KONFOUND: Stata Module to Quantify Robustness of Causal Inferences

Ran Xu[1*], Kenneth A. Frank[2], Spiro J. Maroulis[3], Joshua M. Rosenberg[4]

[1] Department of Industrial and Systems Engineering, Virginia Tech, Falls Church, Virginia, USA
[2] Department of Counseling, Educational Psychology and Special Education, Michigan State University, East Lansing, Michigan, USA
[3] School of Public Affairs, Arizona State University, Tempe, Arizona, USA
[4] College of Education, Health, and Human Sciences, University of Tennessee, Knoxville, Tennessee, USA
[*] Correspondence to ranxu@vt.edu

Abstract. Statistical methods that quantify the discourse about causal inferences in terms of possible sources of biases, also known as “robustness or sensitivity analysis,” are becoming increasingly important to a variety of social-science fields, such as public policy, sociology, and education. A series of recent works by Frank and colleagues on robustness analysis extends earlier methods. We implement these recent developments in STATA. In particular, we provide STATA routines to quantify the % bias necessary to invalidate an inference from a Rubin causal model framework, as well as the robustness of causal inferences in terms of correlations associated with unobserved variables.

1 Introduction

Statistical inferences are often challenged on the basis of uncontrolled bias. There may be bias due to uncontrolled confounding variables or non-random selection into a sample. Methods for sensitivity analysis have been developed to assess the robustness of inferences to various source of bias and inform debate about causal inference. However, most of the previous methods either only accounted for particular sources of bias (e.g., an unobserved variable) or only applied to certain types of data (e.g. categorical treatment variable; Diprete and Gangl 2004; Gill and Robins 2001; Robins 1987; Robins, Rotnisky and Scharfstein, 2000; Rosenbaum 1986, 2002; Scharfstein, 2002; Vanderweele 2010, 2011). In a series of papers, Frank and colleagues (2000, 2004, 2007, 2013) have extended previous work and developed two robustness analysis frameworks. The first uses Rubin's causal model to interpret how much bias there must be to invalidate an inference in terms of replacing observed cases with counterfactual cases or cases from an unsampled population. The second quantifies the robustness of causal inferences in terms of correlations associated with unobserved variables in a regression framework.

In this paper, we introduce the "konfound" package that implements the two robustness analysis methods described above in STATA. Specifically, the konfound command can be used to implement the robustness analysis for the user's model; the mkonfound command can be used to implement the robustness analysis for multiple studies; and the pkonfound command can be used to implement the robustness analysis for single published study. Next, we briefly discuss the foundations of these two methods and describe how to use the "konfound" package in STATA. For a longer introduction of the methods and more technical details readers should refer to Frank (2000), Pan and Frank (2004), Frank and Min (2007), Frank et al. (2008; 2013).

2 Robustness of an Inference

2.1 Impact Threshold for an Omitted Confounding Variable

In observational studies and quasi-experiments, a key concern pertaining to causal inference is the omitted variable bias problem. That is, there are some unobserved confounding variables that may be correlated with both the outcome and the predictor of interest, which will bias the estimates of the model and thus invalidate inferences. To quantify the impact of a confounding variable necessary to alter a statistical inference, Frank (2000) defined the impact of a confounding variable as $r_{x\,cv}r_{y\,cv}$, where $r_{x\,cv}$ is the correlation between the unobserved confound and the predictor of interest and $r_{y\,cv}$, is the correlation between the unobserved confound and the outcome. For example, if the relationship of interest is between Father's occupation (X) and One's own educational attainment (Y), an omitted confounding variable might be Father's education (cv). And the index developed by Frank (2000) allows us to quantify the impact of father's education in terms of its correlation with the predictor father's occupation, as well as its correlation with the outcome – educational attainment. Frank (2000) then shows how strong an omitted confounding variable (cv) would have to be correlated with the predictor

(Father's Occupation, X) as well as the outcome (Educational Attainment, Y) to invalidate an inference of the effect of X on Y.

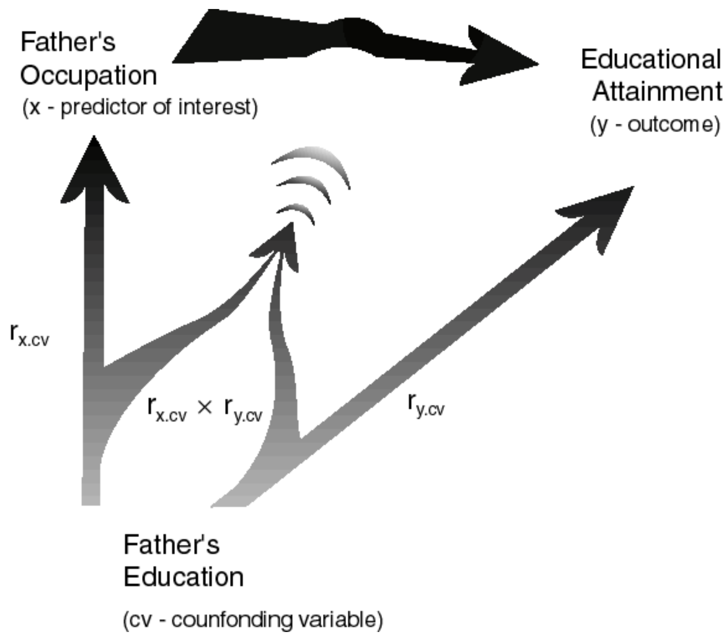

Figure 1: The Impact of a Confounding Variable on a Regression Coefficient

Formally, the calculations follow a partial correlation framework (for more details see Frank (2000), Pan and Frank (2004)). For a bivariate regression

$$Y = \beta_0 + \beta_1 X + e, \qquad (1)$$

the correlation between X and Y – $r_{y.x}$, can be obtained following

$$t(\hat{\beta}_1) = \frac{r_{y.x}}{\sqrt{\frac{1 - r_{y.x}^2}{n - q - 1}}} \qquad (2)$$

Where t is the t-ratio of $\hat{\beta}_1$ to its standard error from an estimated model, n is the sample size, q is number of parameters estimated (other than the intercept) in the model. And the partial correlation between X and Y after controlling for the confounding variable cv – $r_{y.x|cv}$, can be represented as

$$r_{y.x|cv} = \frac{r_{y.x} - r_{y.cv} r_{x.cv}}{\sqrt{1 - r^2{}_{y.cv}}\sqrt{1 - r^2{}_{x.cv}}} \qquad (3)$$

To invalidate an inference we consider the conditions necessary to reduce $r_{xy|cv}$ below a threshold, $r^{\#}$, for making an inference. Here,

$$r^{\#} = \frac{t_{critical}}{\sqrt{(n-q-1)+t^2_{critical}}} \quad (4)$$

Where $t_{critical}$ is decided by the significance level (e.g., for significance level of .05 and a two-sided test with degrees of freedom > 200 $t_{critical}$=1.96).

