Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 23 Number 2, February 2018 ISSN 1531-7714

Using Instrumental Variable Estimation to Evaluate Randomized Experiments with Imperfect Compliance

Francis L. Huang, *University of Missouri*

Among econometricians, instrumental variable (IV) estimation is a commonly used technique to estimate the causal effect of a particular variable on a specified outcome. However, among applied researchers in the social sciences, IV estimation may not be well understood. Although there are several IV estimation primers from different fields, most manuscripts are not readily accessible by researchers who may only be familiar with regression-based techniques. The manuscript provides a conceptual framework of why and how IV works in the context of evaluating treatment effects using randomized evaluations. I discuss the issue of imperfect treatment compliance, explain the logic of IV estimation, provide a sample dataset, and syntax for conducting IV analysis using R. A goal of the current manuscript is to demystify the use of IV estimation and make evaluation studies that use this technique more readily understood by researchers.

Among econometricians, instrumental variable (IV) estimation is a commonly used technique to estimate the causal effect of a particular variable on a specified outcome. IV estimation has been described as the "most powerful weapon" in an economist's arsenal of statistical tools (Angrist & Pischke, 2008, p. 114). However, among applied researchers in the social sciences, IV estimation may not be well understood. This lack of understanding may be evident by the number of primers on IV estimation in diverse fields such as developmental psychology (Gennetian, Magnuson, & Morris, 2008), education (Pokropek, 2016), social work (Rose & Stone, 2011), medicine (Baiocchi, Cheng, & Small, 2014), political science (Sovey & Green, 2011), and criminology (Angrist, 2006). However, most of the articles, though targeted towards novice users of the technique, are often laden with various types of notation, equations, and proofs that may get in the way of developing an intuitive understanding of IVs. Although IV estimation can be used with certain types of observational, nonexperimental data in order to establish some form of causality, the focus of this manuscript is to provide a conceptual framework of how IV works in the context of evaluating treatment effects using a randomized experiment (RE) or a randomized control trial (RCT). I discuss the issue of imperfect compliance in experiments, explain the logic of IV estimation, provide an example, and share syntax for conducting IV analysis using R (R Core Team, 2017). In addition, I clarify some terms that are often encountered when reading articles that make use of IV estimation with the goal of making these articles more readily comprehensible to a broader audience who may know basic regression but are unfamiliar with IV estimation.

Noncompliance in Treatment Assignment

RCTs are considered the 'gold standard' in evaluation research (Sullivan, 2011; Ye, Beyene, Browne, & Thabane, 2014). When conducting an impact evaluation using an RCT or an experiment with random assignment, participants are randomly assigned to either treatment or control conditions. If all participants follow their treatment assignment perfectly (i.e., only those assigned to the treatment received the treatment and those assigned to control did not), only a t-test on the mean differences in outcomes between groups would be needed to obtain the causal effect (Murnane & Willett, 2011). The random assignment assumes that participants are

approximately equivalent at baseline on observed/measured (e.g., gender, GPA) and unobserved/unmeasured (e.g., motivation) characteristics and that any differences in outcomes are due to the treatment.

However, at times, participants may not always follow their treatment assignment. For example, in an education context, in order to evaluate the effectiveness of charter schools, lotteries are often used when there are more students who want to enroll in a school than there are available seats (Angrist, Dynarski, Kane, Pathak, & Walters, 2010). In the presence of oversubscribed seats, spaces are raffled off in a lottery to a pool of interested participants. The lottery is a form of random assignment which allows comparisons to be made based on the outcomes of lottery winners (who were offered a seat at the school) vs. lottery losers (who were not offered a seat at the school). The attendance of the charter school in this case is the treatment and an important distinction to be made is that there is only an offer to attend the school. After an offer is made, parents may then elect to enroll their child at that school. Researchers cannot force offered participants to take up the treatment which is why at times, these are also referred to as randomized encouragement (West et al., 2008) or promotion (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2016) designs. Based on a study of charter school evaluations using lotteries, 78% of students offered a seat took the seat and at the same time, 15% of students who were not offered a seat still wound up attending a charter school (Clark, Gleason, Tuttle, & Silverberg, 2015).

