," *Journalof Econometrics*, 30:

239-267.

Hurwicz, Leonid (1950) “Prediction and Least Squares,” in Tjalling Koopmans, ed., *Statistical Inference in Dynamic Economic Models*, New York: John Wiley and Sons, 1950.

Kennedy, Peter (1985) *A Guide to Econometrics*, Second Edition, Oxford: Basil Blackwell Ltd, 1985.

Kneisner, Thomas, Marjorie McElroy, and Steven Wilcox (1988a)“Getting into Poverty Without a Husband, and Getting Out, With orWithout,” *AEA Papers and Proceedings*, May 1988, 78: 86-90.

Kneisner, Thomas, Marjorie McElroy, and Steven Wilcox (1988b)“Individuals and Families in Transition: Understanding ChangeThrough Longitudinal Data,” Papers Presented at the Social ScienceResearch Council in Annapolis, Maryland, March 16-18, 1988, U.S.Dept. of Commerce, Bureau of the Census.

Lalonde, Robert (1986) “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *American Economic Review*, 76: 604-620.

Lee, Lung-Fei (1978) “Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables,” *International Economic Review*, 19: 415-433.

Lucas, Robert E. (1976) “Econometric Policy Evaluation: A Critique,” *Journal of Monetary Economics*, 1976 Special Supplement on the Phillips Curve, pp. 19-46.

Maddala, G. (1977) *Econometrics*, New York: McGraw-Hill, Inc., 1977.

Moffit, Robert, (1992) “Incentive Effects of the U.S. WelfareSystem: a Review,” *Journal of Economic Literature*, March 1992,30: 1-61.

Mundlak, Y. (1961) “Empirical Production Functions Free of Management Bias,” *Journal of Farm Economics* 443 (February 1961), pp. 44-56.

Nelson, Charles and Richard Startz (1990), “Some Further Results onthe Exact Small Sample Properties of the Instrumental VariablesEstimator,” *Econometrica*, 58: 967-876, July 1990.

Orr, Lloyd (1992) “Cross-Section Multiple Program Variance inWelfare Benefits,” working paper, Indiana University Department ofEconomics, June 1992.

Peltzman, Sam (1976) “Toward a More General Theory of Regulation,” *Journal of Law and Economics* 19 (August 1976), pp. 211-40.

Social Security Administration (1989) *Annual StatisticalSupplement* to the *Social Security Bulletin*, Washington:Superintendent of Documents, U.S. Government Printing Office.

Varian, Hal (1992) *Microeconomic Analysis, Third Edition.* NewYork: W.W. Norton & Company.

Woodbury, Stephen & Robert Spiegelman (1987) “Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois,” *American Economic Review,* September 1987, pp. 513-530.