Observed Choice and Optimism in Estimating the Effects of Government Policies

(PUBLISHED: *Public Choice* , (1998) 97: 65-91. This is an older version. See the corrections at the end, just before the references.)

Eric Rasmusen

Abstract

A policy will be used more heavily in a time and place where its 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 weighted instrumental variables. Using state-level data, the technique substantially increases the estimated responsiveness of illegitimacy to transfers.

Rasmusen: Professor of Business Economics and Publicy Policy and Sanjay Subheadar Faculty Fellow, Indiana University, Kelley School of Business, BU 456, 1309 E 10th Street, Bloomington, Indiana, 47405-1701. Office: (812) 855-9219. Fax: 812-855-3354. Email: Erasmuse@indiana.edu. Web: http://www.rasmusen.org.

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, James Heckman, Hashem Pesaran, Simon Potter, Sunil Sharma, Hal Varian, and seminar participants at Indiana University, the University of Michigan, the University of Rochester, and the Wharton School for comments. Much of this work was completed while the author was an Olin Fellow at the Center for the Study of the Economy and the State, University of Chicago, and on the faculty of UCLA's Anderson Graduate School of Management. Carl Gwin provided research assistance. The data can be found via my homepage on the World Wide Web, http://www.rasmusen.org.

1. Introduction.

A common task is to judge the effect of a policy by looking at data on its use and impact in various times and places– the effect of transfers on poverty, of unemployment insurance on unemployment, or tax rates on revenue. Let the hypothesized relationship be $Impact = \beta \cdot Policy$, or

$$y = \beta x. \tag{1}$$

The observed-choice problem, occurs when $x = x(\beta)$. Policies are chosen in recognition of their costs and benefits in particular times and places, so x should depend on β, which differs across observations. If policies are used more where they are more effective at the margin, then both casual empiricism and ordinary least squares estimates are biased towards optimism about the policies. This is not like typical sources of bias which can cause bias in either direction (e.g. simultaneity, omission of relevant variables). Rather, it is like measurement error with one regressor, which generates a predictable bias.

The mathematics of the observed-choice problem are relatively simple, relying on well-established theories of instrumental variables and random coefficients. Nor is the idea that individuals make decisions based on costs and benefits new; this is the heart of economics. What this paper will contribute is a combination of these ideas, leading to the observation that when decisions are made by rational actors, cross-section estimation of the effects of government policies will be biased systematically in favor of government activism.

Section 2 will set up the estimation problem and the bias that results (subsection 2.1), show the sign of the bias (2.2), devise a consistent estimator (2.3), and discuss a different approach suggested by Garen (2.4). Section 3 will explain the problem more intuitively (3.1), distinguish it from other econometric problems (3.2), discuss related examples with discrete variables and nonlinearities (3.3), and compare policymaking with prediction (3.4). Section 4 will apply the analysis in a particular context, the effect of government transfer payments on illegitimacy. 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 in estimation, 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 the true function might be convex, concave, or wavy. 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, approximation (3) might not be misleading, since in the absence of other considerations the regression of y on x does give an unbiased estimate of the average β. To see this, suppose that the true specification for β_i in equation (4) is

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

where $v \sim (0, \sigma_v^2)$and is independent of ϵ. 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 term for observation i is $x_i^2\sigma_u^2 + \sigma_\epsilon^2$, from equation (6), which varies with x_i. Although $E(x_i v_i) = 0$, observation i's disturbance depends 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, usually ameliorated by some form of weighted least squares.[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 probably 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.

Costs and benefits are relevant regardless of the details of policy motivation. If the legislators aim to maximize social welfare, it is obvious they will weigh costs and benefits. But even if their primary concern is to please special 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 every state's tobacco taxes are too low for maximizing social welfare because of corporate lobbying, but states where the cost of the tax is low and the benefit is high will have the highest taxes, nonetheless, because lobbyists would have to spend more there to obtain a given tax reduction.

This logic says that x_i depends on β_i and on other factors which will be incorporated as an exogenous variable z, 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

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

where it will be assumed that: (i) $\gamma_1 + \gamma_2\overline{\beta} + \frac{\gamma_3 \sum z_i}{N} > 0$, (ii) $\overline{\beta} > 0$, (iii) z and $\overline{\beta}$ are nonstochastic, (iv) ϵ, 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, saying that the average value of x is positive and the policy has a positive impact value, whether the impact be desirable or not. 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]

System (10) to (12) violates the OLS assumptions 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) (so $\beta_i = \overline{\beta}$) has stochastic regressors, but OLS is unbiased. Like binary nerve gas, the two problems are harmless individually, but dangerous in combination.

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

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

The critical middle term in 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),

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

The summed quantity in the numerator can be written as

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

which equals

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

the expectation of which equals

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

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

Expression (18) has the same sign as $\gamma_2[\gamma_1 + \gamma_2\overline{\beta} + \gamma_3 z_i]$. Summed across the n observations, this takes the same sign as γ_2, since the term in square brackets is positive by assumption (i).

The parameter γ_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 γ_2, so the conclusion would be that β is overestimated if y is desirable and underestimated if y is undesirable. Whether γ_2 takes those signs is not obvious, however, and will be analyzed in Section 2.2.

2.2. The Sign of γ_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 γ_2 in equation (12), which is repeated here:

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

What can be said about γ_2 in general, without knowing the particular application? Is the policy used more where it is more effective, so that γ_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 give the problem an interior solution). Assume the benefit and the cost to 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 chosen when the cost parameter is big, but not necessarily when the benefit parameter is small. There are two implications for the bias of the OLS estimates:[3]

(a) If y is undesirable, a cost of the policy, then $\gamma_2 < 0$. A bigger β_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 γ_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 β_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 policy is used more intensively where it is more effective, in which case OLS overestimates $\overline{\beta}$, the positive marginal impact. It is also possible, however, that the policy is used more intensively where it is less effective. The policymaker may wish to attain a threshold benefit, for example, which requires greater use of the policy if it is less effective.

It may be helpful to think of the policy x as an expenditure, PQ^d, and the beneficial 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 is elastic, and underestimate it if demand is inelastic.

