Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. http://www.jstatsoft.org/

ivmodel: An R Package for Inference and Sensitivity Analysis of Instrumental Variables Models with One Endogenous Variable

Yang Jiang[*]
University of Pennsylvania

Hyunseung Kang[*]
University of Pennsylvania

Dylan S. Small
University of Pennsylvania

Abstract

We present a unified R software **ivmodel** for analyzing instrumental variables with one endogenous variable. The package implements a general class of estimators, k-class estimators, and two confidence intervals that are fully robust to weak instruments. The package also provides power formulas. Finally, the package contains methods for sensitivity analysis to examine the sensitivity to the the instrumental variables assumptions. We demonstrate the software on the data set from Card (1995), looking at the causal effect of levels of education on log earnings where the instrument is the proximity to a four-year college.

Keywords: econometrics, instrumental variables, power, sensitivity analysis, weak instruments.

1. Introduction

The instrumental variables (IV) method is a popular method to estimate the casual effect of a treatment, exposure, or policy on an outcome when there is concern about unmeasured confounding (Angrist *et al.* 1996; Angrist and Krueger 2001; Baiocchi *et al.* 2014). IV methods have been widely used in many field including statistics (Angrist *et al.* 1996), economics (Angrist and Krueger 2001), genomics and epidemiology (Davey Smith and Ebrahim 2003), sociology Bollen (2012), psychology (Gennetian *et al.* 2008), political science (Sovey and Green 2011), and countless others. We also note that instrumental variables have been used to correct for measurement errors (see Fuller (2006) for a full treatment on measurement errors).

[*]These authors contributed equally to this work

Informally speaking, IV methods rely on having variables called instruments which are related to the exposure and are exogenous. An instrument is exogenous if it only affects the outcome through the pathway of affecting the exposure (i.e. the instrument has no direct effect on the outcome) and is independent of unmeasured confounders (see Section 2.3 for details). Typically, instruments either come from (i) natural experiments whereby the instruments were naturally assigned to individuals at random or (ii) an actual randomized experiment whereby the actual randomization mechanism is used as an instrument. For example, in a field known as Mendelian randomization, natural genetic variations have been used as an instrument to answer causal questions in epidemiology (Davey Smith and Ebrahim 2003, 2004; Lawlor *et al.* 2008). Another example of the use of an instrument is in the study of the effect of pregnant mother's smoking on birth weight by Sexton and Hebel (1984) and Permutt and Hebel (1989). Here, the instrument was the actual randomized encouragement assignment of mothers to one of the two groups, the first group where the healthcare provider encouraged the mothers to stop smoking and the second group where the healthcare provider did not provide such encouragement. Table 1 illustrates other examples of instrumental variables, divided based on the source of the instruments. For more examples, see Angrist and Krueger (2001) and Baiocchi *et al.* (2014).

Outcome	Exposure	Instruments	Reference
	Natural experiments / Mendelian randomization		
Earnings	Years of schooling	Proximity to college when growing up	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Metabolic phenotypes	C-reactive protein (CRP)	SNPs rs1800947, rs1130864, rs1205	Timpson *et al.* (2005)
Blood pressure	Alcohol intake	Alcohol dehydrogenase (ALDH2) genotype	Chen *et al.* (2008)
	Randomized experiments / Encouragement designs		
Birth weight	Mother's smoking	Randomized encouragement to stop smoking	Sexton and Hebel (1984) and Permutt and Hebel (1989)
Test scores	Class size	Randomized assignment to different class sizes	Krueger (1999)

Table 1: Application of instrumental variables methods based on source of instruments. Natural experiments/Mendelian randomization refer to instrumental variables studies where the instruments come from natural sources, such as genes or calendar years. Randomized experiments/encouragement designs refer to instrumental variables studies where the instruments represent actual randomization mechanisms.

Software for running instrumental variables methods varies widely depending on the programming language. For example, in STATA, there are comprehensive and unified programs to handle the most popular instrumental variables methods, most notably **ivreg2** (Baum *et al.* 2003, 2007) and STATA's default program **ivregress**. In R, different instrumental variables methods are implemented across different packages, for instance **AER** by Kleiber and Zeileis (2008), **sem** by Fox *et al.* (2014), and **lfe** by Gaure (2013). Unfortunately, these packages do not include (i) modern instrumental variables methods that provide confidence intervals that are fully robust to weak instruments (see Section 4), (ii) sensitivity analysis methods that examine sensitivity of inference to violations of IV assumptions (see Section 5), and (iii) power calculations for IV analysis (see Section 6).

The goal of the paper is to present a package **ivmodel** that integrates and unifies R functions to conduct a comprehensive instrumental variables analysis when there is one exposure/endogenous variable. These functions include a general class of estimators known as k-class estimators (see Section 3) and the corresponding standard errors, confidence intervals, and p-values. The functions also integrate more modern approaches to IV analysis, including two methods for confidence intervals that are fully robust to weak instruments (Stock *et al.* 2002), the Anderson and Rubin confidence interval (Anderson and Rubin 1949) and the conditional likelihood ratio confidence interval (Moreira 2003). The package includes functions to calculate power. Finally, the package includes methods to conduct sensitivity analysis to examine the sensitivity to the IV assumptions not holding. All these functions are integrated into an R software package called **ivmodel**.

2. Instrumental Variables Model for One Endogenous Variable

2.1. Notation

Let there be n individuals indexed by $i = 1, \ldots, n$. For each individual i, we observe the outcome $Y_i \in \mathbb{R}$, the exposure $D_i \in \mathbb{R}$, the L instruments $Z_{i.} \in \mathbb{R}^L$, and the p covariates $X_{i.} \in \mathbb{R}^p$. Let $Y = (Y_1, \ldots, Y_n) \in \mathbb{R}^n$ denote the vector of outcomes, $D = (D_1, \ldots, D_n) \in \mathbb{R}^n$ denote the vector of exposures, $Z \in \mathbb{R}^{n \times L}$ denote the matrix of instruments where the ith row corresponds to the vector $Z_{i.}$, and $X \in \mathbb{R}^{n \times p}$ denote the matrix of covariates where the ith row corresponds to the vector $X_{i.}$. Let $W = [Z : X]$ where W is an n by $L + p$ matrix that concatenates the matrices Z and X.