The noncompliance of treatment assignment occurs in various fields. Individuals provided housing vouchers to move from high to low poverty neighborhoods may choose not relocate (Leventhal & Brooks-Gunn, 2003). Police officers who were randomized to separate domestic assault suspects (the treatment) wound up arresting them instead (Sherman & Berk, 1984). Subjects in medical trials may not always take the medicine prescribed to them and those who do not get the experimental treatment may find some other alternative medication (Sussman & Hayward, 2010). Teachers assigned to attend training to help improve children's outcomes may not show up.

Although the offer or assignment of treatment is random (A = 1, treatment offered; A = 0, treatment not offered), the actual take up of the treatment (T = 1, treatment received; T = 0, no treatment received) may not be[1]. Continuing from the charter school example, parents of students who were not offered seats but still wound up enrolling their child in a charter school (whether the study school or another nearby charter school) may possess some extra motivation compared to other parents who complied with their control status and sent their kids to their local public school. In such a case, simply comparing the outcomes of those who actually received the treatment with those who did not may produce biased results.

Estimating the Intention-to-Treat Effect

Given imperfect compliance and the potential for biased results, one common strategy for estimating unbiased effects of the treatment offer is to regress the outcome on treatment assignment and not actual take up. The causal estimand (i.e., the quantity that defines the causal effect for a particular population) in this case is referred to as intention-to-treat (ITT) or the ITT effect (Hollis & Campbell, 1999). The ITT principle adheres to the original random assignment used in the experiment though ITT effects are often diluted because of treatment noncompliance (Gupta, 2011). As compliance rates go up, the dilution effect decreases. In an IV framework, the ITT effect is at times referred to as a "reduced form" equation (Angrist, 2006).

Understanding Compliers and Non-compliers

The ITT effect though is based solely on treatment assignment (A = 1 vs. A = 0) and not if the treatment was actually received or delivered (T = 1 vs. T = 0). As evaluators though, the effect of interest may be the impact of the treatment on those who actually complied with the treatment assignment. In order to estimate the effect on compliers, an understanding of the different compliance types is required.

In a population of individuals, four conceptual compliance styles are generally identified (Angrist, Imbens, & Rubin, 1996). "Compliers" are those who

[1] Some may refer to the treatment assignment variable as Z and the treatment delivered variable as D (Angrist, 2006).

comply with their treatment assignment (i.e., will take the treatment if assigned to it or will not take the treatment if assigned to the control group). However, there are those who will take the treatment, regardless of whether they are assigned to the treatment or control groups. Such individuals are referred to as "always-takers". On the other hand, there are those who will never take the treatment, regardless of treatment assignment and those individuals are referred to as "never-takers". A final group is referred to as "defiers" who only take the treatment if assigned to the control group or do not take the treatment if assigned to the treatment group. However, defiers (who do the opposite of what they are assigned) are assumed to be rare or nonexistent (Angrist & Pischke, 2014).

If an individual is assigned to the treatment group and takes the treatment, such a person could be either a complier or an always-taker. If an individual is assigned to the control group and does not take the treatment, that person could be either a complier or a never-taker. However, to isolate the treatment effect on the compliers, we must be able to estimate, out of the population of individuals, what percent were compliers. In order to do so, knowledge of the proportion of always-takers and never-takers is required.

Due to random assignment of individuals to the treatment or control groups, the assumption is that each group has an approximately equal proportion of compliers, always-takers, and never-takers. In other words, in the pool of individuals who participated in the project (before random assignment), a proportion of them would be compliers, always-takers, and never-takers (we assume that no defiers exist[2]). After random assignment, we can expect an equal proportion of compliers, always-takers, and never-takers in both treatment and control groups.

For those assigned to the treatment group we can only observe the proportion of never-takers (i.e., those who were assigned to the treatment but did not show up). For those assigned to the control group, we can only observe the proportion of always-takers (i.e., those who received the treatment despite being in the control group). Due to the assumption that each assigned group has the same proportion of never-takers and always-takers, we can assume that the proportion of never takers found in the treatment assigned group has the same proportion of never takers in the control group. In the same manner, the percent of always takers should be the same as the those found in the assigned treatment group. So, knowing the proportion of noncompliers (e.g., always takers and never takers) in one group allows researchers to estimate what percent of the participants were noncompliers in the other group.

For example, in an experiment, out of 200 participants with 100 randomly assigned to either a treatment or control group, 80% of treatment assigned individuals took the treatment. This suggests that 20% of individuals were never-takers (i.e., those who did not take the treatment). In the control assigned group, we observe that 10% of individuals took the treatment through some means, which suggests that 10% of individuals were always-takers. We then assume that 70% of participants were compliers (i.e., 100% of individuals – 20% never-takers – 10% always-takers = 70%) and 30% were noncompliers (always-takers and never-takers). We can also say that this is a type of two-sided noncompliance where we have crossovers coming from both assigned groups.

Estimating the Treatment Effects for Compliers

Knowing the percent of compliers allows evaluators to estimate the local average treatment effect (LATE) or the treatment effect for those who complied (i.e., local to compliers). Sometimes, for compliers, this is referred to as the complier average causal effect (CACE) or complier average treatment effect (CATE). The LATE for compliers is computed as the ratio of the ITT estimate to the proportion of compliers. In other words, LATE = ITT / proportion compliers which results in the 'full' effect of the treatment considering that ITT is discounted or diluted by the presence of the noncompliers.

In addition to two-sided noncompliance, one-sided noncompliance is also possible in instances where there is no way control group participants can receive the treatment (i.e., there is a strict adherence to the treatment assignment and no one can sneak into the treatment condition). In such cases, there are no

[2] This is referred to as the monotonicity assumption (Angrist, 2006).

directly observed always-takers (since no one in the control-group can get the treatment) but there can still be never-takers (sometimes referred to as a failure to treat) which we observe as those that were assigned treatment but do not take it. The estimation of the effect is still the same but the causal estimand is referred to as the treatment effect on the treated or TOT (Gertler et al., 2016).

A way of thinking about how we compute the treatment effect on compliers is to think of the ITT as an effect discounted (or diluted) due to noncompliers. If we purchase a product which was discounted by 25% for a sale price of $22.50, to estimate the full cost of the product, we take the sale price divided by (1 – discount rate). In this case, the full price was $30 or 22.50 / (1 - .25).

Manuscripts evaluating treatment effects with imperfect compliance (see Angrist et al., 2010; Leventhal & Brooks-Gunn, 2003) will often present results based on the ITT as well as the LATE or TOT effect (depending on the type of noncompliance). The actual LATE or TOT effect is estimated using an IV approach.

Using Instrumental Variable Estimation

Given imperfect compliance in some experiments or RCTs, IV estimation can be used to recover the treatment effect for those who complied with their treatment assignment. An IV is a variable that has a causal effect on the measured outcome but only indirectly through a second variable of treatment receipt. In the RCT context, treatment assignment is the IV that pushes participants to take the treatment which ultimately affects the measured outcomes.

In the charter school example, seat assignment (or winning/losing the lottery) is the IV. The actual treatment is attending the charter school. It is reasonable to assume that differences in outcomes between winners and losers of the seating lottery (which is a form of random assignment) only result from the student attending or not attending the charter school, not merely because he/she won the lottery (see Figure 1)[3]. In other words, path c is nonexistent. Although conceptually, the model is a full mediation model, the effect is not estimated using path analysis or structural equation modeling (SEM) as is commonly done in education or psychology (i.e., the indirect path is not path a x path b). Instead, IV effects are estimated using what is referred to as two-stage least squares regression (2SLS or TSLS) which is also a series of regression equations (but with a slight twist).

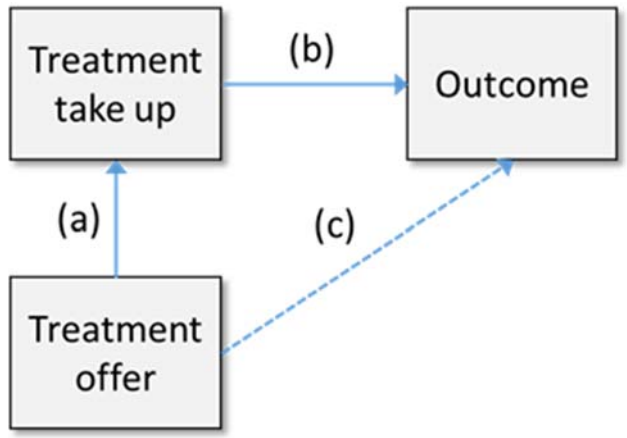

Figure 1. Instrumental Variable Illustration using Two-Stage Least Squares Regression.
Note. The dashed line implies that there is no direct relationship between treatment assignment and the outcome.

The first stage of IV estimation focuses on whether the instrument (e.g., the offer of the treatment) pushes participants to take up the treatment (e.g., attend the charter school). This can be estimated by regressing treatment take up on treatment offer. The resulting coefficient (a) for the assignment variable in the first stage can be interpreted as the compliance rate. One simple way of computing the IV effect is by dividing the ITT effect by the compliance rate (though this will not provide standard errors which are needed for statistical inference tests).

The second stage of the IV estimation involves running a regression predicting the outcome using the predicted take up values from the first stage regression. The use of the predicted values is an important distinction which separates TSLS estimation from standard mediation analysis although graphically, they may look similar. In a standard path analysis (which again is incorrect), the outcome would be regressed on the original treatment take up variable but not the predicted treatment take up values. Although in the

[3] This is referred to as the exclusion restriction assumption (Angrist, 2006).

2nd stage regression, standard errors are computed, they are still not correct and often, the whole IV chain will be estimated using commands found in popular statistical software which also compute correct standard errors.

The purpose of the reminder of the article is to walk readers through the steps to conduct a basic IV analysis using freely available R software together with readily downloadable data[4]. By providing a simple to understand example, together with data and syntax for analysis, I hope to make the use of IV analysis more accessible and interpretable to applied evaluation researchers.

Conducting an IV Analysis

The data for analysis come from a fictitious experiment comprised of 200 participants, 100 who were randomly assigned to a treatment group and 100 who were randomly assigned to a control group. Subjects in the treatment group are told to attend a seminar (i.e., the treatment) and if they attend the seminar, they will automatically receive somewhere from $7 to $13 (an average of $10) in return[5]. The outcome here is simply the money that participants will get from attending the seminar which we know should be around $10. Participants assigned to the control group are not told about the seminar. However, on the day of the seminar, only 78% of those assigned to the treatment group attend the seminar. In addition, some participants in the control group (9%) hear about the seminar and attend as well (the experimenter, maybe unwisely, does not bar anyone from attending). What then is the impact of the seminar in terms of dollar amount received? Due to imperfect compliance, we need to define the impact for a particular group of individuals.

Using R, we first read in the data and check some basic cross tables, using the crosstab function in the descr (Aquino, 2016) package to assess compliance rates[6]. In the dataset, y is the outcome, assign is the assignment variable (1 = assigned to treatment, 0 = assigned to control), and takeup indicates actual treatment status (1 = attended seminar, 0 = did not attend seminar).

```
> dat <-
read.csv('http://pareonline.net/sup/v23n2.csv')
> summary(dat)
     assign         takeup             y
 Min.   :0.0   Min.   :0.000   Min.   : 0.000
 1st Qu.:0.0   1st Qu.:0.000   1st Qu.: 0.000
 Median :0.5   Median :0.000   Median : 0.000
 Mean   :0.5   Mean   :0.435   Mean   : 4.375
 3rd Qu.:1.0   3rd Qu.:1.000   3rd Qu.:10.000
 Max.   :1.0   Max.   :1.000   Max.   :13.000

> library(descr) #for the crosstab function
> crosstab(dat$assign, dat$takeup, prop.r = T)
   Cell Contents
|-------------------------|
|                   Count |
|             Row Percent |
|-------------------------|

================================
             dat$takeup
dat$assign      0     1   Total
--------------------------------
0              91     9     100
               91%    9%     50%
--------------------------------
1              22    78     100
               22%   78%     50%
--------------------------------
Total         113    87     200
================================
```

Using ITT analysis, we run an OLS regression regressing y on assignment status, $y_i = B_0 + B_1(Assign_i) + e_i$. The coefficient for B_1 is the ITT effect for treatment assignment which is $6.95. On average, participants assigned to the treatment group received $6.95 more than the participants in the control group. Participants in the control group received $0.90, on average, as some in the control group managed to attend the seminar and collect as well.

```
> summary(lm(y ~ assign, data = dat))

Call:
lm(formula = y ~ assign, data = dat)

Residuals:
   Min     1Q Median     3Q    Max
 -7.85  -0.90  -0.90   2.15  11.10

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9000     0.3671   2.452   0.0151 *
assign        6.9500     0.5191  13.388   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

However, we know that the experimenter did not merely give out $7.85 (i.e., 6.95 + 0.90) on average and

[4] The sample dataset is available at http://pareonline.net/sup/v23n2.csv.

[5] Note: The range is only used to establish some variability in the outcome.

[6] Another way to estimate compliance rates (though less readable) in base R is to use: prop.table(xtabs(~assign + takeup, data = dat), 1)

the stated average amount should be around $10. In order to recover the full effect for the compliers, we need to assess what percent of participants complied with their treatment assignment.

Using the cross tabs already provided, we estimate that 22% of participants were never-takers (as they were assigned to the treatment but did not show up). At the same time, we also see that 9% were always-takers as some individuals in the assigned control group showed up to receive the treatment. The noncompliance rate is then 31% resulting in a compliance rate of 69%.

The compliance rate can also be directly estimated by regressing treatment take up on treatment assignment (the first stage of the regression). The resulting coefficient of 0.69 shows the compliance rate as well. Often, IV users will want to see signs of a "strong instrument" which gives a firm push to participants to take up the treatment and t values above approximately 3 or F values greater than 10 suggest the absence of a weak instrument (Angrist, 2006).

```
> stage1 <- lm(takeup ~ assign, data = dat)
> summary(stage1)

Call:
lm(formula = takeup ~ assign, data = dat)

Residuals:
   Min     1Q Median     3Q    Max
 -0.78  -0.09  -0.09   0.22   0.91

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09000    0.03578   2.515   0.0127 *
assign       0.69000    0.05060  13.636   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

Knowing the ITT ($6.95) and the compliance rate (69%) allows us then to estimate the LATE for compliers which is 6.95/.69 = $10.07. This adjustment has also been referred to as a "compliance adjusted ITT analysis" (Sussman & Hayward, 2010). This can be also estimated using the second stage regression where we predict y using the predicted (or fitted) values based on the first stage regression. The resulting coefficient is also $10.07.

```
> stage2 <- lm(y ~ fitted(stage1), data = dat)
> summary(stage2)

Call:
lm(formula = y ~ fitted(stage1), data = dat)

Residuals:
   Min     1Q Median     3Q    Max
 -7.85  -0.90  -0.90   2.15  11.10

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.0065    0.4176  -0.016    0.988
fitted(stage1) 10.0724    0.7523  13.388   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

The computations shown to derive the $10 is helpful for pedagogical purposes. However, in actuality, using R, the LATE can be estimated using the `ivreg` function which can be found in the AER (Kleiber & Zeileis, 2008) package or the `tsls` function in the sem (Fox, Nie, & Byrnes, 2017) package. Either package must be installed and loaded for the respective functions to work. With the functions, the second stage must first be specified with the instrument specified after the comma (see syntax below). A benefit of running the 2SLS regression is that standard errors will be estimated correctly and covariates can easily be included in the model to improve model power (see function documentation). In addition, with `ivreg`, robust and cluster robust standard errors may also be obtained using the ivpack (Jiang & Small, 2014) package.[7]

```
> library(AER)
> iv1 <- ivreg(y ~ takeup, ~assign, data = dat)
> summary(iv1)

Call:
ivreg(formula = y ~ takeup | assign, data = dat)

Residuals:
      Min        1Q    Median        3Q       Max
-3.065942 -0.065942  0.006522  0.006522  2.934058

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.00652    0.085095  -0.077    0.939
takeup      10.07246    0.153269  65.718   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

Conclusion

Although IV estimation methods are commonly used in evaluations performed by researchers with an econometric background, IV estimation is not often used by applied researchers who may be familiar with standard regression techniques. In the end, IV estimation is a form of regression analysis which has its own set of associated terms (e.g., ITT, LATE, TSLS) that may be confusing. If noncompliance to treatment assignment is purely random or if compliance is

[7] Note: Robust standard errors may also be obtained using summary(iv1, vcov = sandwich).

perfect, then standard OLS techniques which regress the outcome on treatment take up will yield the same effects as IV estimation. However, with noncompliance to treatment assignment, evaluators should perform their due diligence by testing the robustness of their findings using IV estimation as well. Although I do not discuss how IV estimation can be used to analyze secondary datasets based on natural experiments (e.g., Angrist & Krueger, 1991; Dee, 2004; Li & Konstantopoulos, 2016), understanding the logic of IV in the context of RCTs, as presented in the current manuscript, should aid researchers understand other papers that describe IV estimation in much further detail.

References

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology*, 2, 23–44. https://doi.org/10.1007/s11292-005-5126-x

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., & Walters, C. R. (2010). Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review*, 100, 239–243.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455. https://doi.org/10.2307/2291629

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? The *Quarterly Journal of Economics*, 106, 979–1014. https://doi.org/10.2307/2937954

Angrist, J. D., & Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton university press.

Angrist, J. D., & Pischke, J.-S. (2014). Mastering 'metrics: The path from cause to effect. Princeton University Press.

Aquino, J. (2016). descr: Descriptive Statistics. Retrieved from https://CRAN.R-project.org/package=descr

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Tutorial in biostatistics: Instrumental variable methods for causal inference. *Statistics in Medicine*, 33, 2297–2340. https://doi.org/10.1002/sim.6128

Clark, M. A., Gleason, P. M., Tuttle, C. C., & Silverberg, M. K. (2015). Do charter schools improve student achievement? *Educational Evaluation and Policy Analysis*, 37, 419–436. https://doi.org/10.3102/0162373714558292

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, 88, 1697–1720. https://doi.org/10.1016/j.jpubeco.2003.11.002

Fox, J., Nie, Z., & Byrnes, J. (2017). sem: Structural Equation Models. Retrieved from https://CRAN.R-project.org/package=sem

Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, 44, 381–394. https://doi.org/10.1037/0012-1649.44.2.381

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). Impact evaluation in practice (2nd ed.). Washington, DC: World Bank Publications. Retrieved from http://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, 2, 109–112. https://doi.org/10.4103/2229-3485.83221

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *The BMJ*, 319, 670–674.

Jiang, Y., & Small, D. (2014). ivpack: Instrumental Variable Estimation. Retrieved from https://CRAN.R-project.org/package=ivpack

Kleiber, C., & Zeileis, A. (2008). Applied Econometrics with R. New York: Springer-Verlag. Retrieved from https://CRAN.R-project.org/package=AER

Leventhal, T., & Brooks-Gunn, J. (2003). Moving to opportunity: An experimental study of neighborhood effects on mental health. *American Journal of Public Health*, 93, 1576–1582.

Li, W., & Konstantopoulos, S. (2016). Class size effects on fourth-grade mathematics achievement: Evidence from TIMSS 2011. *Journal of Research on Educational Effectiveness*, 9, 503–530. https://doi.org/10.1080/19345747.2015.1105893

Murnane, R. J., & Willett, J. B. (2011). Methods matter: Improving causal inference in educational and social science research. New York, NY: Oxford University Press.

Pokropek, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-Scale Assessments in Education*, 4. https://doi.org/10.1186/s40536-016-0018-2

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rose, R. A., & Stone, S. I. (2011). Instrumental variable estimation in social work research: A technique for estimating causal effects in nonrandomized settings. *Journal of the Society for Social Work and Research*, 2, 76–88. https://doi.org/10.5243/jsswr.2011.4

Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, 261–272.

Sovey, A. J., & Green, D. P. (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, 55, 188–200. https://doi.org/10.1111/j.1540-5907.2010.00477.x

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education*, 3, 285–289. https://doi.org/10.4300/JGME-D-11-00147.1

Sussman, J. B., & Hayward, R. A. (2010). An IV for the RCT: Using instrumental variables to adjust for treatment contamination in randomised controlled trials. *The BMJ*, 340. https://doi.org/10.1136/bmj.c2073

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., … Mullen, P. D. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98, 1359–1366. https://doi.org/10.2105/AJPH.2007.124446

Ye, C., Beyene, J., Browne, G., & Thabane, L. (2014). Estimating treatment effects in randomised controlled trials with non-compliance: A simulation study. *The BMJ Open*, 4. https://doi.org/10.1136/bmjopen-2014-005362

Citation:

Huang, Francis L. (2018). Using Instrumental Variable Estimation to Evaluate Randomized Experiments with Imperfect Compliance. *Practical Assessment, Research & Evaluation*, 23(2). Available online: http://pareonline.net/getvn.asp?v=23&n=2

Corresponding Author

Francis L. Huang, Assistant Professor
Statistics, Measurement, & Evaluation in Education Program
Educational, School, and Counseling Psychology
University of Missouri
16 Hill Hall
Columbia, MO 65211

email: huangf [at] Missouri.edu