The same problem arises in predicting how input use changes following innovation. If the cost of labor goes up, one can confidently predict that labor use will fall. If the effectiveness of labor goes up, theory cannot predict whether more or less labor will be used. We believe that usually more is used, but this is an empirical question.

2.3. A Consistent Estimator for the Observed-Choice Problem

One way to attack the observed-choice problem when the equation to be estimated is linear is using instrumental variables, even though this is not a conventional simultaneity problem.[4] Begin with the system above: equations (10), (11), and (12). Equations (10) and (11) were combined to give (13), $x_i = \gamma_1 + \gamma_2\overline{\beta} + \gamma_2 v_i + \gamma_3 z_i + u_i$, which can be rewritten as

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

where $\overline{z}$ is the sample mean of z. Using $(z_i - \overline{z})$ as an instrument for x_i, the instrumental variables estimator is[5]

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

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

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

Substituting for x_i from equation (26) gives, because of the separability of x and ϵ,

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

Thus, a consistent estimator can be obtained for $\overline{\beta}$ if an instrument, $(z - \overline{z})$, is available for x.[6]

2.4. The Garen Technique

Garen (1984) solves a problem similar to the present one without using instrumental variables, though his procedure is equivalent to 2SLS in some examples (see Garen [1987]).

Let us assume that z 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 replace identification-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 omitted-variable problem, and including a consistent estimate of vx as a separate regressor would eliminate the bias asymptotically. The analyst can estimate v_i by $\widehat{v_i} = x_i - \overline{x} = \gamma_2 v_i$. 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. The analyst can therefore regress y on x and $\widehat{v}x$ to obtain a consistent 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, $\widehat{v_i} = x_i - \overline{x} = \gamma_2 v_i + u_i$, which is correlated with x_i because x_i and u_i are correlated. Because of the correlation with x_i, $\widehat{v}_i x_i$ is not a consistent estimator even of $\gamma_2 v_i x_i$, and a regression of y on x and $\widehat{v}_i x_i$ would not produce a consistent estimate of $\overline{\beta}$. Equation (30) can be rewritten as

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

Thus, if y were regressed on x and $\widehat{v}_i x_i$, 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 the policy, x, does require the analyst to have precise knowledge of the variables that determine the policy.

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 is biased, but yields little intuition as to why. Diagrams and examples can show that the result is indeed intuitive, and robust.

Figures 1a, 1b, and 2 each show two localities with their own relationships between policy x and impact y, depicted as rays through the origin. Localities 1 and 2 have slopes β_1 and β_2, an average slope of $\overline{\beta} = \frac{(\beta_1+\beta_2}{2}$. 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 1a, y is a benefit of x and the more effective a policy is in a locality, the *more* intensely it is used. γ_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, y is also a benefit of x, but the more effective a policy is in a locality, the *less* intensely it is used. γ_2 is negative, and a steeper slope 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 still ends up with a greater benefit than Nevada.) 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 *negative* slope. OLS underestimates the marginal benefit, and in fact gives an impossible result.