For any matrix M, denote its transpose M^T. Also, for any matrix M, let $P_M = M(M^T M)^{-1} M^T$ be the orthogonal projection matrix onto the column space of M and R_M be the residual projection matrix so that $R_M + P_M = I$, where I is an n by n identity matrix. We assume that M has a proper inverse, so that $(M^T M)^{-1}$ is well-defined, unless otherwise stated.

2.2. Model

We assume the following linear structural model between the observed quantities, $Y_i, D_i, Z_{i.}$, and $X_{i.}$

$$Y_i = D_i \beta + X_{i.}^T \kappa + \epsilon_i, \quad \mathsf{E}(\epsilon_i | Z_{i.}, X_{i.}) = 0, \mathsf{VAR}(\epsilon_i | Z_{i.}, X_{i.}) = \sigma^2 \qquad (1)$$

This is the standard, single equation homoscedastic linear structural model in econometrics (Wooldridge 2010). Note that this is not the usual regression model in the sense that D_i is correlated with ϵ_i. The parameter of interest is β, which is the causal effect of the exposure D_i on the outcome Y_i (see next paragraph for more details on causal effects). The parameter

κ relates the covariates to the outcome. Note that $X_{i.}$ can contain a value of 1 to represent the intercept.

The parameters in model (1) can be given a causal interpretation under the potential outcomes notation (Rubin 1974) where (1) represents the additive linear, constant effects (ALICE) model in Holland (1988). Let $Y_i^{(d,z)}$ be the potential outcome if individual i were to have exposure d, a scalar value, and instruments z, an L dimensional vector. Let $D_i^{(z)}$ be the potential exposure if the individual had instruments z. For each individual, only one possible realizations of $Y_i^{(d,z)}$ and $D_i^{(z)}$ is observed, denoted as Y_i and D_i, respectively, based on his/her observed instrument values $Z_{i.}$ and exposure D_i. Then, for two possible values of the exposure d', d and instruments z', z, we assume the following potential outcomes model

$$Y_i^{(d',z')} - Y_i^{(d,z)} = (d' - d)\beta \quad \mathsf{E}(Y_i^{(0,0)} \mid Z_{i.}, X_{i.}) = X_{i.}^T \kappa \tag{2}$$

In model (2), β represents the causal effect (divided by $d' - d$) of changing the exposure from d' to d on the outcome. The parameter κ represents the impact of covariates on the baseline potential outcome $Y_i^{(0,0)}$. If we further define $\epsilon_i = Y_i^{(0,0)} - \mathsf{E}(Y_i^{(0,0)} \mid Z_{i.}, X_{i.})$, we obtain the observed data model in (1), thus providing the parameters in the observed model in (1) a causal interpretation.

Note that in many works in econometrics literature, one makes additional assumptions about the relationship between the endogenous variable D_i, the instruments $Z_{i.}$, and the covariates $X_{i.}$, specifically

$$D_i = Z_{i.}^T \gamma + X_{i.}^T \tilde{\kappa} + \eta_i, \quad \mathsf{E}(\eta_i | Z_{i.}, X_{i.}) = 0, \mathsf{VAR}(\eta_i | Z_{i.}, X_{i.}) = \omega^2 \tag{3}$$

This "first stage" model in (3) is not necessary for all our methods in the **ivmodel** package. In particular, the k-class estimators in Section 3 and the confidence interval for the Anderson and Rubin test in Section 4 are valid without the first stage modeling assumption in (3). However, the other methods presented in the paper require this assumption and we introduce it in this section.

Similar to equation (2), we can provide a causal interpretation of the first stage model in (3) as follows.

$$D_i^{(z')} - D_i^{(z)} = (z' - z)\gamma \quad \mathsf{E}(D_i^{(0)} \mid Z_{i.}, X_{i.}) = X_{i.}^T \tilde{\kappa} \tag{4}$$

In model (4), γ represents the causal effect (divided by $z' - z$) of changing the IV from z' to z on the exposure. The parameter $\tilde{\kappa}$ represents the impact of covariates on the baseline potential outcome $D_i^{(0)}$. As before, if we further define $\eta_i = D_i^{(0)} - \mathsf{E}(D_i^{(0)} \mid Z_{i.}, X_{i.})$, we obtain the observed data model in (3).

Without loss of generality and throughout the paper, we will use the simplified version of the models in equations (1) and (3) where we project out the covariates X by the Frisch-Waugh-Lovell Theorem (Davidson and MacKinnon 1993). Specifically, models (1) and (3) are equivalent to

$$Y_i^* = D_i^* \beta + \epsilon_i^* \tag{5}$$
$$D_i^* = Z_{i.}^* \gamma + \eta_i^* \tag{6}$$

where

$$Y^* = R_X Y, \quad D^* = R_X D, \quad Z^* = R_X Z, \quad \epsilon^* = R_X \epsilon, \quad \eta^* = R_X \eta$$

The superscripts Y^*, D^*, Z^* represent the outcome, the exposure, and the instruments after controlling for the covariates X by the residual orthogonal projection R_X defined in Section 2.1. The equivalent models (5) and (6) allow us to concentrate on the target parameter of interest, β, and simplify the derivations and expressions of the instrumental variables methods presented in the paper. Note that as before, the model in (6), the simplified version of the first stage model in (3), is not necessary for k-class estimators and the Anderson and Rubin confidence intervals.

2.3. Assumption of Instrumental Variables

Under the model in (1), we make the standard assumptions in the instrumental variables literature below (Wooldridge 2010).

(A1) $\mathsf{E}(W^T W)$ is full rank.

(A2) Conditional on the covariates X, the instruments Z are associated with the exposure D, $\mathsf{E}(Z^T R_X D) \neq 0$

(A3) W is exogenous, $\mathsf{E}(W^T \epsilon) = 0$

Assumption (A1) is a standard moment condition on the matrix of exogenous variables that include the covariates and the instruments. Assumption (A2) states that conditional on the covariates X, the instruments are associated with the exposure. There are many ways to test this assumption in practice, the most popular being the F statistic for the coefficients on the variables in Z being 0 when regressing D on X and Z. A strong association between the instruments Z and the exposure D is desired to reduce the precision of an IV estimator (Stock *et al.* 2002). Instruments with strong associations are considered to be strong instruments while instruments with weak associations are considered to be weak instruments. For example, in the case of one instrument, an instrument is considered weak if the F statistic is less than 10 (Stock *et al.* 2002).

