



War and local collective action in Sierra Leone: A comment on the use of coefficient stability approaches☆



Felipe González^a, Edward Miguel^{a,b,*}

^a Department of Economics, University of California, Berkeley, United States

^b National Bureau of Economic Research

ARTICLE INFO

Article history:

Received 29 August 2014

Received in revised form 6 May 2015

Accepted 14 May 2015

Available online 14 June 2015

JEL classification:

C4

D74

Keywords:

Selection

Unobservables

Civil war

Africa

ABSTRACT

In a study of the effect of civil war exposure on local collective action outcomes in Sierra Leone, Bellows and Miguel (2009) employ a coefficient stability approach to assess the importance of omitted variable bias building on Altonji et al. (2005a). Here we clarify the econometric assumptions underlying Bellows and Miguel (2009), and extend their analysis using data on dependent variable reliability ratios and the method developed in Oster (2015).

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Bellows and Miguel (2009) employ the Altonji et al. (2005a) coefficient stability approach to assess the importance of omitted variable bias, in a study of the effect of civil war exposure (called c below) on local collective action outcomes (y) in Sierra Leone. Specifically, they show that coefficient estimates on civil war exposure change little across regression specifications with and without additional covariates and, following a long tradition in applied economics, argue that omitted variable bias is unlikely to be driving their results.¹

There is an implicit statistical assumption made in Bellows and Miguel (2009), namely, that the variances of the observable covariates ($x'\beta$) predicting local collective outcomes and of the unobservables (\hat{q}) are equal:

$$\sigma_{x'\beta} = \sigma_{\hat{q}}. \quad (1)$$

☆ We would like to thank Raj Chetty, Emily Oster, and three anonymous referees for comments and suggestions.

* Corresponding author at: Evans Hall #3880, University of California, Berkeley, CA 94720-3880

E-mail address: emiguel@berkeley.edu (E. Miguel).

¹ Beyond applying the approach in Altonji et al. (2005a), Bellows and Miguel also present a variety of other evidence that the estimated effects of civil war exposure on local collective action outcomes are causal.

This is not exactly the same assumption that Altonji et al. (2005a) make in their analysis.² If Eq. (1) does not hold, then it follows that the stability of coefficient estimates across specifications with and without observed covariates need not represent convincing evidence of limited omitted variable bias.

As noted by Altonji et al. (2005a), the inherent difficulty that researchers face in establishing a plausible range of values for the ratio of $\sigma_{x'\beta}$ to $\sigma_{\hat{q}}$ limits the practical utility of their approach. In a recent paper, Oster (2015) shows that if observables and unobservables have the same explanatory power in y (after taking into consideration any measurement error in y), then the following is a consistent estimator of the effect of c on y :

$$\hat{\hat{\alpha}} = \hat{\alpha}^* - (\hat{\alpha} - \hat{\alpha}^*) \times \frac{R_{max} - R^*}{R^* - R} \quad (2)$$

where $\hat{\alpha}^*$ and R^* are the coefficient estimate and R^2 from the regression including observable covariates, and $\hat{\alpha}$ and R are the coefficient and R^2 from the uncontrolled regression. In addition, R_{max} is the R^2 in a regression of y on all observable and unobservable controls, which is clearly unknowable (given its reliance on unobservables).

² See condition 4 on page 175 of Altonji et al. (2005a), which states that “the relationship between [the treatment variable and the unobservables] is the same as the relationship between [the treatment variable and the observables], after adjusting for differences in the variance of these distributions” (our italics).

Although several recent papers have employed the approach laid out in Bellows and Miguel (2009), there is no compelling reason to assume that their assumption in Eq. (1) holds in general.³ Therefore, the formula in Bellows and Miguel (2009) should not generally be applied to directly assess the degree of selection; instead, one should apply the corrected formula in Eq. (2) or the equivalent formulas in Oster (2015). Nevertheless, in order to apply this correction we need to have some sense of R_{max} , an unknown parameter.

The best we can hope for is to place plausible bounds on R_{max} . It is immediate that R^* is a lower bound on R_{max} , i.e., $R_{max} \in [R^*, 1]$. In addition, we note that R_{max} is bounded below one when there is classical measurement error in the dependent variable. Evidence on the extent of measurement error in y is thus a potentially useful way to generate an upper bound on R_{max} .⁴ For example, in many low income country households datasets, it is well-known that income, consumption and business profits are measured with considerable error, in which case assuming that $R_{max} = 1$ or that it is close to 1 is likely to be far too conservative. For instance, McKenzie (2012) “demonstrates that for many economic outcomes, the autocorrelations are typically lower than 0.5, with many around 0.3”, with values ranging from 0.2 to 0.8. McKenzie (2012) and De Mel et al. (2009) both suggest that measurement error in the outcome variables is substantial. As a point of reference, in relatively high quality U.S. survey data Angrist and Krueger (1999) conclude that reliability ratios are typically between 0.7–0.9 for the most commonly studied labor market outcomes.

One approach to begin quantifying the extent of measurement error is through the reliability ratio for a variable that should be fixed over time. For example, if the survey–resurvey reliability ratio of an outcome variable is 0.8 (such as in reliably measured labor market outcomes in U.S. datasets), then the maximum attainable R^2 (with any set of controls) is just 0.8.⁵ In what follows we apply the procedure in Oster (2015), using survey–resurvey reliability ratio data, to adjust the coefficient estimates in Tables 3, 4, and 5 in Bellows and Miguel (2009).

2. Impacts of war exposure

2.1. Bounds on R_{max}

The formula in Eq. (2) delivers much wider bounds on the potential bias due to unobservables. In order to obtain more informative bounds, one can gauge the amount of residual variance that comes purely from measurement error by using multiple measures of the same quantity. We adjust coefficients in Bellows and Miguel (2009) using four different approaches: (1) the Bellows and Miguel approach ($R_{max} = 2R^* - R$), (2) the reliability ratio approach developed in this paper, (3) the Oster approach ($R_{max} = \min\{2.2R^*, 1\}$), and (4) the most conservative case ($R_{max} = 1$). Each of these approaches gives different estimates of R_{max} and, therefore, a different estimate of the bounds.

As we need estimates of the reliability ratio in the second approach, it deserves additional explanation. Although there is no information on reliability for the Sierra Leone data used in Bellows and Miguel (2009), Baird et al. (2008) contains an analysis of survey–resurvey data from Kenya.⁶ By focusing on variables where we have reason to believe that

responses should be unchanging over time, the setup in Baird et al. (2008) allows them to estimate how much “noise” there is in reported outcomes in survey data in a rural African sample. We focus on the educational attainment of the respondent's father as a variable that should be fixed over time and where any variation should presumably be due to reporting error.⁷ As reported in Baird et al. (2008), the correlation between stated father's educational attainment across the two survey rounds collected three months apart was moderate: the pairwise correlation coefficient is 0.80. This implies that if father's educational attainment were an outcome variable, the maximum attainable R^2 (with any set of controls) would be just 0.8. Therefore, we use $R_{max} = 0.8$ when obtaining bounds using the reliability ratio approach. We also present the results using the reliability ratio approach using the less conservative assumption of 0.5, roughly based on the data in McKenzie (2012).

2.2. Results

Table 1 first reproduces results for the key coefficient on civil war exposure in Bellows and Miguel (2009) in columns 1 and 2. We do this for each of the three outcome variables presented in Tables 3, 4 and 5 of Bellows and Miguel, namely, community meeting attendance in the last year (the top panel of Table 1); membership in a social group (the middle panel); and membership in a political group (the bottom panel). We then proceed to present five versions of the results in Bellows and Miguel (2009) using different statistical assumptions regarding R_{max} , based on the discussion above. Building on Oster (2015), the result of each extension is an interval of values for the war exposure coefficient that are consistent with the degree of omitted variable bias accommodated by the R_{max} value, the difference between the R^2 in the uncontrolled and controlled regression, and the change in the estimated coefficient across those specifications.

Specifically, columns 1 and 2 present results from the uncontrolled and the controlled regressions, respectively, in Bellows and Miguel (2009), and the corresponding R^2 values. The increase in the R^2 across these two columns captures the amount of variation in the dependent variable that is explained by the observed covariates. This, together with the maximum amount of variation that can be potentially explained (R_{max}), gives rise to an estimate of the lower bound for the interval of coefficients. If this lower bound is greater than zero, it would provide further evidence that the true causal effect of civil war exposure on the outcome is indeed likely to be positive.

In column 3, we again reproduce the analysis in Bellows and Miguel (2009) using the R_{max} implied by their assumption in Eq. (1) laid out above. In this case, the R_{max} is simply the R^2 from the controlled regression plus the difference between the R^2 in the uncontrolled and controlled regressions, i.e., $R_{max} = R^* + (R^* - R)$. The intervals of coefficient values in this case are equivalent to the results presented in the original paper for all three outcome variables, with intervals containing strictly positive values for all three outcome variables.⁸

The next set of results in columns 4 and 5 generate different values of R_{max} using the reliability ratio figure in Baird et al. (2008), namely, $R_{max} = 0.8$ in column 4, and the assumption of $R_{max} = 0.5$ in McKenzie (2012) (both are higher values of R_{max} than that implicitly assumed in BM-09). The interval of coefficient values is broader as a result, increasing from [0.060, 0.065] to [0.048, 0.065] for the community meeting attendance indicator, although even in this case, the lower bound remains greater than zero, providing some evidence of a positive

³ See Cavalcanti et al. (2010), Essaji and Fujiwara (2012), de Brauw and Mueller (2012), Hermes et al. (2012), Sampaio et al. (2013), Breuer and McDermott (2013), Minoiu and Shemyakina (2014), Wong (2014), Jiraporn et al. (2014), and Kosec (2014), among others.

⁴ Oster acknowledges the practical difficulty in establishing an empirically grounded value for R_{max} . The use of prior information about the determinants of observed y to inform the extent of bias has been previously suggested by Altonji et al. (2005a,b, 2008).

⁵ To see this more clearly, note that in the presence of measurement error ε the maximum R^2 in a regression of y on observables x and unobservables \tilde{q} is $R_{max} = \text{Var}(\alpha + x'\beta + \tilde{q}) / \text{Var}(\alpha + x'\beta + \tilde{q} + \varepsilon)$, which corresponds to the reliability ratio.

⁶ We feel the use of estimates from another rural African setting is a useful starting point for illustrating the approach, and can produce suggestive results, although it would clearly be preferable to use data from within the study sample itself.

⁷ In our view, this is an appropriate variable to establish a plausible upper bound on the reliability ratio because it should be stable over time, relatively simple, and is salient in the local context. As noted above, this is just suggestive since it comes from a different setting than the data used in Bellows and Miguel (2009).

⁸ Bellows and Miguel (2009) employ a moderate number of covariates, as described in the note to Table 1.

Download English Version:

<https://daneshyari.com/en/article/969649>

Download Persian Version:

<https://daneshyari.com/article/969649>

[Daneshyari.com](https://daneshyari.com)