4

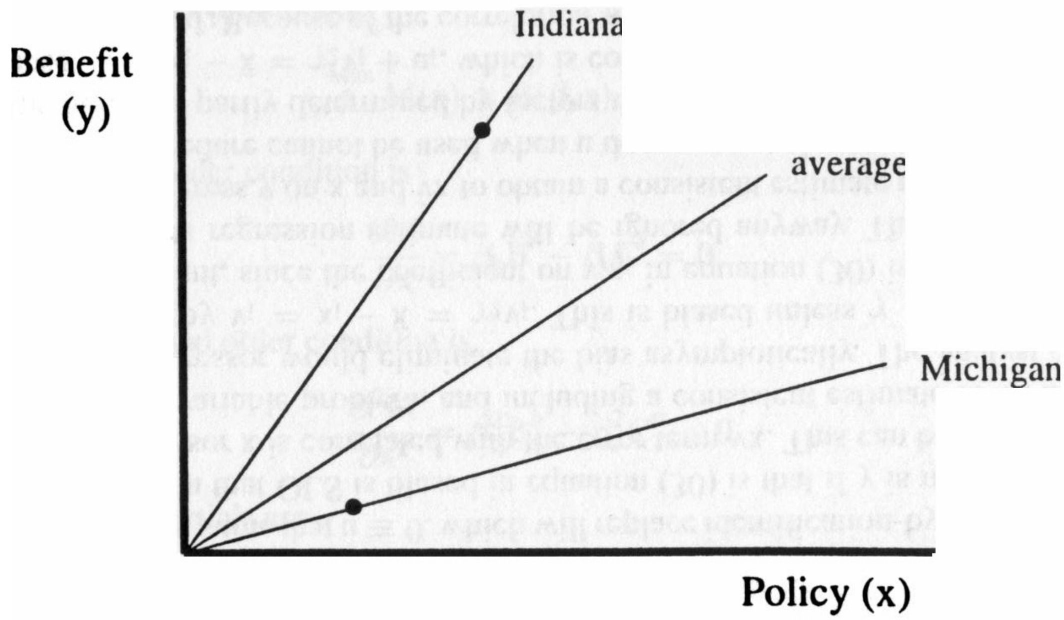

Figure 1a. **A more effective policy is used more**

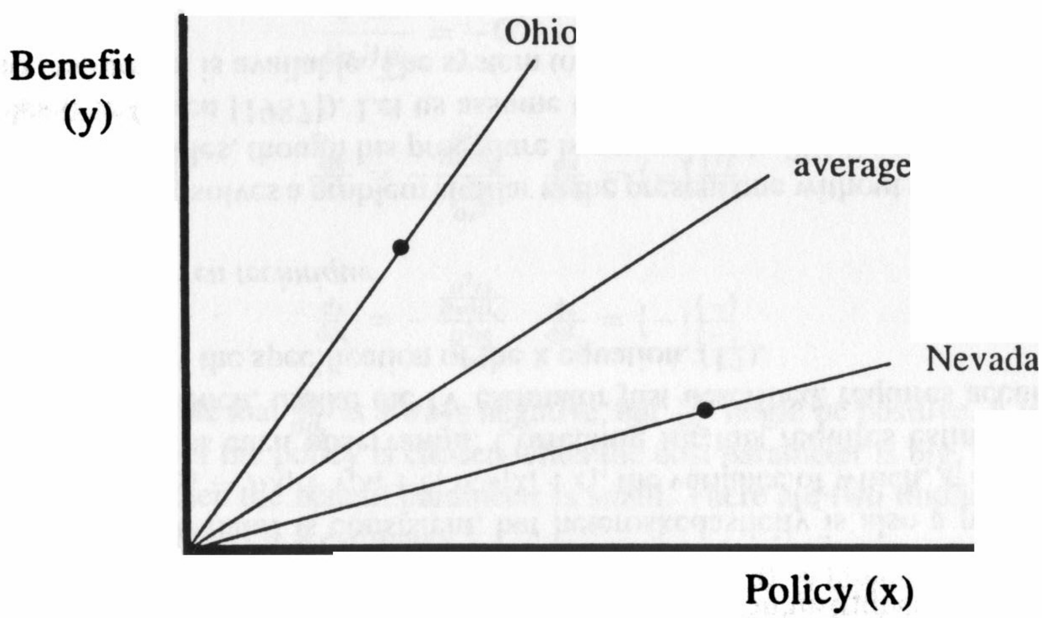

Figure 1b. **A more effective policy is used less**

In Figure 2, y is a *cost* of x, and a steeper slope makes a policymaker choose a lower level of x: γ_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.

75

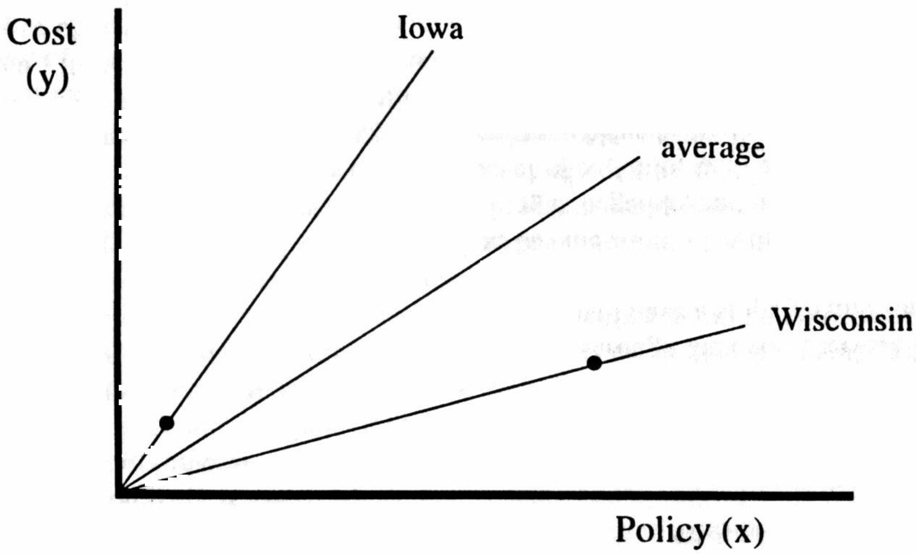

Figure 2. **Estimating the marginal cost of a policy**

3.2 Other Problems, to be Distinguished from the Observed-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.

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 z such that $x = x(z)$ and $y = y(z)$. If richer cities have better roads and fewer high-school dropouts, the correlation between good roads (x) and fewer dropouts (y) is positive because of income (z). 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, since the true impact is zero.

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 over- or underestimation, depending on the relationships 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 β: $\beta = \beta(x)$. 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 over- or underestimation, depending on how β changes in response to x.

The observed-choice problem is not the mutual cause problem, because y does depend on x. It is not simultaneity, because x does not depend on y. And it is not the Lucas critique, because β does not depend on x.

The observed-choice problem is most closely related to the "selection bias" or "self-selection" found in binary-choice models. The observed-choice problem can be considered a form of selection bias, because in both problems the level of the policy— here, continuous rather than just participate/refrain— depends on other variables or disturbances in the model. What is special about the observed-choice problem is that it will be present whenever decisionmakers are rational and coefficients vary between observations, rather than depending on the particular situation being modelled. Section 3.3 will compare the two problems using examples.

3.3 Examples with Discrete Choice, Nonlinearities, and Selection Bias.

In the following four examples, the policy takes just two levels, adoption or rejection.

Example 1: Hotel tax revenue, a desirable impact. A state's hotel tax is either high or low, trading off revenue against harm to tourism. 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 is zero. He has overestimated the benefit of increasing the policy's intensity.

Example 2: Welfare mothers, an undesirable impact. (See also Section 4.) 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, as Table 1 shows, and those states set transfers equal to 2. 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 do not affect illegitimacy and recommends that transfers be increased to 3 everywhere. Doing so would in fact increase illegitimacy considerably, because the true average increase in illegitimacy is 50 (= [25(100) + 25(0)]/50) going from transfers of 2 to 3. He has underestimated the cost of increasing policy intensity.

TABLE 1 GOES HERE

Example 3: The potential for bias is especially strong for policy intensities outside the sample range. Add another transfer level to Example 2: amount 4, which would result in illegitimacy of 600. The low-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 the sample range and less trustworthy because of possible nonlinearity in the effect of transfers. But he will add that although this reduces the reliability of the prediction, it could with equal likelihood result in either over- or underestimate. That is wrong. The very reason why the transfer level of 4 is not in the sample is that the effect is nonlinear in the particular direction unfavorable to more intense 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 may represent the entire population of policy choices, it is not random; there is a reason why the data is in the middle part of the curve.