For assumption (A3), in the ALICE model, (A3) is satisfied if Z has no direct on D and Z is independent of unmeasured confounders. Assumption (A3) is generally untestable in that it's impossible to check whether the exogenous variables Z and X are uncorrelated with the structural error ϵ_i, which is never observed. However, methods exist to partially test this assumption if there are more than one instruments, $L > 1$, the most popular being the Sargan's test (Sargan 1958). Under all the three assumptions (A1)-(A3), standard econometric arguments show that the the model (1) is identified (Wooldridge 2010).

Typically, practitioners assume that they have found instruments that satisfy (A1)-(A3) (Angrist and Krueger 2001). However, violations of these assumptions occur, especially (A2) and (A3), and there has been progress in the literature to handle these violations (Angrist and Krueger 2001; Murray 2006). For (A2), even if it is satisfied, but only weakly, which is known as the weak instrument problem, the most commonly used instrumental variables estimation method, two stage least squares (TSLS), produces biased estimates of β in (1) (Nelson and Startz 1990; Staiger and Stock 1997; Stock *et al.* 2002). Thankfully, many statistical methods exist to provide robust and honest estimates of the parameters in model (1) with weak instruments (Stock *et al.* 2002) (see Section 4 for details). Violations of (A3), known as the invalid instrument problem, is the case where the instruments Z may have a direct effect on the outcome or when the instruments are correlated with ϵ_i. This problem has received less attention than the weak instrument problem (Murray 2006), but has recently been considered by Kolesár *et al.* (2013), Kang *et al.* (2015), and Jiang *et al.* (2015).

Throughout the paper, we assume that our instruments Z satisfy assumptions (A1)-(A3). However, we discuss violations of (A2) and (A3) in Sections 4 and 5, respectively and provide methods that can handle these violations.

3. k-Class Estimation and Inference

3.1. Definitions and General Properties

A class of estimators for β, called the k-class estimator and denoted as $\hat{\beta}_k$, is defined as follows.

$$\hat{\beta}_k = (D^{*T}(I - kP_{Z^*})D^*)^{-1}D^{*T}(I - kP_{Z^*})Y^* \tag{7}$$

Table 2 lists all the estimators that are k-class estimators, including the ordinary least squares (OLS), two-stage least squares (TSLS), limited information maximum likelihood (LIML), and Fuller's estimator (FULL). In Table 2, k_{LIML} is the minimum value of k of the following equation

$$det \begin{pmatrix} Y^{*T}(I - kR_{Z^*})Y^* & Y^{*T}(I - kR_{Z^*})D^* \\ D^{*T}(I - kR_{Z^*})Y^* & D^{*T}(I - kR_{Z^*})D^* \end{pmatrix} = 0 \tag{8}$$

k	Name
$k = 0$	Ordinary least squares (OLS)
$k = 1$	Two-stage least squares (TSLS)
$k = k_{LIML}$	Limited information maximum likelihood (LIML)
$k = k_{LIML} - \frac{b}{n-L-p}, b > 0$	Fuller's estimator (FULL)

Table 2: Different types of k-class estimator

Each k yields an estimator with unique properties, which will be discussed in detail in Section 3.2. However, for all k-class estimators in equation (7), an estimate for the standard error of $\hat{\beta}_k$ is

$$\widehat{\mathsf{VAR}}(\hat{\beta}_k) = \hat{\sigma}^2(D^{*T}(I - kP_{Z^*})D^*)^{-1}, \quad \hat{\sigma}^2 = \frac{(Y^* - D^*\hat{\beta}_k)^T(Y^* - D^*\hat{\beta}_k)}{n - L - p} \tag{9}$$

As long as L and p are fixed, all k-class estimators are consistent so long as $k \to 1$ as $n \to \infty$ in probability (Davidson and MacKinnon 1993). In addition, for fixed L and p, as long as $\sqrt{n}(k-1) \to 0$ in probability as $n \to \infty$, the k-class estimator has the following asymptotic Normal distribution (Amemiya 1985)

$$\frac{\hat{\beta}_k - \beta}{\sqrt{\widehat{\mathsf{VAR}}(\hat{\beta}_k)}} \to N(0,1) \tag{10}$$

The asymptotic distribution in (10) allows us to test the hypothesis

$$H_0 : \beta = \beta_0, \quad H_a : \beta \neq \beta_0 \tag{11}$$

by comparing the standardized deviate in (10) to the standard Normal (or the t distribution with degrees of freedom $n - L - p$). We can also create $1 - \alpha$ confidence intervals for β based on $\hat{\beta}_k$, i.e.

$$\left(\hat{\beta}_k - z_{1-\alpha/2}\sqrt{\widehat{\mathsf{VAR}}(\hat{\beta}_k)}, \quad \hat{\beta}_k + z_{1-\alpha/2}\sqrt{\widehat{\mathsf{VAR}}(\hat{\beta}_k)}\right)$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard Normal distribution. We can alternatively use the $1 - \alpha/2$ quantile of the t distribution with degrees of freedom $n - L - p$.

3.2. Some Examples of k-class Estimators

The most well-known k-class estimator in instrumental variables is two-stage least squares (TSLS) where $k = 1$ in (7), i.e.

$$\hat{\beta}_1 = (D^{*T} M_{Z^*} D^*)^{-1} D^{*T} M_{Z^*} Y^*$$

In addition to being consistent and having an asymptotic Normal distribution, TSLS is efficient among all IV estimators using linear combination of instruments Z (Wooldridge 2010). In fact, under the asymptotics rates of $\sqrt{n}(k-1) \to 0$ introduced in Section 3.1, all k-class estimators have the same asymptotic Normal distribution as TSLS. Also, when $L = 1$, TSLS and LIML produce identical estimates of β (Davidson and MacKinnon 1993).

However, despite having the same asymptotic Normal distributions, TSLS and other types of k-class estimators, for instance LIML and FULL in Table 2, behave differently in finite-samples. With weak instruments (i.e. near violations of (A2)), TSLS tends to be biased towards OLS in finite sample. Even with large samples, TSLS can provide a very biased estimate of the causal effect in the presence of weak instruments (Bound *et al.* 1995). In contrast, LIML and FULL are more robust to violations of (A2) than TSLS (Stock *et al.* 2002). However, LIML has no finite moments of any order while TSLS has moments of up to $L - 1$. FULL corrects LIML's lack of moments by having moments so long as the sample size is large enough (Davidson and MacKinnon 1993).