To calculate the correlations associated with an omitted confounding variable necessary to invalidate an inference, assume the component correlations are equal, $r_{x\cdot cv} = r_{y\cdot cv}$, which generates the largest change from r_{xy} to $r_{xy|cv}$ for a given product = $r_{x\cdot cv}$ x $r_{y\cdot cv}$ =impact (Frank, 2000). Then from Frank (2000):

impact= $r_{x\cdot cv}$ x $r_{y\cdot cv}$ = $r_{x\cdot cv}$ x $r_{x\cdot cv}$ = $r_{y\cdot cv}$ x $r_{y\cdot cv}$. Next, set the partial correlation $r_{y.x|cv}$ equal to the threshold $r^{\#}$

$$r_{y.x|cv} = \frac{r_{y.x} - r_{y.cv} r_{x.cv}}{\sqrt{1-r^2{}_{y.cv}}\sqrt{1-r^2{}_{x.cv}}} = \frac{r_{y.x} - impact}{1 - impact} = r^{\#}, \qquad (5)$$

And then solve for impact: $impact = \frac{r_{y.x} - r^{\#}}{1-|r^{\#}|}$. (6)

Thus, to invalidate the inference, impact of the confounding variable ($r_{x\cdot cv}$ x $r_{y\cdot cv}$) must be greater than

$$\frac{r_{y.x} - r^{\#}}{1-|r^{\#}|}. \qquad (7)$$

Furthermore, the approach also applies to estimated coefficients that are less than their thresholds $r^{\#}$. To alter an inference, (assuming r and $r^{\#}$ take the same sign)

$$\begin{aligned} &if: r > r^{\#} \Rightarrow impact = \frac{r_{y.x} - r^{\#}}{1 - r^{\#}}, \\ &if: r < r^{\#} \Rightarrow impact = \frac{r_{y.x} - r^{\#}}{1 + r^{\#}}. \end{aligned} \qquad (8)$$

Thus, (8) quantifies the smallest impact of confounding variable necessary to invalidate a statistical inference based on the threshold $r^{\#}$.[1]

The above calculations can also be extended to models that control for observed covariates as in multiple regression, where the interpretation of the impact and the correlation can be conditioned on other covariates in the model. In the multiple regression case, the raw component correlation before conditioning on covariates can also be derived, readers who are interested should see Frank (2000). This is available in the konfound module.

2.2 % Bias Necessary to Invalidate an Inference

---

[1] In a case of suppression, smaller products of unequal correlations may produce comparable changes in inference.

A second approach starts by assessing what proportion of an estimate must be due to bias to invalidate an inference (Frank et al., 2013). The proportion is then interpreted in terms of the proportion of observed cases that would have to be replaced with null hypothesis cases to invalidate the inference. These replacement cases can come from counterfactual data as in Rubin's causal model (RCM; Rubin, 1974), or from a population from which observed cases were not sampled. This framework enables researchers to identify a "switch point" (Behn & Vaupel, 1982) where the bias is large enough to undo one's belief about an effect (e.g., from inferring an effect to inferring no effect). Using the switch point, this framework addresses the concerns pertaining to external validity, such as the extent to which the sampling process has to be biased to invalidate the inference, or concerns pertaining to internal validity, such as the extent to which bias due to uncontrolled preexisting differences can invalidate the inference of the treatment effect.

The approach begins when one compares an estimate with a threshold to represent how much bias there must be to switch the inference. For example, consider Figure 2, in which the treatment effect from Hypothetical Study A (with an estimated effect of 6) and B (with an estimated effect of 8) each exceeds the threshold for making an inference of 4. But note that the estimated effect from Study B exceeds the threshold by more than does the estimate from Study A (assuming that the estimates were obtained with similar levels of control for selection bias in the design of the study and similar levels of precision). Therefore we state that the inference from Study B is more robust than that from Study A because a greater proportion of the estimate from Study B must be due to bias to invalidate the inference.

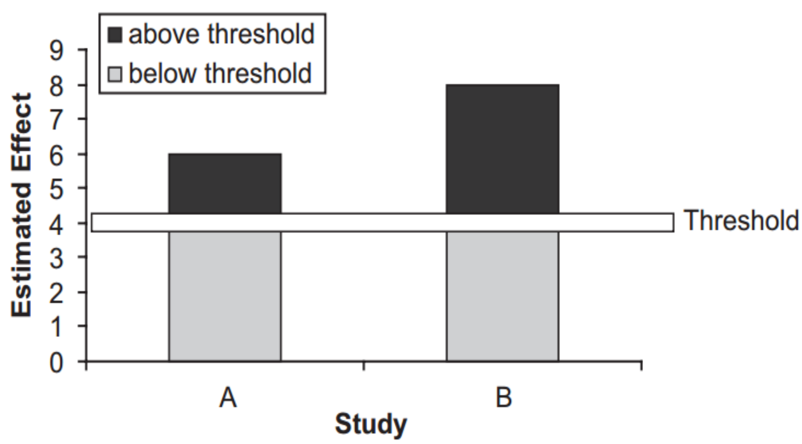

Figure 2: % Bias Necessary to Invalidate an Inference

To formally derive the % bias necessary to invalidate an inference, define a population effect as δ, the estimated effect as $\hat{\delta}$, and the threshold for making an inference as $\delta^{\#}$. An inference about a positive effect is invalid if:

$$\hat{\delta} > \delta^{\#} > \delta . \quad (9)$$

That is, an inference is invalid if the estimate is greater than the threshold while the population value is less than the threshold (a symmetric argument applies for negative effects).

To express how much bias there must be in the estimate to invalidate the inference, we can rewrite the above equation as:

$$\hat{\delta} - \delta > \hat{\delta} - \delta^{\#} > 0 . \quad (10)$$

Defining bias as $bias(\hat{\delta}) = \hat{\delta} - \delta$, to invalidate the inference bias must be larger than the difference between the estimate and the threshold. To express bias as a proportion of the original estimate, we can write:

$$bias(\hat{\delta}) \text{ to invalidate} > \hat{\delta} - \delta^{\#}$$

$$\%bias(\hat{\delta}) \text{ to invalidate} = \frac{bias(\hat{\delta})}{\hat{\delta}} > \frac{\hat{\delta} - \delta^{\#}}{\hat{\delta}} = 1 - \frac{\delta^{\#}}{\hat{\delta}} .$$

For example, in the hypothetical study A in Figure 2, % bias to invalidate the inference = 1 - (4/6) = 1/3. Thus 33% of the estimate would have to be due to bias to invalidate the inference. While in study B, 1 - (4/8) = 50% of the estimate would have to be due to bias to invalidate the inference. Readers should also see Frank et al. (2013), Frank and Min (2007) for other extensions as well as more details of the derivations following the Rubin causal model.

3 The konfound command

3.1 Syntax

```
konfound varlist, [sig(#) nu(#) onetail(#) uncond(#) rep_0(#)
non_li(#)]
```

3.2 Description

konfound calculates the impact of an omitted confounding variable necessary to invalidate/sustain an inference for a regression coefficient from user's model. It also assesses how strong an omitted variable has to be correlated with the outcome and the predictor of interest to invalidate/sustain the inference. After estimating a model (example: linear regression), the user can provide a list of variable names, and konfound will produce the impact of an omitted variable (Frank, 2000) necessary to invalidate/sustain an inference for each variable. The command will also provide the impact table for all observed covariates in the user's previous model. These can be used as a benchmark against which to evaluate the impact of an omitted confounding variable necessary to invalidate an inference.

konfound also calculates how much bias there must be in an estimate to invalidate/sustain an inference from the immediately preceding model. After running a model (such as a linear regression), users can provide the list of variable names, and konfound will calculate the % bias needed to invalidate/sustain the inference for each variable in the variable list. The command will also provide sensitivity plots for those variables that are statistically significant in the user's model.

