Strengthening Weak Instruments by Modeling Compliance*

Marc Ratkovic† Yuki Shiraito‡

November 25, 2014

Abstract

Instrumental variable estimation is a long-established means of reducing endogeneity bias in regression coefficients. Researchers commonly confront two problems when conducting an IV analysis: the instrument may be only weakly predictive of the endogenous variable, and the estimates are valid only for observations that comply with the instrument. We introduce Complier Instrumental Variable (CIV) estimation, a method for estimating who complies with the instrument. CIV uses these compliance probabilities to strengthen the instrument through up-weighting estimated compliers. As compliance is latent, we model each observation's density as a mixture between that of a complier and a non-complier. We derive a Gibbs sampler and Expectation Conditional Maximization algorithm for estimating the CIV model. A set of simulations shows that CIV performs favorably relative to several existing alternative methods, particularly in the presence of small sample sizes and weak instruments. We then illustrate CIV on data from a prominent study estimating the effect of property rights on growth. We show how CIV can strengthen the instrument and generate more reliable results. We also show how characterizing the compliers can help cast insight into the underlying political dynamic.

*We thank Kosuke Imai, John Londregan, Luke Keele, Jacob Montgomery, Scott Abramson, Michael Barber, Graeme Blair, Kentaro Hirose, In Song Kim, Alex Ruder, Carlos Velasco Rivera, Jaquilyn Waddell Boie, Meredith Wilf, and participants of Princeton Political Methodology Seminar for useful comments.

†Assistant Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 608–658–9665, Email: ratkovic@princeton.edu, URL: http://www.princeton.edu/∼ratkovic

‡Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08544. Email: shiraito@princeton.edu

1 Introduction

Social scientists commonly confront the problem of endogeneity, whereby the dependent and independent variables have a direct causal effect on each other. In the presence of this endogeneity, the regression coefficient is a biased estimate of the causal effect. An *instrumental variable* (IV), with a direct effect on the endogenous variable and no direct effect on the outcome, can be used to reduce this bias (Sovey and Green, 2011). The instrumental variable provides an experimental handle in the observational study, allowing the researcher to follow the effect of an exogenous shock through the endogenous variable and onto the outcome. Instruments are most commonly incorporated through two-stage least squares estimation (TSLS). In the first stage of TSLS, the endogenous variable is regressed on the instrument. In the second stage, the dependent variable is regressed on the fitted values from the first stage. As the first stage estimates are not polluted by the endogeneity, the second stage estimate recovers an unbiased estimate of the true causal effect.

Standard IV methods pose two problems to the applied researcher. First, the instrument may be only weakly predictive of the endogenous variable. With weak instruments, the noise in the first-stage estimate can overwhelm the instrument's signal, producing estimates that are, at best, noisy and, at worst, arbitrarily misleading (Bound *et al.*, 1995; Staiger and Stock, 1997; Hahn and Hausman, 2003; Hahn *et al.*, 2011; Kuersteiner and Okui, 2010). If the instrument is sufficiently weak, the sampling distribution of the TSLS estimate may have such fat tails that the asymptotic normal approximation dramatically understates the width of the confidence interval. Second, the instrument may not have a direct effect on *every* observation. IV methods identify a causal effect only for the observations affected by the instrument ("compliers," see Imbens and Angrist, 1994; Angrist *et al.*, 1996; Imbens and Rubin, 1997b). The resultant estimates are of a local average treatment effect (LATE), averaged across these compliers. Estimating local effects, while informative (Imbens, 2010), raises clear questions of scope: for what observations, exactly, has the researcher estimated a causal effect (Deaton, 2010; Heckman and Urzua, 2010)?

We introduce a method, *complier instrumental variable estimation* (CIV), that strengthens instruments through estimating, and down-weighting, non-compliers in the sample. Since the causal effect is not identified for non-compliers, maintaining these observations simply weakens the instrument. We show that estimating compliance weights, and then down-weighting estimated non-compliers, serves to both strengthen instruments and produces more precise and reliable second-stage estimates. Furthermore, estimating the compliance model allows the researcher to make claims about which types of observations complied with the instrument.