Other types of k-class estimators exist and no single k-class estimator uniformly dominates another in all settings (Davidson and MacKinnon 1993). However, in practice, the most popular estimators are TSLS and LIML, with LIML having better robustness properties with regards to weak instruments (Stock *et al.* 2002; Mariano 2003; Chao and Swanson 2005)

4. Dealing with Weak Instruments: Robust Confidence Interval Estimation

In this section, we discuss the case when the instruments Z may nearly violate (A2), also known as the weak instrument problem, and discuss two methods that are fully robust to near violations of (A2).

Let M be an n by 2 matrix where the first column contains Y^* and the second column contains D^*. Let $a_0 = (\beta_0, 1)$ and $b_0 = (1, -\beta_0)$ to be two-dimensional vectors and $\hat{\Sigma} = M^T R_{Z^*} M/(n - L - p)$. Let $\hat{S}$ and $\hat{T}$ be two-dimensional vectors defined as follows.

$$\hat{S} = \frac{(Z^{*T} Z^*)^{-1/2} Z^{*T} M b_0}{\sqrt{b_0^T \hat{\Sigma} b_0}}, \quad \hat{T} = \frac{(Z^{*T} Z^*)^{-1/2} Z^{*T} M \hat{\Sigma}^{-1} a_0}{\sqrt{a_0^T \hat{\Sigma}^{-1} a_0}}$$

We also define the following scalar values, $\hat{Q}_1, \hat{Q}_2$, and $\hat{Q}_3$.

$$\hat{Q}_1 = \hat{S}^T \hat{S}, \quad \hat{Q}_2 = \hat{S}^T \hat{T}, \quad \hat{Q}_3 = \hat{T}^T \hat{T}$$

Based on $\hat{Q}_1$, $\hat{Q}_2$, and $\hat{Q}_3$, we define two tests of the hypothesis in equation (11) that are fully robust to violations of (A2), the Anderson and Rubin test (Anderson and Rubin 1949), and the conditional likelihood test (Moreira 2003).

$$AR(\beta_0) = \hat{Q}_1 / L \tag{12}$$

$$CLR(\beta_0) = \frac{1}{2}(\hat{Q}_1 - \hat{Q}_3) + \frac{1}{2}\sqrt{(\hat{Q}_1 + \hat{Q}_3)^2 - 4(\hat{Q}_1\hat{Q}_3 - \hat{Q}_2^2)} \tag{13}$$

Much work has shown that these two tests are fully robust to weak instruments (Staiger and Stock 1997; Stock *et al.* 2002; Moreira 2003; Dufour 2003; Andrews *et al.* 2006). Between the two tests, there is no uniformly most powerful test under weak instruments, but Andrews *et al.* (2006) and Mikusheva (2010) suggest using (13) due to its generally favorable power compared to (12) in most cases when weak instruments are present. However, the Anderson-Rubin test is the simplest of the two tests in that under a Normality error assumption (see next paragraph), it can be written as a standard F-test in regression where the outcome is $R_{Z^*}(Y - D\beta_0)$, the regressors are Z^*, and we are testing whether the coefficients associated with Z^* are zero or not with the standard F-test. Also, unlike the Anderson and Rubin test in (12), the conditional likelihood ratio test in (13) requires the first stage modeling assumption in (3) (Dufour 2003).

We can invert both tests in equation (12) and (13) to obtain $1 - \alpha$ confidence intervals that are fully robust to weak instruments, i.e. $\{\beta : AR(\beta_0) \leq F_{L,n-L-p,1-\alpha}\}$ for the Anderson and Rubin confidence interval and $\{\beta : CLR(\beta_0) \leq q_{1-\alpha}\}$ for the conditional likelihood ratio test. Here, $F_{L,n-L-p,1-\alpha}$ is the $1 - \alpha$ quantile of the F distribution with degrees of freedom L and $n - L - p$ and $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the the conditional likelihood ratio test. The F distribution for the Anderson and Rubin test is based on an assumption about Normality of the errors in model (1) and our package **ivmodel** currently uses the F distribution. However, one can also use the χ^2 distribution as an asymptotic approximation should the Normality assumption be unreasonable in the data. As for the distribution that underlies the conditional likelihood ratio test and the details on $q_{1-\alpha}$, see Andrews *et al.* (2007).

5. Dealing with Possibly Invalid Instruments: Sensitivity Analysis

Morgan and Winship (2007) showed that assumption (A3) cannot be completely tested. However, there is often concern that a putative IV is invalid in applications. In these cases, a sensitivity analysis can be used to examine the sensitivity of inferences to violations of (A3). Here we assume that there is only one IV ($L = 1$) in the study and this IV may be invalid to some degree.

There is some previous work on sensitivity analysis for IV studies, see DiPrete and Gangl (2004), Small (2007), Kolesár *et al.* (2011) and Conley *et al.* (2012). These papers all use test statistics which are based on the TSLS estimator having an approximately normal distribution, which breaks down in the presence of weak instruments (instruments that are weakly associated with the exposure), see Nelson and Startz (1990). Our sensitivity analysis uses the AR test statistic because of the following properties: the AR test is robust to a weak instrument; the AR test is uniformly most powerful among all unbiased tests (Moreira 2001) and the AR test is uniformly most powerful among all invariant similar tests (Andrews *et al.* 2006).