3.3 Options

sig(#) Significance level of the test; default is 0.05 sig(.05). To change the significance level to .10 use sig(.1).

nu(#) The null hypothesis against which to test the estimate. The default is 0: nu(0).

onetail(#) One-tail or two-tail test; the default is two-tail onetail(0); to change to one-tail use onetail(1).

uncond(#) Calculate the impact and component correlations before or after conditioning on covariates in the model. The default is to calculate the impact and component correlations after conditioning on covariates uncond(0). To change the calculation to before conditioning on covariates use uncond(1).

rep_0(#) For % bias, this controls the effect in the replacement cases; the default is null effect (which may or may not be 0) rep_0(0). When the null hypothesis is not zero, one can still force the replacement cases to have an effect of zero by assigning rep_0(1).

non_li(#) Basis for interpreting % bias to invalidate/sustain an inference for non-linear models (e.g., logit or probit). Default is to use the original coefficient non_li(0); to change the calculation based on average marginal effects use non_li(1).

3.4 Example

To illustrate the use of konfound command, we use two example datasets from Hamilton (1992). The first example comes from a water use survey reported by Hamilton (1983) from Concord, New Hampshire. The outcome of interest is household water usage in the summer of 1981 (`water81`). Independent variables include household water usage in the summer of 1980 (`water80`), household income (`income`), years of education (`educ`), whether the head of the household has retired (`retire`), and number of people in the household in 1980 (`peop80`).

First, we will regress the outcome on all of the independent variables:

```
. use http://stats.idre.ucla.edu/stat/stata/examples/rwg/concord1

(Hamilton (1983))

. reg water81 water80 income educat retire peop80

      Source |       SS           df       MS      Number of obs   =       496

-------------+----------------------------------   F(5, 490)       =    194.82

       Model |   727354309         5   145470862   Prob > F        =    0.0000

    Residual |   365884401       490  746702.858   R-squared       =    0.6653
```

```
-------------+----------------------------------   Adj R-squared   =    0.6619
       Total |  1.0932e+09       495  2208563.05   Root MSE        =    864.12
------------------------------------------------------------------------------
     water81 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     water80 |   .4943149   .0268001    18.44   0.000     .4416577    .5469722
      income |   22.60311   3.502279     6.45   0.000     15.72177    29.48445
      educat |  -44.25776   13.43811    -3.29   0.001     -70.6612   -17.85433
      retire |   155.4727   96.33892     1.61   0.107    -33.81568     344.761
      peop80 |   225.1984   28.70482     7.85   0.000     168.7987    281.5981
       _cons |   299.7437   210.0136     1.43   0.154    -112.8947    712.3821
------------------------------------------------------------------------------
```

The estimated effect of the number of people in the household (`peop80`) is statistically significant (*p* < .001). To quantify the robustness of the inference with respect to omitted variables, or to quantify the percent of the bias necessary to invalidate the current inference, we use the konfound command:

```
. konfound peop80

------------------

% Bias Necessary to Invalidate/Sustain the Inference

For peop80:

To invalidate the inference 74.96% of the estimate would have to be due to bias; to
invalidate the inference 74.96% (372) cases would have to be replaced with cases for
which there is an effect of 0.

------------------

Impact Threshold for an Omitted Confounding Variable

For peop80:

An omitted variable would have to be correlated at 0.519 with the outcome and at 0.519
with the predictor of interest (conditioning on observed covariates) to invalidate an
inference.

These thresholds can be compared with the impacts of observed covariates below.


Observed Impact Table for peop80

+---------------------------------------------------+
|   Zero-Order |  Cor(v,X) |  Cor(v,Y) |    Impact |
|--------------+-----------+-----------+-----------|
```

water80	.5339	.7648	.4083
income	.2845	.4178	.1188
retire	-.3584	-.2731	.0979
educat	.0571	.0404	.0023

Partialled	Cor(v,X)	Cor(v,Y)	Impact
water80	.458	.726	.3325
income	.0714	.2868	.0205
educat	-.0545	-.1567	.0085
retire	-.225	-.0066	.0015

```
X represents peop80, Y represents water81, v represents each covariate.

The first table is based on unconditional correlations. The second table is based on
partialled correlations.
```

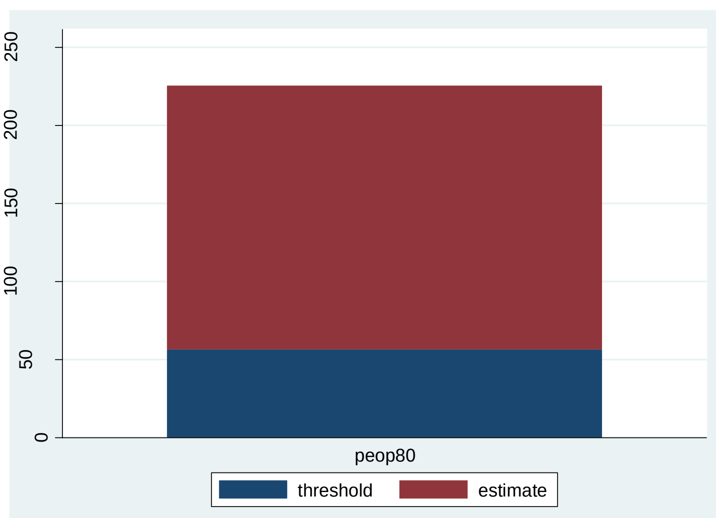

Figure 3. % Bias to Invalidate the Inference for the Effect of `peop80` on `water81`

The first part of the output calculates the % bias needed to invalidate the inference for peop80. As it shows, to invalidate the inference 74.96% (372 cases) would have to be replaced with cases for which there is an effect equal to zero. A graphical illustration is shown in figure 3.