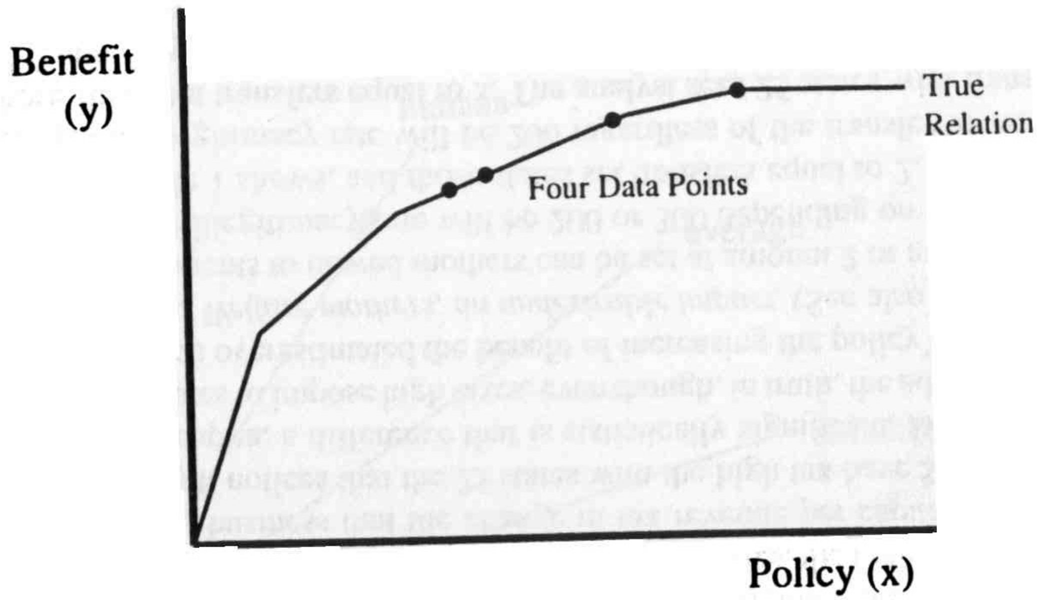

***Figure 3*. The nonlinearity problem of observed choice**

Example 3 has some similarity to the Lucas Critique, 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 context in which economists have worried about selection bias. (see, e. g. Heckman and Robb [1985a, 1985b], Heckman and Smith [1995]). This takes a variety of forms, some of which exemplify the observed-choice problem and some of which do not. Suppose half of a group of unemployed people had wages of 100 in their previous jobs and half had wages of 120. They are all offered training, but only those with past wages of 120 accept it, for some exogenous reason. The training makes no difference in productivity. Afterwards, however, the trained workers earn wages of 120 and the untrained earn 100. If the naive analyst does not know the previous wages, he concludes that training raises wages 20 percent. Just as easily, though, it could happen that only those who earned 100 accepted training, in which case the bias would have been pessimistic.[7]

The observed-choice version of the problem is different, because it arises out of heterogeneous effects of training rather than heterogeneous initial wages. Suppose that all the unemployed had previous wages of 100, but half would get a benefit of 0 from training and half would get 20. Those that would benefit from the training accept it. Afterwards, the trained workers have wages of 120 and 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 information on the direction of the bias.

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. Prediction for policymaking, however, is different from prediction in general.[8] Policymaking asks, "What will happen to y_i if x_i is changed by forces outside the model?" Pure prediction asks, "What will happen to y_i if x_i changes?" This is the difference between "What will happen after I change the policy" and "What will happen after the policy changes?"

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 recommending that roads be improved to reduce the dropout rate, but it would correctly predict that a city with good roads will 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 using the correlation between police and crime to predict a large amount of crime may be correct even though the causal link is that police reduce crime. If the analyst wants to make policy, he needs causation; if he just wants to predict, he needs only correlation.

Prediction in 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 $\widehat{y} = \widehat{\beta}_{IV} x_i$, even though $\widehat{\beta}_{IV}$ 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 GOES HERE

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 a tax decrease in a low-tax state. He will predict \$50, the naive analyst will predict \$100, and the true decrease will be \$0. For high-tax states, the sophisticated analyst predicts a \$50 revenue loss, the naive analyst, \$100, and the truth is \$100. 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 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 endogenous changes or out-of-sample predictions.

4. An Empirical Example: Illegitimacy and Aid to Families with Dependent Children

As an empirical example, let us consider the problem of estimating the effect of welfare on illegitimacy. Economics predicts unambiguously that if transfer payments are made to women contingent on their being single mothers, the number of single mothers will increase. The question is how much. A survey by Elwood and Crane (1990) on the state of the black family suggests that the answer is "very little". As Table 3 shows, the levels of transfer payments do not show any clear relation to the percentage of black children living with a single parent. Since Aid For Dependent Children (AFDC) levels vary across states, cross-section estimates have also been made, both reduced-form and structural, but "In general, both methods reveal only weak to moderate effects of welfare" (Elwood and Crane, 1990: 74). A 1990 study by Darity and 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 is just 0.075. This is a general finding from time-series and cross-sectional studies. In his *Journal of Economic Literature* survey, Moffit (1992: 31) says, "The failure to find strong benefit effects is the most notable characteristic of this literature."[9] At the same time, one longitudinal study, that of Kneisner, McElroy and Wilcox (1989), does find a significant effect of monetary incentives on illegitimacy: greater AFDC payments increases the number of women who become single mothers. The general conclusion, oddly enough is that it seems the AFDC level in a state does not much affect illegitimacy there, but at the level of the individual, AFDC does affect the decision to become a single mother.