We revise the model in Section 2.2 to add the feature of invalid IV. For model (2), assume that Z_i violates the assumption (A3) so there is another term $\delta\sigma(z' - z)$ on the equation's right side:

$$Y_i^{(d',z')} - Y_i^{(d,z)} = (d' - d)\beta + \delta\sigma(z' - z), \quad \mathsf{E}(Y_i^{(0,0,0)} \mid Z_{i.}, X_{i.}) = X_{i.}^T\kappa \tag{14}$$

Here σ is the standard variation of $\epsilon_i = Y_i^{(0,0,0)} - \mathsf{E}(Y_i^{(0,0,0)} \mid Z_{i.}, X_{i.})$, which is a scaling parameter. δ measures how much the IV $Z_{i.}$ violates the assumption (A3). Further assume the sensitivity parameter is within a known range, $\delta \in (\underline{\delta}, \bar{\delta})$. Then the model for sensitivity analysis becomes:

$$Y_i = D_i\beta + X_{i.}^T\kappa + \delta\sigma Z_{i.} + \epsilon_i, \quad \mathsf{E}(\epsilon_i|Z_{i.}, X_{i.}) = 0, \quad \mathsf{VAR}(\epsilon_i|Z_{i.}, X_{i.}) = \sigma^2, \quad \delta \in (\underline{\delta}, \bar{\delta}) \tag{15}$$

If the error term has a normal distribution $\epsilon_i \sim N(0, \sigma^2)$, then hypothesis (11) can be tested by using the AR test statistic $AR(\beta_0)$ in equation (12). Under H_0, $AR(\beta_0)$ has a non-central F distribution :

$$AR(\beta_0) \sim F_{1,n-p-1,\delta^2 {Z^*}^T Z^*} \tag{16}$$

Although δ is unknown and consequently we don't known exact the distribution of $AR(\beta_0)$ under H_0, we can define $\Delta = \max(|\underline{\delta}|, |\bar{\delta}|)$ and construct an interval that provides at least $1 - \alpha$ confidence:

$$CI_{1-\alpha} = \{\beta : AR(\beta_0) < F_{1,n-p-1,\Delta^2 {Z^*}^T Z^*;1-\alpha}\} \tag{17}$$

For our sensitivity analysis, equation (17) is used for the hypothesis test. More details are provided in Jiang *et al.* (2015).

6. Power

If the research goal is to find evidence for an exposure effect, then we would like to know the power of rejecting the null hypothesis $H_0 : \beta = \beta_0$ when the true exposure effect is under the alternative $\beta - \beta_0 = \lambda \neq 0$. With a power formula, researchers can decide how large a sample to collect to achieve a certain power. Freeman *et al.* (2013) presents a power formula for using the asymptotic normal distribution of TSLS estimator to do hypothesis test. Jiang *et al.* (2015) provides a power formula for the AR test and sensitivity analysis. These three different power formulas are included in **ivmodel**. By inverting the power formula, we can calculate the sample size needed to achieve a certain power. These sample size calculation functions are also included in **ivmodel**. The three different approaches to a power formula rely on different assumptions, which will be discussed in the following paragraphs.

Freeman *et al.* (2013) assumes there is only one IV($L = 1$) and there are no observed covariates $X(p = 0)$, which is model (1) with $\kappa = 0$. Asymptotically, the TSLS estimator has a normal distribution:

$$\hat{\beta}_{TSLS} \sim N\left(\beta, \quad \frac{\sigma^2}{n \cdot \mathsf{VAR}(D) \cdot \rho_{ZD}}\right) \tag{18}$$

If the true exposure effect is $\beta - \beta_0 = \lambda$, then the power of testing hypothesis (11) is:

$$\text{Power} = 1 + \Phi\left(-z_{\alpha/2} - \frac{\lambda \rho_{ZD}\sqrt{n \cdot \mathsf{VAR}(D)}}{\sigma}\right) - \Phi\left(z_{\alpha/2} - \frac{\lambda \rho_{ZD}\sqrt{n \cdot \mathsf{VAR}(D)}}{\sigma}\right) \tag{19}$$

where α is the desired significance level of the test (usually 0.05), Φ is the cumulative distribution function of the standard normal distribution, z_α is the value satisfies $\Phi(-z_\alpha) = \alpha$ and ρ_{ZD} is the correlation between Z and D.

The AR test is based on model (1). In order to calculate the exact power, we need to have another model between the exposure and IV, as stated in (3) and make the bivariate normality assumption for the error (ϵ_i, η_i). The extended model is summarized as follows.

$$\begin{aligned} Y^* &= D^*\beta + \epsilon^* \\ D^* &= Z^*\gamma + \eta^* \\ Y^* = R_X Y, \quad D^* = R_X D, \quad Z^* = R_X Z, \quad \epsilon^* &= R_X \epsilon, \quad \eta^* = R_X \eta \\ (\epsilon, \eta) \perp Z; \quad (\epsilon_i, \eta_i)^T \sim N(\mathbf{0}, \Sigma); \quad \Sigma &= \begin{pmatrix} \sigma^2 & \rho\sigma\omega \\ \rho\sigma\omega & \omega^2 \end{pmatrix} \quad \text{rank}(X) = p \end{aligned} \tag{20}$$

If the true exposure effect is $\beta - \beta_0 = \lambda$, then the power of testing hypothesis (11) is:

$$\text{Power} = 1 - \Psi_{1, n-p-L, \frac{(\gamma^T Z^{*T} Z^* \gamma)\lambda^2}{\sigma^2 + 2\rho\sigma\omega\lambda + \omega^2\lambda^2}} \left(F_{1, n-p-L; 1-\alpha}\right) \tag{21}$$

where $F_{a,b;1-\alpha}$ is the $1-\alpha$ quantile of F distribution with degrees of freedom a and b. $\Psi_{a,b,k}(\cdot)$ is the cumulative distribution function of the non-central F distribution with degree of freedom a, b and non-central parameter k.

The sensitivity analysis relies on model (15) with one possibly invalid IV. To calculate its power, the model (3) and normality assumption is still needed. The extended model is similar to (20).

$$\begin{aligned} Y^* &= D^*\beta + \delta\sigma Z^* + \epsilon^* \\ D^* &= Z^*\gamma + \eta^* \\ Y^* = R_X Y; \quad D^* = R_X D; \quad Z^* = R_X Z; \quad \epsilon^* &= R_X \epsilon; \quad \eta^* = R_X \eta; \\ (\epsilon, \eta) \perp Z; \quad (\epsilon_i, \eta_i)^T \sim N(\mathbf{0}, \Sigma); \quad \Sigma &= \begin{pmatrix} \sigma^2 & \rho\sigma\omega \\ \rho\sigma\omega & \omega^2 \end{pmatrix} \quad \text{rank}(X) = p \end{aligned} \tag{22}$$

If the true exposure effect is $\beta - \beta_0 = \lambda$ and it is a favorable situation where the instrument is valid ($\delta = 0$) but we want to allow for the possibility that the instrument is invalid in the range $\delta \in (-\Delta, \Delta)$ (Rosenbaum 2010), then the power of being able to reject the null hypothesis for all $\delta \in (-\Delta, \Delta)$ is:

$$\text{Power} = 1 - \Psi_{1, n-p-1, \frac{\lambda^2\gamma^2 Z^{*T} Z^*}{\sigma^2 + 2\rho\sigma\omega\lambda + \omega^2\lambda^2}} \left(F_{1, n-p-1, \Delta^2 Z^{*T} Z^*; 1-\alpha}\right) \tag{23}$$

where $F_{a,b,c;1-\alpha}$ is the $1-\alpha$ quantile of the non-central F distribution with degree of freedom a, b and non-central parameter c. (23) is called the power of sensitivity analysis for sensitivity Δ.