The second part of the output calculates the impact of an omitted variable necessary to invalidate/sustain an inference. First, it shows the impact (.2697) and the component correlations (.519) between the omitted variable and the outcome (`water81`) as well as with the predictor of interest (`peop80`) that are necessary to invalidate the inference, conditional on other covariates. To calculate impact and component correlation before conditioning on covariates, type

```
. konfound peop80, uncond(1)
```

Next, two observed impact tables are shown. For each observed covariate in the model, the first table contains its correlation with the predictor of interest (`peop80`) and with the outcome (`water81`) before conditioning on other covariates; similarly, the second table contains the correlation between each covariate and the predictor of interest (`peop80`) and the outcome (`water81`) after conditioning on other covariates. These two tables can be used to evaluate the robustness of the inference by comparing the impact of the omitted variable necessary to invalidate the inference with the impact of the observed covariates. For example, figure 4 depicts how the partial correlation between `peop80` and `water81` would change when we add an omitted confounding variable in the regression. It shows that the impact of an omitted confounding variable necessary to invalidate the inference (ITCV: red line) would have to be much larger than the impact of income (as well as education and retire). Furthermore, if the impact of an omitted confounding variable equaled that of prior water usage, the inference would be invalid.[2]

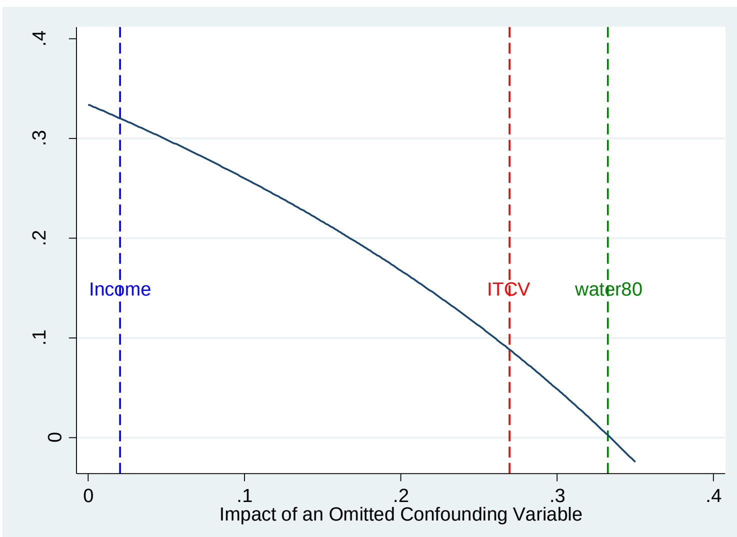

Figure 4. Visualization of the impact of an omitted confounding variable on the partial correlation between `peop80` and `water81`

Finally, there are several things to note:

---

[2] STATA code to reproduce the figure (the partial correlation between `peop80` and `water81` is 0.3341. x represents the impact):

```
twoway function (0.3341 - x)/(1-x), ra(0 0.35)  ytitle("Partial Correlation between peop80 and water81")
/// xtitle("Impact of an Omitted Confounding Variable") xline(0.0205, lcolor(blue) lpattern(dash))
/// xline(0.2697, lcolor(red) lpattern(dash)) xline(0.3325, lcolor(green) lpattern(dash))
/// text( 0.15 0.3325  "water80", color(green) ) text( 0.15 0.269  "ITCV", color(red) ) text( 0.15 0.0205  "Income", color(blue) )
```

1. First-time users of konfound need to install 3 other user-written packages -- moss, indeplist and matsort.
2. Users must run the original regression each time before applying konfound command.
3. Bar graphs are generated only for variables that are statistically significant.
4. Users can evaluate the robustness of inference for multiple variables at the same time; in the previous example to evaluate the robustness of inference of two variables -- `peop80` and `retire`, type

```
. konfound peop80 retire
```

The previous example illustrates how the konfound command can be applied to linear regression models.[3] The next example illustrates how konfound can be applied to non-linear models. Note that for a non-linear model, the impact of an omitted variable necessary to invalidate an inference should not be used, as it is correlation based and thus only applies to linear cases. The % bias to invalidate the inference can still be applied in this case. However, to calculate the % bias to invalidate the inference in a non-linear model, it is recommended to base the calculation on the average marginal effect (also known as average partial effect – Wooldridge, 2010) instead of the original regression coefficient, such that the calculation is robust to different functional form of the model (e.g. logit vs probit).

The next example we use comes from Hamilton (1992), which is from survey data concerning toxic waste in Williamstown, Vermont (Hamilton, 1985). The outcome of interest is a dichotomous variable indicating whether the respondent believed the contaminated school should be closed (`close`). The independent variables include how many years the survey respondent has lived in Williamstown (`lived`), years of education received (`educ`), whether the respondent attended more than two health and safety committee meetings (`hsc`), and whether the respondent is female (`female`).

First, we run a logistic regression using `close` as outcome:

```
. use https://stats.idre.ucla.edu/stat/stata/examples/rwg/toxic, clear

(Hamilton (1985))

. logit close lived educ contam hsc female

Iteration 0:   log likelihood = -104.60578

Iteration 1:   log likelihood = -73.509565

Iteration 2:   log likelihood = -73.284048

Iteration 3:   log likelihood = -73.283842

Iteration 4:   log likelihood = -73.283842
```

[3] konfound is also compatible with random and fixed effects models, as well as linear models with different methods to calculate standard errors, such as robust or clustered standard errors (Wooldridge, 2010). Users who are interested in robustness analysis with alternative methods to calculate the degree of freedom (df) (e.g. using number of level 2 units as the df for level 2 variables) can use pkonfound command or the web-app (konfound-it.com) and manually input the corresponding values for the df.

```
Logistic regression                               Number of obs     =        153
                                                  LR chi2(5)        =      62.64
                                                  Prob > chi2       =     0.0000
Log likelihood = -73.283842                       Pseudo R2         =     0.2994
------------------------------------------------------------------------------
       close |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lived |  -.0433669    .015164    -2.86   0.004    -.0730878    -.013646
        educ |  -.1684151   .0904774    -1.86   0.063    -.3457475    .0089174
      contam |   1.185863   .4641455     2.55   0.011     .2761551    2.095572
         hsc |   2.287901   .4836289     4.73   0.000     1.340006    3.235796
      female |   .7286153   .4422411     1.65   0.099    -.1381614    1.595392
       _cons |   1.223659   1.334176     0.92   0.359    -1.391278    3.838595
------------------------------------------------------------------------------
```

The results show that the estimated effect of `hsc` is statistically significant (p<.001). To calculate the % bias necessary to invalidate the inference for `hsc`, we use konfound with a non-linear model option as below:

```
. konfound hsc, non_li(1)
Average marginal effects                          Number of obs     =        153
Model VCE    : OIM
Expression   : Pr(close), predict()
dy/dx w.r.t. : hsc
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hsc |    .356942   .0532873     6.70   0.000     .2525008    .4613832
------------------------------------------------------------------------------
The following calculation is based on Average Marginal Effect:
------------------
% Bias Necessary to Invalidate/Sustain the Inference
For hsc:
```

```
To invalidate the inference 70.50% of the estimate would have to be due to bias; to
invalidate the inference 70.50% (108) cases would have to be replaced with cases for
which there is an effect of 0.

------------------
```