Specifically, we model the data-generation process as a system with three equations. The first, the *compliance model*, determines each observation's latent type, complier or non-complier, as a function of observed covariates (Frangakis and Rubin, 2002; Hirano *et al.*, 2000). The second equation, the *first-stage model*, varies by compliance type. For compliers, the endogenous variable is a function of exogenous covariates and the instrument. For non-compliers, the endogenous variable is only a function of the exogenous covariates, and not of the instrument. The third equation, the *second-stage model*, models the outcome as a function of the endogenous variable and exogenous covariates. Each observation's compliance status is latent, inducing a mixture structure in the likelihood. To fit the mixture we derive and implement a Gibbs Sampler. We also derive an implementation within the ECM framework.

The basic innovation of CIV is to add the extra layer of the compliance model to the standard simultaneous equation model, in order to strengthen the instrument and model compliers. We borrow the insight from the potential outcomes literature that only some observations may comply with the instrument. We build on these methods (Hirano *et al.*, 2000; Imbens and Rubin, 1997a) by placing them within a structural equation framework, allowing for continuous endogenous variables and instruments. While these methods were developed in the experimental setting where the weak instrument problem is not a concern, we show how they are useful to strengthen weak instruments as well. Our mechanism for handling compliance status is similar, algebraically, to switching regression models (Heckman and Hotz, 1989; Heckman, 1978; Roy, 1951). The key difference is that compliance status is latent, while the switching regressions were developed for observed treatment values. Similar to Conley *et al.* (2012) we embed IV analysis in a Bayesian framework (see, e.g., Chib, 2003; Crespo-Tenorio and Montgomery, 2013; Imbens and Rubin, 1997a; Heckman, 1978). Where Conley *et al.* (2012) relax the exclusion restriction, we make this assumption and instead allow a subset of observations to be affected by the instrument. Finally, we work within the cross-sectional setting, so as to avoid concerns about the number of instruments growing in the sample size (e.g., Arellano and Bond, 1991).

We illustrate the method's use and efficacy on a prominent dataset and a set of simulations. We first reanalyze the data in Acemoglu, Johnson, and Robinson (Acemoglu *et al.* 2001, hereafter AJR). In a cross-national study, AJR argue that early European settler mortality can serve as an instrument for the effect of private property rights on current growth rates. We show that CIV produces more precise and reliable estimates than both standard TSLS and several other methods designed to handle weak instruments. Second, we present simulation evidence comparing CIV to TSLS, a rank-based estimator that is robust to outliers (Imbens and Rosenbaum, 2005; Betz, 2013), and a jack-knife estimator (Hahn *et al.*, 2004). CIV generally outperforms the jack-knife and TSLS estimators, while performing comparably to the rank-based estimator.

The paper progresses in four parts. First, we introduce the proposed method. Second, we apply the method to the data from a recent study. Third, we present evidence from a simulation study. Fourth, a conclusion follows.

2 The Proposed Method

In this section, we introduce the proposed method. Second, we discuss two methods for estimation. Finally, we compare the proposed method to several existing IV methods.

2.1 The Model

Assume a simple random sample of size N, from which we observe: an outcome, Y_i; a set of K_X exogenous covariates, X_i; a set of K_D endogenous covariates D_i; and a set of K_Z instruments, Z_i, and that all covariates have been standardized to mean zero and standard deviation one. We assume that K_X, K_Z, and K_D are all finite. We will let 0_A denote a vector of A zeroes, vectors will be denoted by a subscript, and matrices will be in bold.