When the sample size is small or moderate and the IV is weak, the asymptotic test (18) based on the two stage least squares estimator can have highly inflated Type I error and should be avoided (Jiang *et al.* 2015). For these settings, Jiang *et al.* (2015) recommend using the AR test and its associated power formula (21).

7. Application

In this section, we illustrate the application of **ivmodel** with the data set from Card (1995). The data is from the National Longitudinal Survey of Young Men (NLSYM), which has $n = 3010$ individual observations and 35 variables. We want to estimate the causal effect of education (`educ`) on log earnings (`lwage`). The IV is a binary variable indicating whether the individual grew up in a place with a nearby 4-year college (`nearc4`). There are also some exogenous variables included in the data (`exper`, `expersq`, `black`, `south`, etc).

7.1. Ivmodel Class and the Basic Usage

First, we load **ivmodel** and specify the outcome Y, the exposure D, the IV Z and other exogenous variables X in `card.data`.

```
R> Y=card.data[,"lwage"]
R> D=card.data[,"educ"]
R> Z=card.data[,"nearc4"]
R> Xname=c("exper", "expersq", "black", "south", "smsa", "reg661", "reg662",
R+          "reg663", "reg664", "reg665", "reg666", "reg667", "reg668", "smsa66")
R> X=card.data[,Xname]
R> cardfit=ivmodel(Y=Y, D=D, Z=Z, X=X)
```

The object `cardfit` generated by the function `ivmodel()` contains all the different IV methods. To display the information, use the `summary()` function:

```
R> summary(cardfit)

Call:
ivmodel(Y = Y, D = D, Z = Z, X = X)
sample size: 3010
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

First Stage Regression Result:

F=13.25579, df1=1, df2=2994, p-value is 0.00027634
R-squared=0.004407934,   Adjusted R-squared=0.004075405
Residual standard error: 1.940537 on 2995 degrees of freedom
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Coefficients of k-Class Estimators:
```

```
              k Estimate Std. Error t value Pr(>|t|)
OLS    0.000000 0.074693   0.003498  21.351   <2e-16 ***
Fuller 0.999666 0.127501   0.052708   2.419   0.0156 *
LIML   1.000000 0.131504   0.054964   2.393   0.0168 *
TSLS   1.000000 0.131504   0.054964   2.393   0.0168 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Alternative tests for the treatment effect under H_0: beta=0.

Anderson-Rubin test:
F=5.415279, df1=1, df2=2994, p-value=0.020028
95 percent confidence interval:
 [ 0.0248048359651019 , 0.284823593339036 ]

Conditional Likelihood Ratio test:
Test Stat=5.415279, p-value=0.019961
95 percent confidence interval:
 [0.0248546898484261, 0.284720676631606]
```

There are four sections in the display. The first section is a recall for the ivmodel expression and the sample size. The second section summarizes the first stage regression between the IV and exposure. Here the F statistic is 13.25579, which is greater than 10, indicating the IV is not weak (Stock *et al.* 2002). The third section lists the results for several k-class estimator. The default k's are $k = 0$ (OLS), $k = 1$ (TSLS), and k's associated with LIML and Fuller. They all suggest a significant effect of education on earnings. The OLS estimator is considerably smaller than the other three estimators. However, the other three estimators are consistent when the instrumental variable is valid but OLS is not. The last section provides the AR and CLR confidence intervals, which are robust even for weak IVs (although in this case the IV is not weak).

The function `coef()` applied to the generated object `cardfit` returns the coefficient matrix for the k-class estimators. The function `confint()` on the object `cardfit` returns the confidence interval for different IV methods.

```
R> coef(cardfit)

              k   Estimate  Std. Error   t value   Pr(>|t|)
OLS    0.000000 0.07469326 0.003498346 21.351022 0.00000000
Fuller 0.999666 0.12750110 0.052708406  2.418990 0.01562292
LIML   1.000000 0.13150384 0.054963673  2.392559 0.01679262
TSLS   1.000000 0.13150384 0.054963673  2.392559 0.01679262

R> confint(cardfit)

              2.5%      97.5%
OLS     0.06783385 0.08155266
```

```
Fuller  0.02415275 0.23084946
LIML    0.02373345 0.23927422
TSLS    0.02373345 0.23927422
AR      0.02480484 0.28482359
CLR     0.02485469 0.28472068
```

7.2. Options for k-Class Estimators and Sensitivity Analysis

In the function `ivmodel()`, we can use the option `k =` to specify a specific k-class estimator, for example `k = 0.9`. In addition, if there is concern that the IV may be invalid, we can specify the range of sensitivity parameter δ to perform a sensitivity analysis. Here we assume $\delta \in (-0.03, 0.03)$, which means a unit change in the invalid IV `near4c` will not only changes the potential outcome of `earnings` through the change of exposure `educ`. Since the invalid IV `near4c` violates the assumption (A3), it will also lead to other changes of exposure `educ`, up to 3% of the standard deviations of the potential outcome `earnings`.

```
R> cardfit2=ivmodel(Y=Y, D=D, Z=Z, X=X, k=0.9, deltarange=c(-0.03, 0.03))
R> summary(cardfit2)

Call:
ivmodel(Y = Y, D = D, Z = Z, X = X, k = 0.9, deltarange = c(-0.03,
    0.03))
sample size: 3010
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

First Stage Regression Result:

F=13.25579, df1=1, df2=2994, p-value is 0.00027634
R-squared=0.004407934,   Adjusted R-squared=0.004075405
Residual standard error: 1.940537 on 2995 degrees of freedom
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Coefficients of k-Class Estimators:

              k Estimate Std. Error t value Pr(>|t|)
k-class 0.90000  0.07686    0.01085   7.084 1.74e-12 ***
Fuller  0.99967  0.12750    0.05271   2.419   0.0156 *
LIML    1.00000  0.13150    0.05496   2.393   0.0168 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Alternative tests for the treatment effect under H_0: beta=0.

Anderson-Rubin test:
F=5.415279, df1=1, df2=2994, p-value=0.020028
```

```
95 percent confidence interval:
 [ 0.0248048359651019 , 0.284823593339036 ]

Sensitivity analysis with deltarange [ -0.03 ,  0.03 ]:
non-central F=5.415279, df1=1, df2=2994, ncp=0.4390019, p-value=0.049504
95 percent confidence interval:
 [ 0.000347142651197386 , 0.340944347177351 ]

Conditional Likelihood Ratio test:
Test Stat=5.415279, p-value=0.019961
95 percent confidence interval:
 [0.0248546898484261, 0.284720676631606]
```