TABLE 3 GOES HERE

The observed-choice problem may help explain the discrepancy between aggregate and individual estimates. The problem applies if the explanatory variable is a policy and the dependent variable is a cost. Illegitimacy is one of the chief costs of AFDC, and it is reasonable to suppose that the marginal effect of AFDC differs across states for a variety of cultural and economic reasons that are difficult to pick up in aggregate regressions. One explanation for the time series evidence is that the social breakdown occurring in the 1960s and 1970s increased the marginal impact of AFDC on illegitimacy for any level of AFDC, shifting up the entire curve, so the government reduced the size of AFDC payments. Theory cannot predict whether the final effect of an increase in the marginal impact would be an increase or decrease in illegitimacy; here, it seems to have increased despite the cuts in AFDC. Similarly, the cross-sectional evidence might be the result of states in which AFDC would have a bigger effect on illegitimacy choosing lower levels of AFDC. In longitudinal studies, more variables can be taken into account and the observed-choice problem diminishes, which might explain the greater size and significance of the estimated coefficients.[10]

To illustrate the techniques derived earlier in the paper, I will use state-level data on AFDC and illegitimacy.[11] Table 4 shows the complete dataset. AFDC varies from state to state because the federal government does not pay for the entire amount, and gives states some flexibility in eligibility requirements, or even in whether they wish to participate at all.[12] The variable "AFDC" is defined as the annual AFDC benefit for a woman with two children in the state divided by the mean salary in that state, which adjusts for differences in affluence and cost of living. The 1995 *Statistical Abstract of the United States* provides data on the illegitimacy rate, the percentage of urbanization, and the percentage of the population that is black.[13]

TABLE 4 GOES HERE

A simple regression of illegitimacy on AFDC and a constant yields the following relationship (with standard errors in parentheses):

$$\begin{array}{lll} \textit{Illegitimacy} & = 38.53 & -\mathbf{47.01} * \mathbf{AFDC}, \\ & (3.16) & (\mathbf{15.31}) \end{array} \tag{33}$$

with $R^2 = 0.16$. Equation (33) implies that high AFDC reduces illegitimacy, but this is, of course, misleading because the simple regression leaves out important variables. Regression (34) more appropriately controls for a variety of things which might affect the illegitimacy rate.

$$\begin{array}{lllll} \textit{Illegitimacy} & = 24.0 & +\mathbf{0.47} * \mathbf{AFDC} & +0.63 * \textit{Black} & -4.13 * \textit{South} \\ & (5.38) & (\mathbf{16.38}) & (0.098) & (2.34) \\ & & & & \\ & +0.0000079 * \textit{Income} & -0.0082 * \textit{Urbanization}, & & \\ & (0.00030) & (0.047) & & \end{array} \tag{34}$$

with $R^2 = 0.68$. Equation (34) would leave us with the conclusion that AFDC, with a mean of 0.195 and a coefficient of 0.47, has almost no effect on illegitimacy. Nor, surprisingly, do any of the other variables except race and location in the South have large or significant coefficients. The coefficients are small enough that one might doubt whether increasing the size of the dataset would change the conclusions; the variables are insignificant not because of large standard errors, but small coefficients.

If the theory of this paper is correct, the problem with equation (34) is not just lack of data, but that the coefficient on AFDC, β_{AFDC}, is properly a cause of the level of AFDC. For purposes of estimation, some identifying instrument is needed to replace AFDC. The instrument used here is Michael Dukakis's percentage of the vote in the 1988 presidential election, which is correlated with a state's liberalism and hence with its tendency to prefer higher levels of AFDC.[14] This is a suitable instrument if (i) liberals tend to value the net benefits of AFDC more highly than conservatives, (ii) the presence of Dukakis voters, conditioning on the other variables in the model, is not a direct cause of illegitimacy, and (iii) the presence of Dukakis voters is not a direct result of the current rate of illegitimacy. Also, the decisionmaking model need to be separable in β_{AFDC} and the instrument, as in

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

Equation (35) is the equivalent of the earlier equation (12) . Even if the functions f and g were known, equation (35) could not be estimated, since β_{AFDC} is unknown. But equation (35) does not have to be estimated to use instrumental variables. 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 the instrumental variables estimator is $(Z'X)^{-1}Z'y$ and the estimates become

$$\begin{array}{lllll} Illegitimacy & = 9.10 & +\mathbf{141.97 * AFDC} & +0.95 * Black & +3.13 * South \\ & (13.09) & (\mathbf{95.76}) & (0.27) & (6.06) \\ & & & & \\ & -0.0012 * Income & 0.15 * Urbanization. & & \\ & (0.00093) & (0.13) & & \end{array} \tag{36}$$

In regression (36), the signs on the variables match intuition and theory. AFDC causes more illegitimacy, and higher incomes reduce it. Most of the variables are still statistically significant, but the standard errors are at least smaller than the coefficients. From this regression, one might hope that a larger sample size would bring all the variables into significance.[15]

The average value of AFDC is 19.5%. Increasing this to 20.5% would be a 5.1% increase in the level of the variable. Equation (**??**) says that illegitimacy would rise 1.42% in response, which given the average illegitimacy rate of 30.1% is a 4.7% increase, an elasticity of 0.92. The coefficient on AFDC is thus economically significant.[16]

Notice the contrast with OLS equation (34). The sign has changed on $South$, $Urbanization$, and $Income$, and all coefficients except the constant, $South$, and $Black$ have increased by at least two orders of magnitude, while the estimated elasticity of illegitimacy with respect to AFDC for the average state has risen from 0.0 to 0.92.

Table 5 lists a variety of other regressions, showing that the results are robust to specification.[17] Column (36) is the regression just discussed. Column (36a) applies the same procedure, but with AFDC yearly payments unadjusted for the average salary in the state. Column (36b) replaces AFDC with the ratio of the pretax income equivalent of all welfare payments, including AFDC, food stamps, medical benefits, etc. as computed by Tanner et al. (1995) to the average salary in the state. This addresses the concern of Orr (1992) that overall transfer payments show less variance across states than do AFDC payments, perhaps giving rise to the small cross-sectional effects of AFDC. Columns (34c) and (36c) are regressions that include only $Black$ and $AFDC$. AFDC becomes highly significant in this specification. Although this is shown for only two of them, every specification has the same progression from insignificant and tiny coefficients with OLS to larger more significant coefficients with weighted IV, Correcting for the observed choice problem does make a difference, and might explain why welfare seems to have so little effect on illegitimacy

TABLE 5 GOES HERE

in previous work.

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, OLS has 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 in combination.

The observed-choice problem applies to a variety of policies. Whether the analyst wishes to estimate the effects of transfer payments or speed limits, he should worry about the source of policy variation. If it 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. When decisionmakers are 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, might be fooled into thinking that there is.

Even if the variation in policies does not arise from differences in coefficients, 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 and 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.

