Pitfalls of Using Proxy Variables in Studies of

Guns and Crime

Carlisle E. Moody
Department of Economics
College of William and Mary
Williamsburg, VA 23187-8795
757-221-2373
Fax: 757-221-1175
cemood@wm.edu

Thomas B. Marvell
Justec Research
155 Riding Cove
Williamsburg, VA 23185
757-229-9772
Marvell@cox.net

November 10, 2003

ABSTRACT

The magnitude of the estimated coefficient on a proxy variable in a regression is only useful in determining sign and significance. Two recent papers on the relationship between guns and crime incorrectly state that such coefficients have important policy implications.

When data for a variable are missing, researchers frequently use proxy variables. This presents problems when coefficient size is important – that is, when the proxy is used in place of a target variable in policy research (as opposed to when the proxy is a control or when the research aim is limited to sign and significance). When the proxy is a target variable, the magnitude of its coefficient is generally not interpretable because there is another coefficient relating the proxy to the true but unobserved variable (see e.g., Wooldridge, 2000: 284-290; Kmenta, 1986: 579-581). When the proxy is the dependent variable, none of the regression coefficients are interpretable without additional information.

As examples, we use two recent studies of the impact of guns on crime (Duggan, 2001; Cook and Ludwig, 2003). Duggan (2001) uses circulation of the magazine *Guns & Ammo* as a proxy for gun availability. He regresses homicide on lags of the proxy, and, finding a positive relationship, concludes, "The coefficient estimates suggest that a 10 percent increase in the rate of gun ownership is associated with approximately a 2 percent increase in the homicide rate" (p.1096).

Following Duggan, we measure all variables (guns, G, the *Guns & Ammo* circulation rate, GA, and homicide, H) in logarithms, and we assume that H is a linear function of G, *i.e.*,

$$H = a + bG \tag{1}$$

where *b* could be of any sign. Also, guns are linearly related to *Guns & Ammo* circulation rates,

$$G = \alpha + \beta GA \tag{2}$$

where β is presumably positive. Solving for G as a function of GA and substituting into the homicide equation yields

$$H = (a + b\alpha) + (b\beta)GA \tag{3}$$

Thus the effect of a one-unt increase in GA sales on homicide is the product ($b\beta$). The true effect of guns on homicide is $dH / dG = (dH / dGA)(dGA / dG) = (b\beta)(1/\beta) = b$.

Regressing homicide on GA, rather than on G, yields an estimate of ($b\beta$). We do not know the value of β and *therefore we have no measure of the effect of guns on homicide*, aside from its sign and significance. Duggan implicitly assumes that $\beta = 1$ and that GA is a one-for-one proxy for gun ownership.

The same mistake occurs when estimating reverse causation running from crime to guns. Again using the GA proxy, Duggan finds that, " a ten percent increase in the homicide rate is associated with only a 0.2-0.3 percent increase in gun ownership in the subsequent year.... it appears that gun ownership has a much greater impact on murder rates than murder rates have on gun ownership" (p. 1098).

Suppose that homicides in fact cause people to acquire guns in a linear fashion, e.g.,

$$G = c + dH \tag{4}$$

where d>0. Substituting (2) for G, we get

$$GA = (c - \alpha) / \beta + (d / \beta)H \tag{5}$$

Regressing GA (instead of G) on H yields an estimate of the ratio (d/β). Again, without information concerning the value of β, one cannot estimate the size of any impact of homicide on guns. The true effect is

$dG / dH = (dG / dGA)(dGA / dH) = \beta(d / \beta) = d$.

If we are to make any inferences concerning the magnitudes of these coefficients, it is critically important to know if β is greater or less than one. If it is greater than one, then Duggan is overestimating the effect of guns on homicide and underestimating the effect of homicide on guns. If it is less than one, the reverse is true.

Duggan reports a regression of GA on the General Social Survey (GSS) estimate of gun prevalence (Duggan 2001, Table 3, p. 1093). The GSS proportion of households reporting the presence of a gun is the "gold standard" of gun prevalence measures according to Azrael, Cook, and Miller (2001) and we take it to be the true measure of gun prevalence. Duggan estimates that the elasticity of GA with respect to the GSS estimate of gun ownership is .354 with a standard error of .114 (p. 1093). Since Duggan regresses GA on G, he is estimating the inverse of β, which implies that $\beta = 2.82$. He finds that the long run elasticity of homicide to GA (bβ) is .255, while the long run elasticity of GA with respect to homicide (d/β) is .052 (p. 1097). However, the implied true elasticity of homicide with respect to guns is

$$dH / dG = (b\beta)(1/\beta) = (.255)(.354) = .090$$

Similarly, the true elasticity of guns with respect to homicide is

$$dG / dH = \beta(d/\beta) = (2.82)(.052) = .147 .$$

Duggan, therefore, overestimates the effect of guns on homicide, and underestimates the effect on homicide on guns, by a factor of 3. The estimated elasticity of guns with respect to homicide is actually greater than the elasticity of homicide with respect to guns. However, given the standard errors involved, the implied elasticities are not significantly different from each other.[1]

A second example of the same error occurs when Cook and Ludwig (2003) investigate the potential reciprocal relationship between guns and burglary. People might acquire guns to protect themselves against burglary, but, since guns are valuable loot, the presence of guns in the home might encourage burglary. Again, a proxy is used for guns, in this case firearm suicide as a percentage of total suicide (FS). Employing a Granger causality model in first differences, they estimate an elasticity of burglary with respect to guns of .67 (p. 89) and an elasticity of guns with respect to burglary of .06-.07 (p. 91). They conclude, "Thus it appears that gun prevalence drives burglary but burglary does not drive gun prevalence" (p. 91). Again, this claim is not justified because the authors have not shown that the coefficient on the proxy is the same as the (unobserved) coefficient on gun ownership.

Azreal, Cook, and Miller (2001) regress the log of GSS on the log of FS for nine census regions, estimating a coefficient of 2.35 (standard error = .85). Because GSS is the dependent variable, this is an estimate of β (very close to Duggan's estimate of 2.82). This coefficient is numerically greater than one and affects the calculation of the underlying elasticity between guns and burglary. Performing the same analysis as above, we find that the implied underlying elasticity of burglary with respect to guns is .285 and the implied elasticity of guns with respect to burglary is .141. These two

[1] Duggan did not test whether his coefficients were significantly different from each other. He simply compared their magnitudes with no reference to their variances.

coefficients are not significantly different[2]. Thus, Cook and Ludwig have no basis for their conclusions that guns cause burglary but burglary does not cause guns.

In conclusion, researchers must be careful when using proxies for target variables and dependant variables. Estimated coefficients on such proxies should only be used to determine sign and significance.

[2] Cook and Ludwig also do not test to see if the estimated coefficients are significantly different from each other.

References.

Azreal, Deborah, Philip J. Cook, and Matthew Miller, 2001. State and local prevalence of firearms ownership: measurement, structure, and trends, Working Paper 8570, National Bureau of Economic Research, http://www.nber.org/papers/w8570.

Cook, Philip J. and Jens Ludwig 2003. Guns and Burglary, in: Jens Ludwig and Philip J. Cook, eds., *Evaluating Gun Policy* (Brookings Institution Press, Washington, DC) 74-118.

Duggan, Mark, 2001. More guns, more crime, *Journal of Political Economy* 109, 1086-1114.

Kmenta, Jan 1986. *Elements of econometrics*, second edition (Macmillan, New York).

Wooldridge, Jeffrey M., 2000. *Introductory econometrics: a modern approach* (South-Western, Mason, OH).