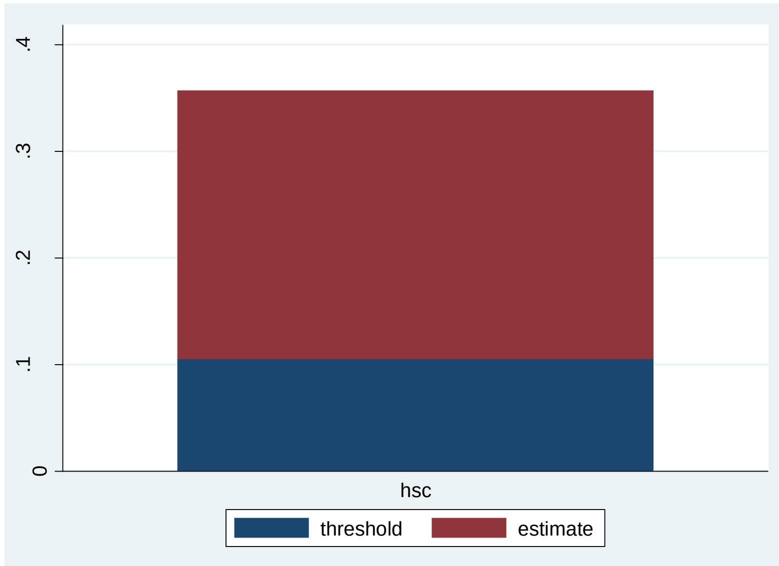

Figure 5. % Bias Necessary to Invalidate the Inference for the Effect of `hsc` on `close`

Results show that to invalidate the inference, 70.5% (108) cases would have to be replaced with cases for which there is an effect of 0. The calculation is based on average marginal effects instead of on the original coefficient; In this case the inference is more robust compared with the calculation based on the original coefficient, which would be 58.23% (89).[4]

4 The mkonfound command

4.1 Syntax

```
mkonfound var1 var2, [sig(#) nu(#) onetail(#) rep_0(#) z_tran(#)]
```

4.2 Description

mkonfound calculates the impact of an omitted confounding variable necessary to invalidate an inference of a regression coefficient for multiple studies. The command also assesses how strong an omitted confounding variable must be correlated with the outcome and with the predictor of interest to invalidate/sustain the inference for each study. Users input two variables: the observed t-ratio and the degrees of freedom for each study. The command mkonfound produces four variables. The first variable is itcv_, indicating the impact of an omitted variable needed to invalidate/sustain the inference.

[4] For some recent developments of the robustness analysis for non-linear models see https://jmichaelrosenberg.shinyapps.io/shinykonfound/

The second variable is r_cv_y, indicating the correlation between the omitted variable and the outcome necessary to invalidate/sustain an inference, conditioning on other covariates. Third variable is r_cv_x, indicating the correlation between the omitted variable and the predictor of interest necessary to invalidate/sustain an inference, conditioning on other covariates. Fourth variable is stat_sig_, indicating if the original regression coefficient is statistically significant; 1 if yes and 0 otherwise.

mkonfound also calculates how much bias there must be in an estimate to invalidate/sustain an inference for multiple studies. The bias necessary to invalidate/sustain an inference is interpreted in terms of sample replacement. The bias necessary to invalidate or sustain an inference is standardized by the size the effect and thus can be compared across studies. Users input two variables: the observed t-ratio and the degrees of freedom in each study. The command mkonfound produces two variables. The first variable is percent_replace, indicating what percent of the original cases must be replaced to invalidate the inference; the second variable is percent_sustain, indicating what percent of the original cases must be replaced to sustain an inference.

4.3 Options

sig(#) Significance level of the test; default is 0.05 sig(.05). To change the significance level to .10 use sig(.1).

nu(#) The null hypothesis against which to test the estimate. The null hypothesis is defined as a correlation, ranging from -1 to 1. The default is 0 nu(0).

onetail(#) One-tail or two-tail test; the default is two-tail onetail(0); to change to one-tail use onetail(1).

rep_0(#) For % bias to invalidate the inference, this specifies the effect in the replacement cases; the default is the null effect (which may or may not be 0) rep_0(0); to force replacing cases with effect of zero use rep_0(1).

z_tran(#) Calculates the % bias to invalidate the inference based on Fisher's z-transformation (only apply to non-zero hypothesis testing); default calculation is based on the original test statistic z_tran(0); to calculate based on Fisher's z use z_tran(1). This option will produce two additional variables based on Fisher's z: percent_replace_z and percent_sustain_z.

4.4 Example

To illustrate the use of mkonfound command, we generate t-ratios and degrees of freedom for 10 studies

```
. clear

. set obs 10

number of observations (_N) was 0, now 10

. drawnorm t,mean(1) sd(3)

. gen df=int(200*uniform())
```

```
. list

     +-----------------+
     |         t    df |
     |-----------------|
  1. |  7.076763    97 |
  2. |  4.127893   174 |
  3. |  1.893137   153 |
  4. | -4.166395    50 |
  5. | -1.187599    33 |
  6. |  3.585478   148 |
  7. |   .281938   196 |
  8. |  2.549647   145 |
  9. | -4.436048   180 |
 10. | -2.045373    52 |
     +-----------------+
```

Next, we calculate the % bias necessary to invalidate/sustain the inference and impact threshold for omitted variables using mkonfound command

```
. mkonfound t df

. list

    +--------------------------------------------------------------------------------+
    |         t    df        itcv_   r_cv_y   r_cv_x   stat_s~_   percen~e   percen~n |
    |--------------------------------------------------------------------------------|
 1. |  7.076763    97      .482843     .695     .695          1      66.15          . |
 2. |  4.127893   174     .1773619     .421     .421          1      50.45          . |
 3. |  1.893137   153    -.0055674     .075    -.075          0          .       4.08 |
 4. | -4.166395    50    -.3253992      .57     -.57          1      46.17          . |
 5. | -1.187599    33     .0993485     .315     .315          0          .      39.36 |
 6. |  3.585478   148     .1462456     .382     .382          1      43.28          . |
 7. |   .281938   196    -.1049747     .324    -.324          0          .      85.57 |
 8. |  2.549647   145     .0541095     .233     .233          1      21.81          . |
```

```
 9. | -4.436048   180   -.1976479     .445    -.445          1       53.65          . |
10. | -2.045373    52   -.0066187     .081    -.081          1        1.76          . |
    +-------------------------------------------------------------------------------+
```

To calculate the impact threshold for omitted variables, mkonfound generates four variables for each study. The first variable is itcv_, indicating the impact of an omitted variable necessary to invalidate/sustain an inference. The second variable is r_cv_y, indicating the correlation between the omitted variable and the outcome necessary to invalidate/sustain an inference, conditioning on other covariates in the model. The third variable is r_cv_x, indicating the correlation between the omitted variable and the predictor of interest necessary to invalidate/sustain an inference, conditioning on other covariates in the model. The fourth variable is stat_sig_, indicating if the original regression coefficient is statistically significant; 1 if yes and 0 otherwise.

To calculate % bias necessary to invalidate/sustain the inference, mkonfound generates two variables for each study as in last two columns: percent_replace (percen~e) and percent_sustain (percen~n). For studies that are statistically significant, percent_replace shows the % of cases that need to be replaced with cases with an effect of 0 to invalidate the inference. For studies that are not statistically significant, percent_sustain shows the % of 0 effect cases that need to be replaced with cases that have an effect at the threshold of inference to sustain the inference.

5 The pkonfound command

5.1 Syntax

```
pkonfound # # # #, [sig(#) nu(#) onetail(#) rep_0(#)]
```

5.2 Description

pkonfound takes the user's input of numerical values (e.g., from a published study) and calculates (1) the % bias in an estimate necessary to invalidate/sustain an inference. The % bias necessary to invalidate/sustain an inference is interpreted in terms of sample replacement; (2) the impact of an omitted confounding variable necessary to invalidate/sustain an inference for a regression coefficient. It also assesses how strong an omitted variable must be correlated with the outcome and with the predictor of interest to invalidate/sustain the inference.[5]

Four numbers must be input. The first number is the estimated value of the effect (e.g., the estimated regression coefficient); the second number is the standard error of the estimated effect (regression coefficient); the third number is the sample size; the fourth number is the number of covariates in the model.

[5] Users who are interested in using pkonfound can also refer to the web-app (konfound-it.com).

5.3 Options

sig(#) Significance level of the test; default is 0.05 sig(.05). To change the significance level to .10 use sig(.1).

nu(#) The null hypothesis against which to test the estimate. The default is 0: nu(0).

onetail(#) One-tail or two-tail test; the default is two-tail onetail(0); to change to one-tail use onetail(1).

rep_0(#) For % bias to invalidate the inference, this controls the effect in the replacement cases; the default is the null effect (which may or may not be 0) rep_0(0); When the null hypothesis is not zero, one can still force the replacement cases to have an effect of zero by assigning rep_0(1).

5.4 Example

To illustrate the use of pkonfound command, assume that in a published study the estimated effect is 10, the standard error of the estimate is 2, the sample size is 100, and the number of covariates is 4. To calculate the % bias necessary to invalidate the inference and the impact threshold for the omitted variable, users type:

```
. pkonfound 10 2 100 4

------------------

Impact Threshold for an Omitted Confounding Variable

An omitted variable would have to be correlated at 0.569 with the outcome and at 0.569
with the predictor of interest (conditioning on observed covariates) to invalidate an
inference.

Correspondingly the minimum impact to invalidate an inference for a null hypothesis of
0 effect is 0.569 x 0.569 = 0.3240

------------------

% Bias Necessary to Invalidate/Sustain the Inference

To invalidate the inference 60.29% of the estimate would have to be due to bias; to
invalidate the inference 60.29% (60) cases would have to be replaced with cases for
which there is an effect of 0.

Note:

For non-linear models, the impact threshold for an omitted confounding variable should
not be used.

The % bias calculation is based on the original coefficient, compare with the use of
average marginal effects as in the [konfound] command.
```

Similar to the konfound command, the results are divided into two parts. The first part of the output shows the impact threshold and component correlations for the omitted confounding variable necessary to invalidate the inference. The second part of the output shows the % bias necessary to invalidate the inference.

6 Examples of Publishable Write-ups

To facilitate the interpretation of the robustness analysis, here we provide some examples of publishable write-ups for correlation based as well as case replacement based robustness analysis. The example of correlation based approach comes from Frank et al. (2008), where the main focus is on whether national board of professional teaching standards (NBPTS) certified teachers provide more instructional help to other teachers:

> While we may be close to exhausting our ability to reduce bias that can be attributed to confounding variables measured in our data, we use Frank's (2000) indices to quantify how much the impact of an *unobserved* confound must be to invalidate the inference that NBPTS certification affects the number of others a teacher helps with instructional matters. Here we base the analysis on the estimate and inference using propensity weighting to estimate the EOTM, the most conservative of the estimates that used the full sample and controlled for covariates.
>
> Given the sample size of 1,131, the threshold for statistical significance, $r^{\#}$, is.058. The observed t-ratio of 4.13 (4.13=.57/.138) translates to a correlation between being and NBPTS and number of others helped of r =.122. From (5), the impact of an unmeasured confound would have to be greater than .068 to invalidate our inference; the impact threshold = $(r-r^{\#})/(1-|r^{\#}|)$=(.122-.058)/(1-|.058|)=.068. Correspondingly, each component correlation would have to be equal to .26. Thus to invalidate the inference that NBPTS certification increases the help provided by a teacher, a confounding variable would have to be correlated with NBPTS certification at 0.26 *and* with help provided at 0.26. These are moderate correlations by social science standards (Cohen & Cohen, 1983). Moreover, these are zero-order correlations, assuming that the unmeasured confound is uncorrelated with the measured covariates (see Frank, 2000). The relevant partial correlations from which the impact of an unobserved confound would be constructed would be smaller than the zero-order correlations because of correlations with existing covariates.
>
> Though the magnitude of the impact threshold for an unmeasured variable can be interpreted in terms of general findings in the social sciences, it is also helpful to compare the threshold to the impacts of measured covariates. The extent to which a teacher believes leadership will enhance teaching has the strongest impact of the measured covariates. Its impact on the coefficient for NBPTS certified teachers on help provided is 0.011 which is the product of the correlation with being an NBPTS certified teacher (0.17) and the correlation with number of other teachers helped (0.06). Thus the impact of an unmeasured confound necessary to invalidate the inference of .068 would have to be more than six times greater than the strongest impact of the measured covariates, .011, to invalidate the inference that NBPTS certification affects the number of colleagues a teacher helps with instruction.

An example of the case replacement based approach comes from Saw et al. (2017), where the focus is the impact of being labeled as a Persistently Lowest Achieving (PLA) School on students' academic performances:

> To inform policy debates and theoretical interpretations of the causal effects of the PLA list, it is useful to quantify the discourse about the robustness of the inferences in this study. We quantify how much bias there must be in our RD estimates to invalidate inferences in terms of replacement data, 16 focusing only on the positive PLA list effects on the average of students' scale scores in writing and the percentage of students who met proficiency level in social studies. As shown in table 5, to invalidate our causal inference of the PLA list effects on the average of students' scale scores in writing, we would need to replace about 25% to 32% of our PLA schools with school samples for which there is no effect of being on the list. These 17 to 22 replacement schools could represent populations not directly in our sample, such as schools from outside of the selected bandwidth. Additionally, to invalidate the inference of an effect of assignment to the PLA list on social studies achievement, we would have to replace 6% to 8.6% of schools with schools in which there was no effect of being on the PLA list.

More write-up examples from other fields can be found in Appendix B.

7 References

Behn, R.D. and Vaupel, J.W., 1982. *Quick analysis for busy decision makers* (p. 3). New York: Basic Books.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S., 1983. Applied multiple regression for the behavioral sciences. Laurence Erlbaum, Hillsdale, NJ.

DiPrete, T.A. and Gangl, M., 2004. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology*, *34*(1), 271-310.

Frank, K.A. 2000. Impact of a Confounding Variable on the Inference of a Regression Coefficient. *Sociological Methods and Research, 29*(2), 147-194

Frank, K. A. and Min, K. 2007. Indices of Robustness for Sample Representation. *Sociological Methodology, 37,* 349-392.

Frank, K.A., Gary Sykes, Dorothea Anagnostopoulos, Marisa Cannata, Linda Chard, Ann Krause, Raven McCrory. 2008. Extended Influence: National Board Certified Teachers as Help Providers.*Education, Evaluation, and Policy Analysis, 30*(1): 3-3

Frank, K.A., Maroulis, S., Duong, M., and Kelcey, B. 2013. What would it take to Change an Inference?: Using Rubin's Causal Model to Interpret the Robustness of Causal Inferences. *Education, Evaluation and Policy Analysis, 35*, 437-460.

Gill, R.D. and Robins, J.M., 2001. Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 1785-1811.

Hamilton, L.C., 1983. Saving water: a causal model of household conservation. *Sociological Perspectives*, *26*(4), 355-374.

Hamilton, L.C., 1985. Concern about toxic wastes: Three demographic predictors. *Sociological Perspectives*, *28*(4), 463-486.

Hamilton, L.C., 1992. Regression with graphics: A second course in applied statistics. Brooks/Cole Pub. Co.

Pan, W., and Frank, K.A. 2004. An Approximation to the Distribution of the Product of Two Dependent Correlation Coefficients. *Journal of Statistical Computation and Simulation, 74,* 419-443

Pan, W., and Frank, K.A., 2004. A probability index of the robustness of a causal inference. *Journal of Educational and Behavioral Statistics, 28*, 315-337.

Robins, J. 1987. A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods. *Journal of Chronic Diseases 40(*2):1395–1615.

Robins, J., A. Rotnisky, and D. Scharfstein. 2000. Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. Pp. 1–95 in Statistical Models in Epidemiology, the Environment and Clinical Trials (The IMA Volumes in Mathematics and Its Applications), edited by E. Halloran and D. Berry. New York: Springer-Verlag.

Rosenbaum, P. R. 1986. Dropping Out of High School in the United States: An Observational Study. *Journal of Educational Statistics 11*(3):207–24.

Rosenbaum, P. R. 2002. Attribution Effects to Treatment in Matched Observational Studies. *Journal of the American Statistical Association, 97*, 183–92.

Rubin, D. B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology, 66,* 688–701.

Saw, G., Schneider, B., Frank, K., Chen, I.C., Keesler, V. and Martineau, J., 2017. The Impact of Being Labeled as a Persistently Lowest Achieving School: Regression Discontinuity Evidence on Consequential School Labeling. *American Journal of Education, 123*(4), 585-613.

Scharfstein, D. A. I. 2002. Generalized Additive Selection Models for the Analysis of Studies with Potentially Nonignorable Missing Data. *Biometrics 59*, 601–13.

VanderWeele, T.J., 2010. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology (Cambridge, Mass.)*, *21*(4), 540.

VanderWeele, T.J. and Arah, O.A., 2011. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology (Cambridge, Mass.)*, *22*(1), 42-52.

Wooldridge, J.M., 2010. *Econometric analysis of cross section and panel data*. MIT press.

Technical Appendix A:

Maximizing the Impact of an Omitted Variable

Formula for partial correlation can be represented as (also applies to regression):

$$r_{x\cdot y|cv} = \frac{r_{x\cdot y} - r_{\mathrm{cv}\cdot y} \times r_{\mathrm{cv}\cdot x}}{\sqrt{1-r_{\mathrm{cv}\cdot y}^2}\sqrt{1-r_{\mathrm{cv}\cdot x}^2}}$$

We want to minimize the function of partial correlation given (assuming all terms are positive):

$$k = r_{\mathrm{cv}\cdot y} \times r_{\mathrm{cv}\cdot x} \Rightarrow \; r_{\mathrm{cv}\cdot y} = \frac{k}{r_{\mathrm{cv}\cdot x}}$$

$$\Rightarrow r_{x\cdot y|cv} = \frac{r_{x\cdot y} - \frac{k}{r_{\mathrm{cv}\cdot x}} \times r_{\mathrm{cv}\cdot x}}{\sqrt{\left(1-\left(\frac{k}{r_{\mathrm{cv}\cdot x}}\right)^2\right)\left(1-r_{\mathrm{cv}\cdot x}^2\right)}} = \frac{r_{x\cdot y} - k}{\sqrt{1-\left(\frac{k}{r_{\mathrm{cv}\cdot x}}\right)^2 - r_{\mathrm{cv}\cdot x}^2 - \left(\frac{k^2}{r_{\mathrm{cv}\cdot x}^2}\right)\left(-r_{\mathrm{cv}\cdot x}^2\right)}} = \frac{r_{x\cdot y} - k}{\sqrt{1-\left(\frac{k}{r_{\mathrm{cv}\cdot x}}\right)^2 - r_{\mathrm{cv}\cdot x}^2 + k^2}}$$

To maximize the impact, we want to minimize the function. This occurs when the denominator is maximized:

$$\frac{d}{dr_{\mathrm{cv}\cdot x}}\left(1-\frac{k^2}{r_{\mathrm{cv}\cdot x}^2} - r_{\mathrm{cv}\cdot x}^2 + k^2\right) = 0$$

$$\Rightarrow 2\frac{k^2}{r_{\mathrm{cv}\cdot x}^3} - 2r_{\mathrm{cv}\cdot x} = 0$$

$$\Rightarrow r_{\mathrm{cv}\cdot x} = \frac{k^2}{r_{\mathrm{cv}\cdot x}^3} \Rightarrow r_{\mathrm{cv}\cdot x}^4 = k^2 \Rightarrow r_{\mathrm{cv}\cdot x} = \pm\sqrt{k},\; r_{\mathrm{cv}\cdot \mathrm{y}} = \pm\sqrt{k}$$

The positive term is used when k is positive. Otherwise the negative root is used for suppression.

Note the second derivative is

$$\frac{d}{dr_{\mathrm{cv}\cdot x}}\left(2\frac{k^2}{r_{\mathrm{cv}\cdot x}^3}-2r_{\mathrm{cv}\cdot x}\right)=-6\frac{k^2}{r_{\mathrm{cv}\cdot x}^4}-2$$

which is less than zero when:

$$-6\frac{k^2}{r_{\mathrm{cv}\cdot x}^4}-2<0\Rightarrow\frac{k^2}{r_{\mathrm{cv}\cdot x}^4}>-\frac{1}{3}$$

This condition always holds, so first derivative above defines a maximum.

Technical Appendix B:

Examples of applications of indices for quantifying the robustness of causal inferences

Correlation framework

Business:

Badertscher, B. A., Katz, S. P., & Rego, S. O. (2013). The separation of ownership and control and corporate tax avoidance. *Journal of Accounting and Economics, 56*(2), 228-250.

Gamache, D. L., McNamara, G., Graffin, S. D., Kiley, J. T., Haleblian, J., & Devers, C. E. (forthcoming). Impression offsetting as an early warning signal of low CEO confidence in acquisitions. *Academy of Management Journal*, (ja).

Harmen, Derek (May 2018). When the Fed Speaks: Arguments, Emotions, and the Microfoundations of Institutions. Administrative Science Quarterly

Hill, A. D., Recendes, T., & Ridge, J. W (forthcoming). Second-order effects of CEO characteristics: How rivals' perceptions of CEOs as submissive and provocative precipitate competitive attacks. *Strategic Management Journal*.

Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics, 49*(3), 186-205.

Quinn, P. J. (2018). Shifting corporate culture: executive stock ownership plan adoptions and incentives to meet or just beat analysts' expectations. *Review of Accounting Studies, 23*(2), 654-685.

Westphal, J. D., & Zhu, D. H. (forthcoming) Under the radar: How firms manage competitive uncertainty by appointing friends of other ceos to their boards. *Strategic Management Journal.* https://www.researchgate.net/publication/328110533_Under_the_radar_How_firms_manage_competitive_uncertainty_by_appointing_friends_of_other_ceos_to_their_boards

Economics

Barontini, R., & Bozzi, S. (2018). Family firm heterogeneity and CEO compensation in Continental Europe. *Journal of Economics and Business*.

Education

Crosnoe, R., & Cooper, C. E. (2010). Economically disadvantaged children's transitions into elementary school: Linking family processes, school contexts, and educational policy. *American Educational Research Journal, 47*(2), 258-291.

Frank, K.A., Gary Sykes, Dorothea Anagnostopoulos, Marisa Cannata, Linda Chard, Ann Krause, Raven McCrory. 2008. Extended Influence: National Board Certified Teachers as Help Providers. *Education, Evaluation, and Policy Analysis, 30*(1): 3-30.

Sun, M., Penuel, W. R., Frank, K. A., Gallagher, H. A., & Youngs, P. (2013). Shaping professional development to promote the diffusion of instructional expertise among teachers. *Educational Evaluation and Policy Analysis, 35(*3), 344-369.

Medicine

Carbonell, F., Bellec, P., & Shmuel, A. (2014). Quantification of the impact of a confounding variable on functional connectivity confirms anti-correlated networks in the resting-state. *Neuroimage, 86,* 343-353.

Political Science

Dietz, T., Frank, K. A., Whitley, C. T., Kelly, J., & Kelly, R. (2015). Political influences on greenhouse gas emissions from US states. *Proceedings of the National Academy of Sciences, 112*(27), 8254-8259.

Psychology

Crosnoe, R. (2009). Family–school connections and the transitions of low-income youths and english language learners from middle school to high school. *Developmental Psychology, 45*(4), 1061.

Pontifex, M. B., Kamijo, K., Scudder, M. R., Raine, L. B., Khan, N. A., Hemrick, B., ... & Hillman, C. H. (2014). V. The differential association of adiposity and fitness with cognitive control in preadolescent children. *Monographs of the Society for Research in Child Development, 79*(4), 72-92.

Public Policy

Frank, K.A., Penuel, W.R., Krause, A (2015). What is a “Good” Social Network for Policy Implementation?: The Flow of Know-How for Organizational Change. *Journal of Policy Analysis and Management. 34*(2), pp. 378-402.

Frank, K., Chen, I. C., Lee, Y., Kalafatis, S., Chen, T., Lo, Y. J., & Lemos, M. C. (2012). Network Location and Policy ‐ Oriented Behavior: An Analysis of Two ‐ Mode Networks of Coauthored Documents Concerning Climate Change in the Great Lakes Region. *Policy Studies Journal, 40*(3), 492-515.

Newman, S., Holupka, C. S., & Harkness, J. (2009). The long ‐ term effects of housing assistance on work and welfare. *Journal of Policy Analysis and Management, 28*(1), 81-101.

Strunk, K. O., Goldhaber, D., Knight, D. S., & Brown, N. (2018). Are There Hidden Costs Associated With Conducting Layoffs? The Impact of Reduction-in-Force and Layoff Notices on Teacher Effectiveness. *Journal of Policy Analysis and Management*, *37*(4), 755-782.

Sociology

Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high schools. *American Sociological Review, 74*(5), 709-730.

Frisco, M. L., Muller, C., & Frank, K. (2007). Parents’ union dissolution and adolescents’ school performance: Comparing methodological approaches. *Journal of Marriage and Family, 69*(3), 721-741.

Gaddis, S. M. (2012). What's in a relationship? An examination of social capital, race and class in mentoring relationships. *Social Forces, 90*(4), 1237-1269.

Lee, J. C., & Klugman, J. (2013). Latino school concentration and academic performance among Latino children. *Social Science Quarterly, 94*(4), 977-1015.

Case Replacement Framework

Environmental Sociology

Asensio, O. I., & Delmas, M. A. (2015). Nonprice incentives and energy conservation. *Proceedings of the National Academy of Sciences, 112*(6), E510-E515.

Masuda, Y. J., Liu, Y., Reddy, S. M., Frank, K. A., Burford, K., Fisher, J. R., & Montambault, J. (2018). Innovation diffusion within large environmental NGOs through informal network agents. *Nature Sustainability, 1*(4), 190.

Education

Allensworth, E. M., Moore, P. T., Sartain, L., & de la Torre, M. (2017). The educational benefits of attending higher performing schools: Evidence from Chicago high schools. *Educational Evaluation and Policy Analysis, 39*(2), 175-197.

Beymer, P. N., Rosenberg, J. M., Schmidt, J. A., & Naftzger, N. J. (2018). Examining Relationships among Choice, Affect, and Engagement in Summer STEM Programs. *Journal of Youth and Adolescence, 47*(6), 1-14.

Jacob, B., Dynarski, S., Frank, K., & Schneider, B. (2017). Are Expectations Alone Enough? Estimating the Effect of a Mandatory College-Prep Curriculum in Michigan. *Educational Evaluation and Policy Analysis, 39*(2), 333-360.

Saw, G., Schneider, B., Frank, K., Chen, I. C., Keesler, V., & Martineau, J. (2017). The Impact of Being Labeled as a Persistently Lowest Achieving School: Regression Discontinuity Evidence on Consequential School Labeling. *American Journal of Education, 123*(4), 585-613.

Wilhelm, A. G., Chen, I. C., Smith, T. M., & Frank, K. A. (2016). Selecting expertise in context: Middle school mathematics teachers' selection of new sources of instructional advice. *American Educational Research Journal, 53*(3), 456-491.

Public Policy

Frank, K.A., Penuel, W.R., Krause, A (2015). What is a "Good" Social Network for Policy Implementation?: The Flow of Know-How for Organizational Change. *Journal of Policy Analysis and Management. 34*(2), pp.378-402.

Sociology

Ressler, R. W., Smith, C., Cavanagh, S., & Crosnoe, R. (2017). Mothers' union statuses and their involvement in young children's schooling. *Journal of Marriage and Family, 79*(1), 94-109.

Foundational Methodological Work

Case replacement

Frank, K.A., Maroulis, S., Duong, M., and Kelcey, B. (2013). What would it take to Change an Inference?: Using Rubin's Causal Model to Interpret the Robustness of Causal Inferences. *Education, Evaluation and Policy Analysis., 35,* 437-460.

Frank, K. A. and Min, K. (2007). Indices of Robustness for Sample Representation. *Sociological Methodology, 37*, 349-392.

Correlational

Frank, K.A., Gary Sykes, Dorothea Anagnostopoulos, Marisa Cannata, Linda Chard, Ann Krause, Raven McCrory. (2008). Extended Influence: National Board Certified Teachers as Help Providers. *Education, Evaluation, and Policy Analysis, 30*(1): 3-30.

Frank, K. (2000). Impact of a Confounding Variable on the Inference of a Regression Coefficient. *Sociological Methods and Research, 29*(2), 147-194