Notice that in the third section, the k-class estimator is for the specified $k = 0.9$. In the last section, the p-value is 0.049, suggesting that education has a significant positive effects towards earnings even if the IV may be invalid at a certain degree.

7.3. Functions for Power Formula and Sample Size

Suppose there is an opportunity to design another study of the question addressed by this data set and we would like to find the needed sample size to obtain a certain power.

First we use the function `para()` to get the estimated $\hat{\beta}_{TSLS}, \hat{\gamma}, \hat{\rho}$, $\hat{\sigma}$(`sigmau`), $\hat{\omega}$(`sigmav`) for the original data set. They may be used as a reference to approximate the parameter values in the future study, which are needed for the power formulas in section (6).

```
R> esti=para(cardfit)
R> esti

$gamma
[1] 0.3198989

$beta
[1] 0.1315038

$sigmau
[1] 0.3882648

$sigmav
[1] 1.940213

$rho
[1] -0.2851473
```

For this data, the sample size is `cardfit$n`. Suppose the null hypothesis is $H_0 : \beta = 0$ and the true effect β is the TSLS estimator `esti$beta`. ρ_{ZD} can be calculated by `cor(cardfit$Zadj, cardfit$Dadj)`, $\mathsf{VAR}(D)$ is `var(cardfit$Dadj)` and $\mathsf{VAR}(Z)$ is `var(cardfit$Zadj)`.

If we use TSLS estimator for the hypothesis test, TSLS.power() gives the power 0.668 of rejecting the null hypothesis. TSLS.size() shows that if we want to correctly reject the null with 0.8 power, a sample size of at least 4125 is needed.

```
R> TSLS.power(cardfit$n, esti$beta, cor(cardfit$Zadj, cardfit$Dadj),
R+            esti$sigmau, var(cardfit$Dadj))
```

```
[1] 0.6676418
```

```
R> TSLS.size(0.8, esti$beta, cor(cardfit$Zadj, cardfit$Dadj),
R+           esti$sigmau, var(cardfit$Dadj))
```

```
[1] 4124.388
```

If we use AR test, AR.power() gives the power 0.643 and AR.test() suggest a sample size of at least 4362 for reaching the 0.8 power.

```
R> AR.power(cardfit$n, cardfit$p, cardfit$L, esti$beta, esti$gamma,
R+          var(cardfit$Zadj),esti$sigmau, esti$sigmav, esti$rho)
```

```
[1] 0.6432517
```

```
R> AR.size(0.8, cardfit$p, cardfit$L, esti$beta, esti$gamma,
R+         var(cardfit$Zadj),esti$sigmau, esti$sigmav, esti$rho)
```

```
[1] 4362
```

If the IV may be invalid and we assume the sensitivity parameter $\delta \in (-0.03, 0.03)$ for the sensitivity analysis. Then under the favourable situation $\delta = 0$, ARsens.power() gives the power 0.502 and ARsens.test() suggest a sample size of at least 6723 for reaching the 0.8 power.

```
R> ARsens.power(cardfit$n, cardfit$p, esti$beta, esti$gamma,
R+              var(cardfit$Zadj), esti$sigmau, esti$sigmav, esti$rho,
R+              deltarange=c(-0.03, 0.03), delta=0)
```

```
[1] 0.5022335
```

```
R> ARsens.size(0.8, cardfit$p, esti$beta, esti$gamma,
R+             var(cardfit$Zadj), esti$sigmau, esti$sigmav, esti$rho,
R+             deltarange=c(-0.03, 0.03), delta=0)
```

```
[1] 6723
```

In this example, to achieve the targeted 0.8 power, the TSLS approach requires a lower sample size. The AR test requires a greater sample size, however, as a trade off, it is robust for weak IV while TSLS is not. The sensitivity analysis requires the largest sample size while allowing the IV assumption to be violated to a certain degree.

8. Summary

The package **ivmodel** provides a unified implementation of instrumental variables methods in the case of one endogenous variable. The package contains a general class of estimators, k-class estimators. The package also contains methods that can deal with violations of instrumental variables assumptions, (A2) and (A3). First, for violations of (A2), the package contains two confidence intervals that are fully robust to weak instruments. For (A3), the package contains methods for sensitivity analysis should the exogeneity instrumental variables assumption. The package also contains power formulas to guide designs of future instrumental variables studies. As our data example in Section 7 demonstrated, our package provides an easy and unified way of conducting a comprehensive instrumental variables analysis with data where there is one endogenous variable, along with ways to assess sensitivity to violations of instrumental variables assumptions and to compute power.

References

Amemiya T (1985). *Advanced Econometrics*. Harvard University Press.

Anderson TW, Rubin H (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics*, **20**, 46–63.

Andrews DWK, Moreira MJ, Stock JH (2006). "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression." *Econometrica*, **74**(3), 715–752.

Andrews DWK, Moreira MJ, Stock JH (2007). "Performance of Conditional Wald Tests in IV Regression with Weak Instruments." *Journal of Econometrics*, **139**(1), 116–132.

Angrist JD, Imbens GW, Rubin DB (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, **91**(434), 444–455.

Angrist JD, Krueger AB (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, **106**(4), 979–1014.

Angrist JD, Krueger AB (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *The Journal of Economic Perspectives*, **15**(4), 69–85.

Baiocchi M, Cheng J, Small DS (2014). "Instrumental variable methods for causal inference." *Statistics in Medicine*, **33**(13), 2297–2340.