The standard model calling for IV analysis can be written as

$$Y_i = D_i^\top \beta + X_i^\top \gamma + \epsilon_i \tag{1}$$

$$D_i = Z_i^\top \delta + X_i^\top \theta + \eta_i \tag{2}$$

where β, γ, δ, and θ are parameters associated with the observed covariates. Correlation between D_i and ϵ_i, such that $\mathbb{E}(D_i \epsilon_i) \neq 0_{K_D}$, renders the least squares estimate of β biased. This correlation may arise due to endogeneity, as when Y_i and D_i co-determine each other, or when D_i is measured with some error. To generate an unbiased estimate of β, we assume we have a set of instruments, Z_i that satisfy three conditions: the exclusion restriction, $\mathbb{E}(Z_i \epsilon_i) = 0_{K_Z}$; the instruments are not weak, $(\delta \neq 0)$; and the span of D_i is at least that of Z_i.

Several assumptions are embedded in this model. First, the model assumes a constant marginal effect of Z_i on D_i. This implies that the instrument has the same marginal impact on each observation (δ). Second, the model assumes monotonicity in the effect of the instrument on the endogenous variable, δ. In practice, though, an instrument may have a differential impact on influencing the endogenous variable across observations. For example, there may be a reason to believe that the probability of an impact of rainfall on conflict (Miguel *et al.*, 2004), geographic distance on trade (Frankel and Romer, 1999), early settler mortality on current risk of property expropriation (Acemoglu *et al.*, 2001), and so on, varies from country to country in a systematic manner.

In order to account for the possibility that not every observation is affected by the instrument, we introduce a latent variable, C_i, which takes a value of 1 if the observation is influenced by the instrument, and a 0 otherwise. The probability of compliance is taken as a function of K_W observed covariates, $\widetilde{W}_i$, and corresponding parameter α. Incorporating compliance produces the following system of equations:

$$Y_i = D_i^\top \beta + X_i^\top \gamma + \epsilon_i \tag{3}$$

$$D_i = \begin{cases} \delta_0^C + Z_i^\top \delta + X_i^\top \theta + \eta_i; & C_i = 1 \\ \delta_0^{NC} + X_i^\top \theta + \eta_i; & C_i = 0 \end{cases} \tag{4}$$

$$\Pr(C_i = 1) = \Phi\left(W_i^\top \alpha\right) \tag{5}$$

The compliance probabilities serve two distinct roles. First, estimating α allows the researcher to characterize the observations for which she has estimated a causal effect (β). Second, including these probabillities strengthens the instrument. Intuitively, the researcher would never include observations in the sample *ex ante* for which the instrument has no effect; including such observations simply adds noise to the estimation. Similarly, *ex post*, we would want to down-weight the observations that we expect to not have been affected by the instrument.

Modeling the latent types, or "principal strata (Frangakis and Rubin, 2002)," allows the researcher to make meaningful claims about compliance types and lead to more efficient effect estimation (Esterling *et al.*, 2011; Barnard *et al.*, 2003; Imbens and Rubin, 1997a,b; Hirano *et al.*, 2000).

Compliance status bears an algebraic similarity to "switching regression" models (Heckman and Hotz, 1989; Heckman, 1978; Roy, 1951). The primary distinction is that compliance status is latent, where switching models are used to model observed choices (e.g., Borjas, 1988). Our notion of compliance is exactly that of Imbens and Angrist (1994, , 467). An observation complies if its value of the endogenous variable is affected by the instrument. We diverge from Imbens and Angrist (1994) and Angrist *et al.* (1996) in two ways. First, we are working within a linear structural equation framework, which allows us to better accommodate continuous instruments and endogenous variables. Second, we use the compliance probabilities to strengthen a weak instrument. Finally, by weak instrument, we are referring to the case where the number of instruments K_Z is fixed, but the instrument is weak in the sense of Staiger and Stock (1997), where the concentration parameter is constant as $N \to \infty$. This type of scenario is encountered when researchers face a dataset in which the first stage regression has a low F-statistic even with a large sample size.