When policies differ, one should ask why. For the economist, as for the Freudian, nothing happens by accident. If policies depend on their potential impacts, then naive estimates of those impacts are biased. This will ordinarily 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 β_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 used to correct the estimates. This can make an important difference in problems such as estimating the effect of AFDC on illegitimacy.

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

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

TABLE 3

TRANSFER PAYMENTS OVER TIME

	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 and Crane (1990). Housing and medical benefits, which increased substantially during the 1980's, are not included.

State	Illegitimacy (%)	AFDC/ Avg. Salary (%)	Black (%)	Urban- ization (%)	Avg. Salary ($/year)	Welfare Income Equivalent/ Avg. Salary (%)	Dukakis Vote (%)	Unexpected Illegitimacy (%)
Maine	25.3	23.2	0.4	35.7	21,618	99.9	44.7	3.41
New Hampshire	19.2	27.0	0.6	59.4	24,426	93.3	37.6	-8.46
Vermont	23.4	34.7	0.3	27.0	22,091	94.6	48.9	**-12.81**
Massachusetts	25.9	23.7	5.7	96.2	29,370	103.8	53.2	-1.37
Rhode Island	29.6	27.2	4.4	93.6	24,426	106.9	55.6	-7.01
Connecticut	28.7	25.1	8.9	95.7	32,477	91.1	48.0	0.08
New York	34.8	26.1	17.9	91.7	32,265	84.6	51.6	-3.51
New Jersey	26.4	15.8	14.6	**100.0**	32,152	82.4	43.8	4.53
Pennsylvania	31.6	19.6	9.6	84.8	25,715	76.6	50.7	3.63
Ohio	31.6	16.5	11.2	81.3	24,787	70.2	45.0	5.97
Indiana	29.5	14.7	8.2	71.6	23,507	80.8	40.2	9.19
Illinois	33.4	15.7	15.6	84.0	27,995	69.3	49.3	8.11
Michigan	26.8	21.2	14.8	82.7	27,633	71.3	46.4	-5.78
Wisconsin	26.1	27.0	5.6	68.1	22,951	84.5	51.4	-9.38
Minnesota	23.0	25.5	2.3	69.3	25,075	83.0	52.9	-4.75
Iowa	23.5	24.5	2.0	43.8	20,825	91.2	54.7	-3.97
Missouri	31.5	15.0	11.0	68.3	23,406	63.7	48.2	8.51
North Dakota	22.6	25.8	0.6	41.6	19,030	92.5	44.0	-7.12
South Dakota	26.6	27.5	0.4	32.6	**18,177**	95.2	47.2	-5.09
Nebraska	22.6	21.0	3.9	50.6	20,843	76.3	39.8	-2.57
Kansas	24.3	23.5	6.7	54.6	21,936	80.2	44.2	-6.38
Delaware	32.6	15.4	18.4	82.7	26,375	81.5	44.1	0.28
Maryland	30.5	16.2	26.9	92.8	27,145	84.0	48.9	-11.63
Dist. of Columbia	**66.9**	13.2	**66.0**	**100.0**	**38,128**	76.3	**82.6**	3.78
Virginia	28.3	16.7	19.3	77.5	25,386	91.0	40.3	-7.22
West Virginia	27.7	13.6	3.0	41.8	21,897	69.4	52.2	13.20
North Carolina	31.3	14.5	22.3	66.3	22,443	74.9	42.0	-5.82
South Carolina	35.5	11.2	30.3	69.8	21,432	75.6	38.5	-6.22
Georgia	35.0	13.7	27.5	67.7	24,467	71.1	40.2	-3.71
Florida	34.2	15.6	14.6	93.0	23,370	77.9	39.1	0.13
Kentucky	26.3	12.6	8.1	48.5	21,697	77.4	44.5	7.19
Tennessee	32.7	9.7	19.5	67.7	22,908	59.8	42.1	5.48
Alabama	32.6	8.9	25.3	67.4	22,149	**58.7**	40.8	0.14
Mississippi	42.9	**7.5**	35.7	34.6	19,120	60.1	40.1	3.69
Arkansas	31.0	12.3	15.6	44.7	19,837	66.5	43.6	3.47
Louisiana	40.2	10.4	31.5	75.0	21,971	77.4	45.7	-1.62
Oklahoma	28.4	18.0	7.7	60.1	21,543	82.2	42.1	0.50
Texas	17.5	8.8	12.1	83.9	25,093	60.6	44.0	-1.19
Montana	26.4	24.7	**0.2**	**24.0**	19,467	83.7	47.9	1.71
Idaho	18.3	18.4	0.4	30.0	20,722	86.9	37.9	3.06
Wyoming	24.0	20.1	0.8	29.7	21,546	88.6	39.5	7.00
Colorado	23.8	16.9	4.2	81.8	25,292	82.6	46.9	4.82
New Mexico	39.5	19.8	1.9	56.0	21,689	85.8	48.1	**18.16**
Arizona	36.2	17.7	3.0	84.7	23,453	60.1	40.0	14.54
Utah	**15.1**	22.8	0.7	77.5	21,811	91.2	**33.8**	-12.42
Nevada	33.3	16.0	6.8	84.8	26,177	77.2	41.1	13.79
Washington	25.3	24.9	3.0	83.0	26,306	77.2	50.0	-2.88
Oregon	27.0	23.2	1.7	70.0	23,766	80.8	51.3	1.33
California	34.3	25.2	7.8	96.7	28,910	83.4	48.9	2.22
Alaska	27.4	**35.4**	4.1	41.8	31,309	102.8	40.4	- 4.63
Hawaii	26.2	32.7	2.9	74.7	26,139	**139.3**	54.3	-11.91
United States	30.1	19.5	12.6	79.7	24,358	81.9	46.6	–

TABLE 4: THE DATA AND RESIDUALS

Extreme values are circled. Sources and definitions are in footnote 13 and the text. Southern states are italicized.