Baum CF, Schaffer ME, Stillman S (2003). "Instrumental Variables and GMM: Estimation and Testing." *Stata Journal*, **3**(1), 1–31.

Baum CF, Schaffer ME, Stillman S (2007). "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing." *Stata Journal*, **7**(4), 465–506.

Bollen KA (2012). "Instrumental Variables in Sociology and the Social Sciences." *Annual Review of Sociology*, **38**, 37–72.

Bound J, Jaeger DA, Baker RM (1995). "Problems with instrumental variables estimation when the correlation between instruments and the endogenous variable is weak." *Journal of the American Statistical Association*, **90**, 443–450.

Card D (1995). "Using Geographic Variations in College Proximity to Estimate the Return to Schooling." In LN Christofides, EK Grant, R Swidinsky (eds.), *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press.

Chao JC, Swanson NR (2005). "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica*, **73**(5), 1673–1692.

Chen L, Davey Smith G, Harbord RM, Lewis SJ (2008). "Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach." *PLoS Medicine*, **5**(3), e52.

Conley TG, Hansen CB, Rossi PE (2012). "Plausibly Exogenous." *Review of Economics and Statistics*, **94**(1), 260–272.

Davey Smith G, Ebrahim S (2003). "'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?" *International Journal of Epidemiology*, **32**(1), 1–22.

Davey Smith G, Ebrahim S (2004). "Mendelian randomization: prospects, potentials, and limitations." *International Journal of Epidemiology*, **33**(1), 30–42.

Davidson R, MacKinnon JG (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

DiPrete TA, 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.

Dufour JM (2003). "Identification, Weak Instruments, and Statistical Inference in Econometrics." *The Canadian Journal of Economics / Revue canadienne d'Economique*, **36**(4), 767–808.

Fox J, Nie Z, Byrnes J (2014). ***sem**: Structural Equation Models*. R package version 3.1-5, URL http://CRAN.R-project.org/package=sem.

Freeman G, Cowling BJ, Schooling CM (2013). "Power and Sample Size Calculations for Mendelian Randomization Studies Using One Genetic Instrument." *International journal of epidemiology*, **42**(4), 1157–1163.

Fuller WA (2006). *Measurement Error Models*. Wiley.

Gaure S (2013). "**lfe**: Linear group fixed effects." *The R Journal*, **5**(2), 104–117. URL http://journal.r-project.org/archive/2013-2/gaure.pdf.

Gennetian LA, Magnuson K, Morris PA (2008). "From Statistical Associations to Causation: What Developmentalists Can Learn From Instrumental Variables Techniques Coupled With Experimental Data." *Developmental Psychology*, **44**(2), 381.

Holland PW (1988). "Causal Inference, Path Analysis, and Recursive Structural Equations Models." *Sociological Methodology*, **18**(1), 449–484.

Jiang Y, Zhang N, Small DS (2015). "Sensitivity Analysis and Power for Instrumental Variable Studies." *Biometrics, under review*.

Kang H, Zhang A, Cai TT, Small DS (2015). "Instrumental Variables Estimation with Some Invalid Instruments and its Application to Mendelian Randomization." *Journal of the American Statistical Association*.

Kleiber C, Zeileis A (2008). *Applied Econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2, URL http://CRAN.R-project.org/package=AER.

Kolesár M, Chetty R, Friedman JN, Glaeser EL, Imbens GW (2011). "Identification and Inference with Many Invalid Instruments." *Technical report*, National Bureau of Economic Research.

Kolesár M, Chetty R, Friedman JN, Glaeser EL, Imbens GW (2013). "Identification and Inference with Many Invalid Instruments." *National Bureau of Economic Research*, p. No. w17519.

Krueger AB (1999). "Experiemtnal Estimates of Education Production Functions." *the Quarterly Journal of Economics*, **114**(2), 497–532.

Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G (2008). "Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology." *Statistics in Medicine*, **27**(8), 1133–1163.

Mariano RS (2003). *Simultaneous Equation Model Estimators: Statistical Properties and Practical Implications*, pp. 122–141. Blackwell Publishing Ltd.

Mikusheva A (2010). "Robust Confidence Sets in the Presence of Weak Instruments." *Journal of Econometrics*, **157**(2), 236–247.

Moreira MJ (2001). *Tests with Correct Size When Instruments Can Be Arbitrarily Weak*. Center for Labor Economics, University of California, Berkeley.

Moreira MJ (2003). "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica*, **71**(4), 1027–1048.

Morgan SL, Winship C (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.

Murray MP (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments." *The Journal of Economic Perspectives*, **20**(4), 111–132.

Nelson CR, Startz R (1990). "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One." *The Journal of Business*, **63**(1), S125–40.

Permutt T, Hebel JR (1989). "Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight." *Biometrics*, **45**(2), 619–622.

Rosenbaum PR (2010). *Design of Observational Studies*. Springer Series in Statistics. Springer, New York.

Rubin DB (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, **66**(5), 688–701.

Sargan JD (1958). "The Estimation of Economic Relationships using Instrumental Variables." *Econometrica*, pp. 393–415.

Sexton M, Hebel RJ (1984). "A Clinical Trial of Change in Maternal Smoking and its Effect on Birth Weight." *Journal of the American Medical Association*, **251**(7), 911–915.

Small DS (2007). "Sensitivity Analysis for Instrumental Variables Regression With Overidentifying Restrictions." *Journal of the American Statistical Association*, **102**(479), 1049–1058.

Sovey AJ, Green DP (2011). "Instrumental Variables Estimation in Political Science: A ReadersâĂŹ Guide." *American Journal of Political Science*, **55**(1), 188–200.

Staiger D, Stock JH (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, **65**(3), 557–586.

Stock JH, Wright JH, Yogo M (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics*, **20**(4).

Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day INM, Palmer LJ, Hattersley AT, Ebrahim S, Lowe G, Rumley A, Davey Smith G (2005). "C-reactive Protein and its Role in Metabolic Syndrome: Mendelian Randomisation Study." *The Lancet*, **366**(9501), 1954–1959.

Wooldridge JM (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. edition. MIT press.

Affiliation:

Yang Jiang, Hyunseung Kang, Dylan S. Small
Department of Statistics
The Wharton School
University of Pennsylvania
400 Jon M. Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104-6340, United States of America
E-mail: yajiang@wharton.upenn.edu, hskang@stanford.edu, dsmall@wharton.upenn.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd