causalsens: Sensitivity Analysis for Causal Effects

Matthew Blackwell*

June 18, 2018

1 Introduction

The cornerstone of causal inference are assumptions about the comparability of groups. Our goal, either by randomization or conditioning, is to find groups of observations that are similar on all variables except the variable of interest. And while most studies simply assert this comparability, careful investigations of causal effects want to understand how deviations from this assumption will affect their results. This package provides a way for scholars to assess violation of a key causal assumption, ignorability, by implementing the sensitivity analysis methods proposed in Blackwell (2013).

2 A Selection Bias Approach to Sensitivity Analysis

Blackwell (2013) introduced an approach to sensitivity analysis for causal effects that directly models confounding or selection bias. To review, the goal of causal inference is to estimate the effect of a treatment, A_i on an outcome Y_i. In this case, we will assume a binary treatment, where we call units with $A_i = 1$ the treated group and the units with $A_i = 0$ the control group. We can write causal effects in terms of *potential outcomes* with $Y_i(1)$ being the outcome value that unit i would take if they were treated and $Y_i(0)$ being the outcome value under control. A consistency assumptions connects these potential outcomes to observed outcomes:

$$Y_i = Y_i(a) \quad \text{if } A_i = a. \tag{1}$$

With these definitions in hand, we can define causal effects at the individual level:

$$\tau_i = Y_i(1) - Y_i(0). \tag{2}$$

Unfortunately, we cannot identify or estimate these individual effects without strong assumptions. A ore estimable quantity is the average of these individual effects, called the average treatment effect (ATT):

$$\tau = E[Y_i(1) - Y_i(0)], \tag{3}$$

---

*Department of Political Science, University of Rochester. web: http://www.mattblackwell.org email: m.blackwell@rochester.edu

where the expectation is over units. In some cases, the average effect among the treated group is a more interesting quantity. We can define this parameter as:

$$\tau_{\text{ATT}} = E[Y_i(1) - Y_i(0)|A_i = 1]. \tag{4}$$

In order to obtain estimates of these quantities, we require assumptions that can ensure the comparability of the treatment and control groups. Intuitively, we need this comparability because we need to use the control group to estimate a counterfactual: how the treated would have responded to control? One major approach is to find a set of covariates that ensure the groups are comparable. This is usually called an *ignorability* assumption and can be stated as

$$Y_i(a) \perp\!\!\!\perp A_i | X_i, \tag{5}$$

for all values of the treatment, a. Here, $B \perp\!\!\!\perp C|D$ indicates that the random variables B and C are independent, conditional on a set of variables, D. This assumptions states that the treated and control groups have similar values of the potential outcomes, conditional on a set of covariates. Of course these groups will differ on the observed outcomes if there is a causal effect. But this assumption allows us to use, say, the control group as a proxy for a counterfactual of the treated group. It would be violated if there were unmeasured variable that affected both the treatment variable and the control variable. The goal of an analyst is to collect as much data as possible to make this assumption as plausible as possible.

Of course, with explicit randomization, this ignorability assumption is obviously suspect. In order to quantify the uncertainty over violations of this assumption, Blackwell (2013) describes the *confounding function*. This function describes specific violations of ignorability. We write this function as

$$q(a, x) = E[Y_i(a)|A_i = a, X_i = x] - E[Y_i(a)|A_i = 1 - a, X_i = x]. \tag{6}$$

Here, q describes how the group with $A_i = a$ differs in their potential outcomes from the group with $A_i = 1 - a$ for some value of the covariates. Thus, it captures how the comparability between treatment groups breaks down. If $q = 0$ for all levels of the covariates, then we have ignorability because there is no difference between treatment and control groups in terms of their potential outcomes.

There a number of choices of the confounding function that represent different choices over the type of selection bias. One, which Blackwell (2013) calls *one-sided bias*, writes the confounding function as a function of a single parameter:

$$q(a, x; \alpha) = \alpha(2a - 1). \tag{7}$$

Here, when $\alpha > 0$ the potential outcomes for the treated group are higher on average than the potential outcomes for the control group. In the context of a job training program, this might be because the treated group is more motivated and has greater ability than those who do not apply for the job training program. An alternative parameterization of the confounding function allows for *alignment bias*:

$$q(a, x; \alpha) = \alpha. \tag{8}$$

Here, when $\alpha > 0$, the observed values of the potential outcomes are higher than they would be if the treatment allocation was reversed. This might be because the treatment effect varies over the population and the treatment was targeted to those people for whom the treatment is the

most effective. The choice of which confounding function to use depends on the most plausible violations of ignorability. Each confounding functions tests violations of ignorability specific to that function, but not others. Thus, researchers and critics both have important roles to play in designing sensitivity analyses.

With a confounding function in hand, we can see how the estimated effects changes as we vary the strength of confounding as summarized by the α values above. As Brumback et al. (2004) explains, we can reapply the same outcome model we used to estimate the original effect on a confounding-adjusted outcome:

$$Y_i^q = Y_i - q(A_i, X_i) \Pr[1 - A_i | X_i]. \tag{9}$$

Here, we are essentially subtracting off the bias due to confounding that is implied by the confounding function. Note that this also requires an estimate of the propensity score, or the probability of treatment, conditional on the covariates. Once we have this adjusted outcome, we can rerun our original outcome model on this adjusted outcome to get an estimate of the causal effect under this assumption about confounding. Clearly, when $q = 0$, this will recover the effect estimated under ignorability.

3   An Illustration

An illustration is helpful to see how this selection bias approach to sensitivity analysis works. The data in this case come from a job training program first studied in LaLonde (1986), which has a long history in the causal inference literature on matching estimators. The data comes from the National Supported Work program, which instituted a randomized experiment to evaluate the effectiveness of a job training program. The `lalonde.exp` dataset in the `causalsens` package is the 445 experimental units with complete data on their pre-treatment income. To begin, we load the package and the data:

```
library(causalsens)
data(lalonde.exp)
```

To begin, we fit a regression model to this data, with the outcome of interest, earnings in 1978 (`re78`), as a function of participating in the program (`treat`) and various pre-treatment covariates.

```
ymodel <- lm(re78 ~ treat+age + education + black + hispanic + married +
             nodegree + re74 + re75 + u74 + u75, data = lalonde.exp)
summary(ymodel)

##
## Call:
## lm(formula = re78 ~ treat + age + education + black + hispanic +
##     married + nodegree + re74 + re75 + u74 + u75, data = lalonde.exp)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
##  -9612  -4355  -1572   3054  53119
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.567e+02  3.522e+03   0.073  0.94193
## treat        1.671e+03  6.411e+02   2.606  0.00948 **
## age          5.357e+01  4.581e+01   1.170  0.24284
## education    4.008e+02  2.288e+02   1.751  0.08058 .
## black       -2.037e+03  1.174e+03  -1.736  0.08331 .
## hispanic     4.258e+02  1.565e+03   0.272  0.78562
## married     -1.463e+02  8.823e+02  -0.166  0.86835
## nodegree    -1.518e+01  1.006e+03  -0.015  0.98797
## re74         1.234e-01  8.784e-02   1.405  0.16079
## re75         1.974e-02  1.503e-01   0.131  0.89554
## u74          1.380e+03  1.188e+03   1.162  0.24590
## u75         -1.071e+03  1.025e+03  -1.045  0.29651
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6517 on 433 degrees of freedom
## Multiple R-squared:  0.05822,Adjusted R-squared:  0.0343
## F-statistic: 2.433 on 11 and 433 DF,  p-value: 0.005974
```

From this, we get an estimate of the treatment effect, which is the coefficient on `treat`. Of course, this effect requires an ignorability assumption, so that conditional on the covariates, the treated and control groups are comparable. In this case, the data is experimental in this case so this assumption is very plausible. But we still may want to check out violations of this assumption might affect our results. To do this, we then need a model for the relationship between the treatment variable and the covariates. This is often called the propensity score model because we use predicted values from this model as estimates of the propensity score. In general, this will be a generalized linear model (GLM) of some sort.

```
pmodel <- glm(treat ~ age + education + black + hispanic + married +
            nodegree + re74 + re75 + u74 + u75, data = lalonde.exp,
            family = binomial())
summary(pmodel)

##
## Call:
## glm(formula = treat ~ age + education + black + hispanic + married +
##     nodegree + re74 + re75 + u74 + u75, family = binomial(),
##     data = lalonde.exp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -1.4884  -0.9934  -0.8708    1.2242    1.7403
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.622e+00  1.102e+00   1.472   0.1410
## age           8.276e-03  1.452e-02   0.570   0.5687
## education    -8.282e-02  7.230e-02  -1.145   0.2520
## black        -2.216e-01  3.684e-01  -0.601   0.5476
## hispanic     -8.557e-01  5.128e-01  -1.669   0.0952 .
## married       1.960e-01  2.784e-01   0.704   0.4813
## nodegree     -8.981e-01  3.146e-01  -2.855   0.0043 **
## re74         -4.466e-05  3.010e-05  -1.483   0.1380
## re75          2.924e-05  4.788e-05   0.611   0.5414
## u74          -1.927e-01  3.765e-01  -0.512   0.6088
## u75          -3.369e-01  3.213e-01  -1.048   0.2945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 604.20  on 444  degrees of freedom
## Residual deviance: 584.26  on 434  degrees of freedom
## AIC: 606.26
##
## Number of Fisher Scoring iterations: 4
```

With these two models in hand, we can run the sensitivity analysis by passing these two models to the `causalsens` function, which will estimate the effect of treatment under a number of different assumptions about the direction and magnitude of unmeasured confounding. We also have to choose a confounding function. Here we use the one-sided confounding function, focused on the ATTparameter. `causalsens` provides a number of confounding functions, including the one-sided and alignment bias functions above, but users can specify their own confounding function as well. We also generate a sequence of values for α to determine the range of possible magnitude of ignorability violations.

```
alpha <- seq(-4500, 4500, by = 250)
ll.sens <- causalsens(ymodel, pmodel, ~ age + education, data = lalonde.exp,
                      alpha = alpha, confound = one.sided.att)
```

Once we have run the sensitivity analysis, we can plot the results to get a sense of the sensitivity of our estimates, which we plot in Figure 1.

```
plot(ll.sens, type = "raw", bty = "n")
plot(ll.sens, type = "r.squared", bty = "n")
```
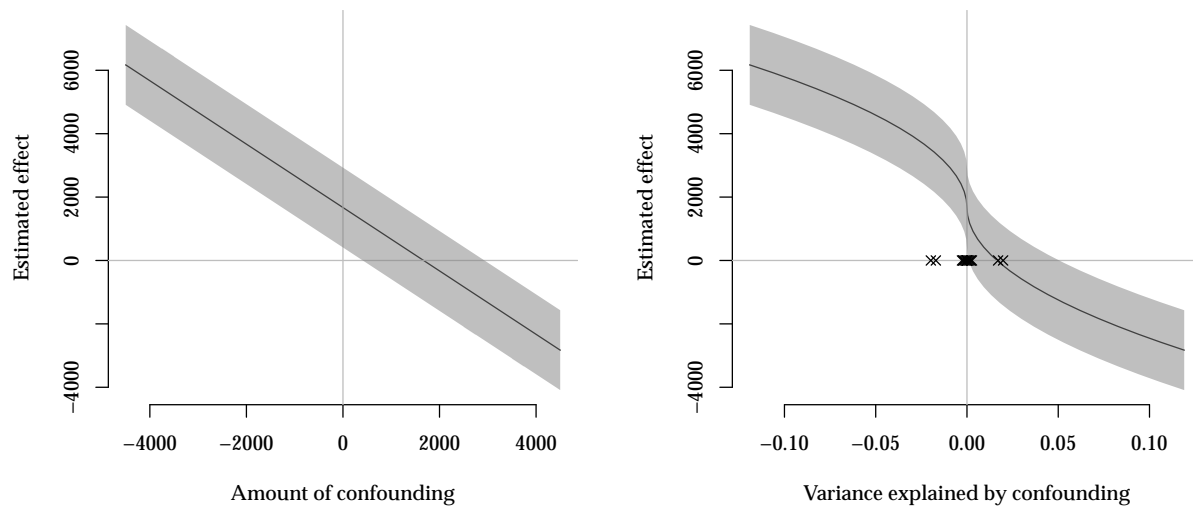
5

Figure 1: Results from LaLonde (1986)

In these plots, we can see the difference between the raw confounding in terms of α and the sensitivity in terms of variance explained by confounding. In addition, we can compare the strength of confounding against the variance explained by the covariates. Obviously, there may be other confounding functions to check and other amounts of confounding to investigate, but this represents a first pass at analyzes the sensitivity of causal estimates in the face of unmeasured confounding.

Bibliography

Blackwell, Matthew. 2013. "A Selection Bias Approach to Sensitivity Analysis for Causal Effects." *Political Analysis* . In Press.

Brumback, Babette A., Miguel A. Hernán, Sebastien J. P. A. Haneuse and James M. Robins. 2004. "Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures." *Statistics in Medicine* 23(5):749–67.

LaLonde, Robert J. 1986. "Evaluating the econometric evaluations of training programs with experimental data." *The American Economic Review* 76(4):604–620.
   URL: *http://www.jstor.org/stable/10.2307/1806062*