Regression:	(33) OLS	(34) OLS	(36) IV	(36a) IV	(36b) IV	(34c) OLS	(36c) IV
AFDC (ratio to salary)	-47.01 (15.31)	0.47 (16.38)	141.97 (95.76)	–	–	14.88 (12.39)	73.35 (29.93)
AFDC ($/year)	–	–	–	0.011 (0.012)	–	–	–
100* (Welfare Income Equivalent)/Income	–	–	–	–	0.59 (0.48)	–	–
Constant	38.54 (3.16)	24.04 (5.38)	9.10 (13.09)	52.30 (33.43)	-11.32 (30.05)	20.2 (3.00)	6.6 (7.02)
Black	–	0.63 (0.098)	0.94 (0.27)	1.33 (0.80)	0.92 (0.29)	0.56 (0.07)	0.75 (0.12)
South	–	-4.13 (2.34)	3.13 (6.06)	6.41 (12.83)	-2.42 (4.45)	–	–
Income	–	0.0000079 (0.00030)	-0.0012 (0.00093)	-0.0046 (0.005)	-0.00094 (0.00093)	–	–
Urbanization	–	-0.0082 (0.047)	0.15 (0.13)	0.30 (0.35)	0.083 (0.11)	–	–

TABLE 5: OTHER SPECIFICATIONS

Dependent variable: Illegitimacy. Sources and definitions are in footnote 13 and the text. Standard errors are in parentheses.

Footnotes

1. On varying-parameter models, see Maddala (1977: 390-393) and Kennedy (1985: 75-89).

2. This assumption is used following equation (17). The bias will exist regardless of whether there is skewness or not, but if $Ev_i^3 \neq 0$, analysis of the sign of the bias becomes more complicated.

3. It is interesting to note that the result on costs leads to the same conclusion as the folk wisdom that estimation problems usually lead to small coefficients.

4. The observed-choice problem can be viewed as a variety of the self- selection problem, which has been attacked in the binary-variable context using not only instrumental variables, but a large number of other estimation approaches. See Heckman and Robb (1985a, 1985b), or Heckman and Smith (1995).

5. The constant is another suitable instrument for x here, since v has mean zero. If a constant is used as an instrument, then z itself can be used, instead of $(z - \overline{z})$. This problem differs from the standard instrumental variables problem, in which the difficulty is that x is correlated with the disturbance ϵ, so, since ϵ has mean zero, the instrument does not itself need to have mean zero. The special difficulty here is the $zv^2\gamma_2$ term. Since $Ev^2 \neq 0$, the instrument must have mean zero or the set of instruments must include a constant.

6. The IV estimator is consistent, but heteroskedasticity is also a problem. The error in $y_i = \overline{\beta}x_i + v_i x_i + \epsilon_i$ 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. Correcting for this requires estimates of σ_v^2 and σ_ϵ^2, which, unlike the IV estimator just described, requires accurate knowledge of the specification of the x equation, (12).

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 on output and is correlated with use of some input, then the analyst will overestimate the effect of the input on output.

For a simple exposition of this story, see Varian (1992: 204-207). This is an example of the observed-choice problem, the heterogeneity in the marginal impact arising from the unobserved input. As Varian explains, a solution for estimating production functions, though one which does not carry over to government policy, is to estimate parameters of the dual cost function instead.

8. The difference between prediction and estimation has long been known. See Haavelmo (1943), Hurwicz (1950: 278) and Mundlak (1961: 56).

9. For a recent exception, which uses state-level data from 1975-1990, see Brinig and Buckley (1995).

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

11. A more thorough analysis would use data on counties or individuals, assemble price indices for each location, try nonlinear specifications, use more instruments, test overidentifying restrictions, test for whether the model should be fully simultaneous, etc.

12. For details of the state and federal responsibilities in funding and eligibility criteria, see the *1993 Green Book*, the annual report on entitlement programs by the House Ways and Means Committee, which contains additional data on maximum possible benefits per family, state shares of the payments, payments over time, and so forth.

13. "AFDC" is "AFDC Benefits" from Table 2 divided by the median wage from Table 12 of Tanner, Moore and Hartman (1995) . "Income" is the median wage. Both are 1995 figures. "Illegitimacy" is "1992 births to unmarried women, percent," p. 77, 1995

Statistical Abstract of the United States. "Black" is the 1995 percentage, calculated from population figures on p. 36. "Urbanization" is "Resident population in metro areas, 1992, percent," p. 39. "Dukakis vote" is calculated from "1988 percent for leading party," p. 246, 1990 *Statistical Abstract* . "South" takes the value of 1 if the state is southern under the *Statistical Abstract's* definition and 0 otherwise (see Table 4). Estimates use the STATA econometrics package (College Station, Texas: Stata Press).

14. The 1992 vote for President Clinton, although more recent, is not so clear a sign of liberalism. The sample correlations of Dukakis Vote with AFDC and Illegitimacy are 0.18 and 0.50.

15. Nelson and Startz (1990) find that when one variable is being instrumented using one instrument, the IV estimator has a central tendency in small samples that is biased in the direction of the OLS estimator—towards too small a coefficient, here. Thus, the small-sample results here are especially encouraging.

16. Recall the caveat earlier: this analysis ignores other welfare benefits such as food stamps, medicaid, and housing subsidies. If they are correlated state by state with AFDC, then what looks like the impact of a 5.1 percent increase in AFDC is actually the impact of a more-than-ten-dollars, 5.1 percent increase in total welfare benefits. If, on the other hand, AFDC and other benefits are negatively correlated, the method here underestimates the effect of additional welfare income. See Equation (36b) in Table 5 for a regression that uses the entire welfare package as an independent variable.

17. The biggest outlier for four variables— the illegitimacy rate, urbanization, percentage of blacks, and vote for Dukakis— is the District of Columbia. When D.C. is excluded, the coefficient and standard error for AFDC in equation (36) are 86.43 and 73.07 rather than 141.97 and 95.77.

November 6, 2000

Mistakes and Extensions for: Eric Rasmusen, "Observed Choice, Estimation, and Optimism About Policy Changes, " Public Choice, (October 1998) 97: 65-91.

(1) There is a mistake on page 73. The paper says,

"If the econometrician draws a line through the origin to go through the two observations and minimize the squared deviations, that line will have a *negative* slope. OLS underestimates the marginal benefit, and in fact give an impossible result"

This is obviously wrong– a line through the origin through the middle of positive datapoints cannot have a negative slope. I was confusing the cases where the intercept is or is not constrained to equal zero. I should have said,

"If the econometrician draws a line through the origin to go through the two observations and minimize the sum of squared deviations, that line will have a slope gentler than the true value. OLS underestimates the marginal benefit."

I could have added:

"If the econometrician does not constrain the line to go through the origin, minimizing the sum of squared deviations will yield a regression line with *negative* slope– the impossible result that more of the policy leads to less benefit."

(2) The paper uses a model with zero intercept and one regressor, so

$$y_i = beta_i x_i + epsilon_i$$

The observed choice bias still exists in the same way if the intercept is allowed to vary among observations, so the model is

$$y_i = alpha_i + beta_i x_i + epsilon_i$$

The slope coefficient betabarhat will be biased as in the simple model with the zero intercept. It doesn't matter if the intercept is assumed to take the same nonzero value for all observations or allowed to vary.

The intercept $\hat{\overline{\alpha}}$ will also be biased, but with the opposite bias of the slope coefficient. This is because it takes the value

$$\hat{\overline{\alpha}} = \overline{x} - \hat{\overline{\beta}} * \overline{\beta}$$

We must, however, make one additional assumption for the presence of an intercept to make no difference: that every observation does use a nonzero value of the policy X. Also, I ignore the information that the intercept must be non-negative, which prior information might perhaps make some technique fancier than OLS optimal. One way to interpret this assumption is that we assume that even if the policy is chosen to be $X = 0$, there is still some benefit or cost Y– a true fixed cost or benefit, as opposed to a cost or benefit that jumps from 0 to some higher level if any policy $X > 0$ is used. This is not a big assumption, and it is, of course, easily checked in the observed data on X.

(3) Typo: On page 67, the first line after equation (3) should have an 'N' so it reads "N(0," rather than "(0,". This is an obvious typo that will not mislead anyone.

(4) Missing information, p. 90: The variable "Welfare Income Equivalent" is taken from Tanner, Moore, and Hartman (1995).

(5) A related article is Summers & Pritchett (1993). They note that countries do not choose policies randomly. They are worried about things such as the tendency for a country that deliberately picks a policy to be in bad shape anyway and to deteriorate despite the policy. See Summers, Lawrence and Lant Pritchett (1993) "The structural-adjustment debate," The American Economic Review, May 1993; 83: 383- 390.

(6) An idea. This paper involves the bother of whether to use expected values or plims, unbiasedness or consistency, in evaluating estimators. I switch to plims and consistency when setting out the instrumental variables estimator because with infinite supports of the normal distribution for regressors, expected values have existence problems. I perhaps should have done that for the OLS part too, since I have stochastic regressors there.

Another way to tackle that problem, I speculate, is to use truncated normal distributions for the error terms instead of normal distributions. That creates problems for small-sample hypothesis testing, perhaps, but not for expected values or consistency

References

Brinig, Margaret and F. Buckley. (1995). *The price of virtue.* Working paper, George Mason University School of Law, August 16, 1995.

Committee on Ways and Means, U.S. House of Representatives. *Overview of Entitlement Programs: 1995 Green Book*, annual. Washington: U.S. Government Printing Office.

Darity, William and Samuel Myers (1990). Impacts of violent crime on black family structure. *Contemporary Policy Issues* 8 (October): 15-29.

Department of Commerce. *Statistical Abstract of the United States*, annual, Washington: Superintendent of Documents, U.S. Government Printing Office.

Ellwood, David and Jonathan Crane (1990). Family change among black Americans: What do we know? *Journal of Economic Perspectives* 4 (Fall) : 65-84.

Garen, John (1984). The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica* 52 (September): 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 aystem of aimultaneous equations. *Econometrica* 2 (January):1-12.

Heckman, James, and Richard Robb (1985a). Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, J. Heckman and B. Singer, eds., New York: Cambridge University Press, pp. 156-245.

Heckman, James, and Richard Robb (1985b). Alternative Methods for Evaluating the Impact of Interventions: An Overview. *Journal of Econometrics* 30: 239-267.

Heckman, James, and Jeffrey Smith (1995). Experimental and non-experimental evaluation. Chapter 1 of *International Handbook of Labour Market Policy and Evaluation*, forthcoming.

Hurwicz, Leonid (1950). Prediction and least squares. In Tjalling Koopmans, ed.

Statistical Inference in Dynamic Economic Models. New York: John Wiley and Sons.

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

Kneisner, Thomas, Marjorie McElroy, and Steven Wilcox (1989). Family structure, race, and the hazards of young women in poverty. In *Individuals and Families in Transition: Understanding Change Thriugh Longitudinal Data*, pp. 33-42. Washington: U.S. Department of Commerce, Bureau of the Census.

Lucas, Robert E. (1976). Econometric policy evaluation: A critique. *Journal of Monetary Economics* 1976 Special Supplement on the Phillips Curve: 19-46.

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

Moffit, Robert (1992). Incentive effects of the U.S. welfare system: A review. *Journal of Economic Literature* 30 (March): 1-61.

Mundlak, Y. (1961). Empirical production functions free of management bias. *Journal of Farm Economics* 443 (February): 44-56.

Nelson, Charles and Richard Startz (1990). Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica* 58 (July): 967-876.

Orr, Lloyd (1992). *Cross-section multiple program variance in welfare benefits*. Working paper, Indiana University Department of Economics, June 1992.

Peltzman, Sam (1976). Toward a more general theory of regulation. *Journal of Law and Economics* 19 (August): 211-40.

Tanner, Michael, Stephen Moore, and David Hartman (1995). *The work versus welfare trade-off: An analysis of the total level of welfare benefits by state.* Washington D.C.: Cato Institute Policy Analysis No. 240 , September 19, 1995. Http://www.cato.org.

Varian, Hal (1992). *Microeconomic Analysis, Third Edition.* New York: W.W. Norton and Company.

Woodbury, Stephen and Robert Spiegelman (1987). Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois. *American Economic Review* (September): 513-530.