The Likelihood and Prior Densities The density of a single outcome, (Y_i, D_i), is a mixture between the compliance and non-compliance densities, as

$$\begin{aligned}\Pr((Y_i, D_i)|\Theta, X_i, Z_i, W_i, \alpha) &= \Pr((Y_i, D_i)|\Theta_C, X_i, Z_i, C_i = 1) \Pr(C_i = 1|W_i, \alpha) \\ &+ \Pr((Y_i, D_i)|\Theta_{NC}, X_i, Z_i, C_i = 0) \Pr(C_i = 0|W_i, \alpha) \qquad (6)\end{aligned}$$

where $\Theta_C = [\beta, \gamma, \delta_0^C, \delta, \theta]$, $\Theta_{NC} = [\gamma, \delta_0^{NC}, \theta]$, and $\Theta = \Theta_C \bigcup \Theta_{NC}$.

This mixture structure leads to the following log-likelihood:

$$
\begin{aligned}
l(\Theta, \alpha | X_i, Z_i, W_i) &= \sum\nolimits_{i=1}^{N} \log \left(\Pr\left((Y_i, D_i) \,|\Theta, X_i, Z_i, W_i, \alpha \right) \right) \\
&= \sum\nolimits_{i=1}^{N} \log \Bigg\{ \Pr((Y_i, D_i)|\Theta_C, X_i, Z_i, C_i = 1) \Pr(C_i = 1|W_i, \alpha) \\
&\qquad + \Pr((Y_i, D_i)|\Theta_{NC}, X_i, Z_i, C_i = 0) \Pr(C_i = 0|W_i, \alpha) \Bigg\}
\end{aligned} \tag{7}
$$

We complete the specification of the likelihood, where $\Phi(\cdot)$ denotes the distribution of a standard normal random variable, with

$$C_i \big| W_i, \alpha \overset{\text{i.i.d.}}{\sim} \text{Bern}(\Phi(\mathrm{W_i^\top} \alpha)) \tag{8}$$

$$(Y_i - D_i^\top \beta, D_i)^\top \big| \Theta_C, X_i, Z_i, C_i = 1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_C, \Omega) \tag{9}$$

$$(Y_i - D_i^\top \beta, D_i)^\top \big| \Theta_{NC}, X_i, Z_i, C_i = 0 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_{NC}, \Omega) \tag{10}$$

$$\mu_C = \begin{bmatrix} X_i^\top \gamma \\ \\ \delta_0^C + Z_i^\top \delta + X_i^\top \theta \end{bmatrix}; \;\; \mu_{NC} = \begin{bmatrix} X_i^\top \gamma \\ \\ \delta_0^{NC} + X_i^\top \theta \end{bmatrix}; \Omega = \mathbb{E}\left(\begin{bmatrix} \epsilon_i \\ \\ \eta_i \end{bmatrix} \begin{bmatrix} \epsilon_i \\ \\ \eta_i \end{bmatrix}^\top \right) \tag{11}$$

We place non-informative priors over $[\beta, \gamma]$ and $[\delta, \theta]$ of mean zero and variance 10^4. We assume

$$\alpha \sim \mathcal{N}(0, \sigma_\alpha^2 I_{K_W}) \tag{12}$$

$$\sigma_\alpha^{-2} \sim \Gamma(1, 1) \tag{13}$$

for the parameters in the compliance model.

Extensions The model admits several useful, and straightforward, extensions. First, our framework accommodates any distribution that can be expressed in terms of latent normal random variables. For example, if Y_i or D_i were binomial, we could model the latent variables Y_i^* or D_i^* such that $\phi(Y_i^*)$ or $\Phi(D_i^*)$ generate the observed Y_i or D_i (Albert and Chib, 1993; Heckman, 1979). Generalizations to any outcome in the exponential family is also possible (Dunson and Herring, 2005; Muthen, 1984). Furthermore, we can handle any outcome that can be expressed as a scale-mixture of normals. The method, therefore, can be extended to include both t- densities, if we want to mix the normals with the same mean, or Bayesian Model Averaging, if we want to mix normals with different means (Karl and Lenkoski, 2012).

2.2 Estimation

We offer two different estimation techniques: a conditional EM estimator and a Gibbs Sampler. We implement the Gibbs Sampler in the examples below, as it provides proper posterior estimates of our parameters of interest and makes it easier to obtain uncertainty estimates. The software will be made free for download as the package `CIV` in the **R** statistical language (R Development Core Team, 2012). For completeness, we also derive and present a conditional EM algorithm.

Gibbs Sampler The posterior distributions of the model's parameters can be found through Markov Chain Monte Carlo simulation. We augment the posterior with indicator variables for whether each observation is a complier (Imbens and Rubin, 1997a; Meng and van Dyk, 1999). After augmentation, all conditional posteriors can be derived analytically, allowing for an efficient Gibbs sampler.

The algorithm begins with $C_i = 1$ for each observation, $\alpha = \delta_0^{NC} = 0$, and all other parameters set to their maximum likelihood estimates. The parameters $[\beta, \gamma]$ and $[\delta_0^C, \delta_0^{NC}, \delta, \theta]$ are given non-informative Normal priors with mean zero and variance 10,000. The parameters in α are given a Normal distribution, with mean zero and variance σ_α^2, and $1/\sigma_\alpha^2$ is given a Gamma prior with rate and scale 1. Ω is given an inverse Wishart prior. A single cycle through the Gibbs sampler proceeds as follows:

1. Update regression parameters given estimated compliers.

(a) Construct augmented covariate matrices,

$$\widetilde{X}_1 = [D_i^\top, X_i^\top]; \quad \widetilde{\beta} = [\beta, \gamma] \tag{14}$$

$$\widetilde{X}_2 | C_i = [C_i, 1 - C_i, Z_i C_i, X_i]; \quad \widetilde{\delta} = [\delta_0^C, \delta_0^{NC}, \delta, \theta] \tag{15}$$

(b) Update $\widetilde{\beta}$ and $\widetilde{\delta}$ conditional on $\widetilde{X}_1$ and $\widetilde{X}_2$ (see Karl and Lenkoski (2012) or Rossi *et al.* (2006) for details). Update the new residuals, ϵ_i and η_i.

(c) Update $\Omega \sim InvWish(N+1, S+I)$ where N is the sample size,

$$S = \sum_{i=1}^{N} \begin{bmatrix} \epsilon_i \\ \\ \eta_i \end{bmatrix} \begin{bmatrix} \epsilon_i \\ \\ \eta_i \end{bmatrix}^\top, \tag{16}$$

and I is the identity matrix of same dimension as S.

2. Find likelihood of compliance and non-compliance, $\Pr(Y_i - D_i^\top \beta, D_i | \Theta_C, X_i, C_i = 1)$ and $\Pr(Y_i - D_i^\top \beta, D_i | \Theta_{NC}, X_i, C_i = 0)$

3. Update prior compliance probabilities

- Sample $C_i^* \sim TN(W_i^\top \alpha, 1, -1^{(1-C_i)})$ where $TN(\mu, V, \pm)$ denotes the truncated normal with mean vector μ, variance V, and truncated below (-) or above (+) 0.
- Sample $\alpha \sim \mathcal{N}\left((W^\top W + \sigma_\alpha^2 I_K)^{-1} W^\top C_i^*, (W^\top W + \sigma_\alpha^2)^{-1}\right)$
- Sample $1/\sigma_\alpha^2 \sim \Gamma(k+1, \sum_{i=1}^k \alpha_i^2 + 1)$
- Update $Pr(C_i = 1 | W_i, \alpha) = \Phi(W_i^\top \alpha)$

4. Update posterior of compliance probabilities

$$\Pr(C_i = 1 | Y_i, D_i, W_i, X_i, \Theta) =$$
$$\frac{\Pr(Y_i - D_i^\top \beta, D_i | \Theta_C, X_i, C_i = 1) \Pr(C_i = 1 | W_i, \alpha)}{\Pr(Y_i - D_i^\top \beta, D_i | \Theta_C, X_i, C_i = 1) \Pr(C_i = 1 | W_i, \alpha) + \Pr(Y_i - D_i^\top \beta, D_i | \Theta_{NC}, X_i, C_i = 0) \Pr(C_i = 0 | W_i, \alpha)} \tag{17}$$

5. Sample $C_i \sim Bern\big(\Pr(C_i = 1|Y_i, D_i, W_i, X_i, \Theta)\big)$

We offer two different measures of uncertainty of the causal effect, β. The first is the posterior credible interval, sampled as described above. This posterior credible interval, though, will not have nominal coverage. The $1-\alpha\%$ credible interval, by construction, will not to contain the true value across $1-\alpha\%$ samples. We illustrate this phenomenon in simulations below. For researchers concerned with coverage, we implement an estimate of the posterior mean of the $1-\alpha\%$ confidence interval for the compliers, the CIV-augmented TSLS interval. We obtain this estimate by calculating the $1-\alpha\%$ TSLS estimate of β for the estimated compliers ($C_i = 1$), across Gibbs samples. We show in simulations that these confidence intervals achieve coverage rates closer to nominal than the CIV posterior interval.

Conditional EM A common estimation approach to mixture models is the Expectation-Maximization (EM) algorithm. Because the difficulty of estimation is caused by each observation's latent class indicator, the EM algorithm treats this variable as missing data. The complete-data log-likelihood of mixture models generally has a simple form so that its maximization has a closed-form solution.

Unfortunately, the maximization step of the standard EM algorithm for our model will require challenging numerical approximation, since each mixture component is a simultaneous equation model. We take advantage of the fact that some of the parameters have closed-form maxima. We replace the M-step with a set of conditional maximization (CM) steps, allowing us to minimize the numerical optimization.

The complete-data log-likelihood is given by

$$\begin{aligned} & l_{\text{comp}}(\Theta, \alpha, \Omega | X, Z, \widetilde{X}, C) \\ = & \sum_{i=1}^{N} \sum_{c=0}^{1} \mathbf{1}_{\{C_i = c\}} \log \Big\{ \Pr((Y_i, D_i)|\Theta, \Omega, X_i, Z_i, C_i = c) \Pr(C_i = c|\widetilde{X}_i, \alpha) \Big\} \end{aligned} \tag{18}$$

In the E-step of the lth iteration of the ECM algorithm, we will compute the Q-function, which is the conditional expectation of (18) conditional on data and the parameter estimates in the $(l-1)$th iteration. We obtain

$$
\begin{aligned}
&Q(\Theta, \Omega, \alpha|\Theta^{(l-1)}, \Omega^{(l-1)}, \alpha^{(l-1)}) \\
=&\sum_{i=1}^{N} \left[\hat{\pi}_i^{(l)} \log \left\{ \Pr((Y_i, D_i)|\Theta_C, \Omega, X_i, Z_i, C_i = 1) \Pr(C_i = 1|\widetilde{X}_i, \alpha) \right\} \right. \\
&\left. + (1 - \hat{\pi}_i^{(l)}) \log \left\{ \Pr((Y_i, D_i)|\Theta_N C, \Omega, X_i, Z_i, C_i = 0) \Pr(C_i = 0|\widetilde{X}_i, \alpha) \right\} \right] \qquad (19)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\pi}_i^{(l)} &= \mathbb{E}\Big[\mathbf{1}_{\{C_i=1\}}|\Theta^{(l-1)}, \Omega^{(l-1)}, \alpha^{(l-1)}, Y_i, D_i, Z_i, X_i, \widetilde{X}_i \Big] \\
&= \mathrm{P}\Big\{ C_i = 1|\Theta^{(l-1)}, \Omega^{(l-1)}, \alpha^{(l-1)}, Y_i, D_i, Z_i, X_i, \widetilde{X}_i \Big\} \\
&= \frac{\mathrm{P}\{C_i = 1|\alpha^{(l-1)}\}\mathrm{P}((Y_i, D_i)|\Theta_C^{(l-1)}, \Omega^{(l-1)}, X_i, Z_i, C_i = 1)\}}{\sum_{c=0}^{1} \mathrm{P}\{C_i = c|\alpha^{(l-1)}\}\mathrm{P}((Y_i, D_i)|\Theta^{(l-1)}, \Omega^{(l-1)}, X_i, Z_i, C_i = c)\}}. \qquad (20)
\end{aligned}
$$

In the maximization step, we partition the parameter space into four distinct subsets and implement maximization of each subset of parameters fixing the other parameters at the current value. The four subsets are $\{\alpha\}$, $\{\beta, \gamma\}$, $\{\delta_0^C, \delta_0^{NC}, \delta, \theta\}$, and $\{\Omega\}$.

CM-step 1 We compute $\alpha^{(l)}$ first. Observing the relevant terms in the Q-function (19), $\alpha^{(l)}$ will be given by

$$
\alpha^{(l)} = \underset{\alpha}{\operatorname{argmax}} \sum_{i=1}^{N} \left\{ \hat{\pi}_i^{(l)} \log \Phi(\widetilde{X}_i^\top \alpha) + (1 - \hat{\pi}_i^{(l)}) \log(1 - \Phi(\widetilde{X}_i^\top \alpha)) \right\}
$$

This step will be implemented by numerical approximation.

CM-step 2 Second, we will obtain the estimates of the second stage coefficients, $\{\beta, \gamma\}$.

Factoring the joint density of (Y, D) into the marginal density of D and the conditional density of Y given D, maximization of the Q-function with respect to $\{\beta, \gamma\}$ reduces to the maximization problem of the conditional density of Y given D. Noting that this conditional density involves the residual correction term, $\{\beta^{(l)}, \gamma^{(l)}\}$ are given by the weighted regression of $\left(\left(Y - \frac{\sigma_\epsilon^{(l-1)}}{\sigma_\eta^{(l-1)}}\rho^{(l-1)}(D - \delta_0^{C(l-1)} - Z\delta^{(l-1)} - X\theta^{(l-1)}\right)^\top, \left(Y - \frac{\sigma_\epsilon^{(l-1)}}{\sigma_\eta^{(l-1)}}\rho^{(l-1)}(D - \delta_0^{NC(l-1)} - X\theta^{(l-1)}\right)^\top\right)^\top$ on $(1_N, D, X)$ with weights $(\hat{\pi}^{(l)\top}, (1-\hat{\pi}^{(l)})^\top)^\top$. See Appendix A.1 for mathematical details.

CM-step 3 In the thrid CM-step, we will estimate the first stage coefficients, $\{\delta_0^C, \delta_0^{NC}, \delta, \theta\}$. The solution to the maximization problem of the joint normal density described in Equations 9, 10, and 11 with respect to these parameters is given by the weighted regression of $(D^\top, D^\top)^\top$ on $((1_N, 0_N, Z, X)^\top, (0_N, 1_N, \mathbf{O}_{N\times K_Z}, X)^\top)$ with weights $(\hat{\pi}^{(l)\top}, (1-\hat{\pi}^{(l)})^\top)^\top$.

CM-step 4 The last CM-step obtains the estimate of $\{\Omega\}$. Since all the coefficient parameters are fixed in this step, the solution is straightforwardly given by the mean of the weighted sum of the residual squares with weights being $(\hat{\pi}^{(l)\top}, (1-\hat{\pi}^{(l)})^\top)^\top$. See Appendix A.1 for more details.

2.3 Comparison to Existing Methods

In this section, we discuss the relationship between CIV and several existing methods. For a more complete overview of the instrumental variable literature, see Flores-Lagunes (2007). CIV offers the prospect of strengthening instruments through up-weighting estimated compliers. Existing methods for addressing weak instruments work through some form of smoothing or transformation so as to reduce the impact of outliers and dampen the tails of the estimate's distribution. CIV offers a new form of leverage, through explicitly modeling the compliers.

Working in the potential outcomes framework, Imbens and Angrist (1994, , Theorem 2, 471) derive the IV estimand for discrete instruments, while Hirano *et al.* (2000) and Imbens and Rubin (1997a) discuss methods for estimating compliance probabilities in the presence of binary endogenous variables and instruments. Abadie (2003) develops a weighting method, where the outcome model is weighted by individual-level estimated compliance probabilities. CIV incorporates these basic insights into the structural equation framework, modeling compliance and using the estimated compliers (at each Gibbs step) to estimate the underlying causal effect. While these methods focus primarily on non-parametric identification of causal effects, we make parametric assumptions so as to better accommodate continuous endogenous variables and instruments.

Imbens and Rosenbaum (2005) propose a solution to the weak instrument problem, where they replace the TSLS moment conditions with corresponding rank-based quantities (see also Betz, 2013). These estimators, dealing with ranks rather than the true values, are robust to outliers. To derive this estimator, first consider the moment conditions for two-stage least squares, that Z_i and X_i are uncorrelated with ϵ_i:

$$\mathbb{E}\left(\begin{bmatrix} Z_i \\ X_i \end{bmatrix}(\epsilon_i\right) = \mathbb{E}\left(\begin{bmatrix} Z_i \\ X_i \end{bmatrix}(Y_i - D_i^\top \beta - X_i\gamma)\right) = 0_{K_X+K_Z}. \tag{21}$$

Setting the moment conditions' sample analog to zero produces the standard TSLS estimator. Imbens and Rosenbaum replace this moment condtion with a robust version, assuming that the *ranks*, as opposed to their observed values, are uncorrelated:

$$\mathbb{E}\Big\{ (rnk\,(Z_i - X_i\beta_Z))\,(rnk\,(Y_i - D_i\beta - X_i\gamma)) \Big\} = \left(\frac{N+1}{2}\right)^2 \tag{22}$$

where $rnk()$ denotes the rank function. The righthand side is the expected rank if the two lefthand terms were, in truth, independent. A central limit theorem for rank statistics allows for the construction of a t-statistic that can be inverted to generate a confidence interval (Hajek and Sidak, 1967). The original papers do not include additional covariates, X_i, in their formulation of the problem; to the best of our knowledge, we are the first to do so.

Several jackknife methods have also been proposed to reduce the bias of TSLS. The jackknife estimate drops each observation in turn, fits TSLS to the remaining $N-1$ observations, and uses this to construct an estimate. Denoting the TSLS estimate with the j^{th} observation omitted $\widehat{\beta}^{[-j]}$, the jackknife estimate is

$$N\hat{\beta}_{TSLS} - \frac{N-1}{N}\sum_{j=1}^{N}\widehat{\beta}^{[-j]} \tag{23}$$

Angrist *et al.* (1999) propose using the jackknife for only the first stage, though later work found poor performance of this estimator (Davidson and MacKinnon, 2006). For that reason, we jackknife both stages in the simulations below (Hahn and Hausman, 2003).

CIV is also related to the REQML estimator proposed by (Chamberlain and Imbens, 2004). Like REQML, CIV starts with the LIML likelihood function, placing a multivariate normal over the joint error distribution. REQML was designed for the case when there are a large number of low-information instruments. The authors place a random effects structure over the instruments so as to avoid the incidental parameter problem. CIV is similar, in that we also use a form of Bayesian smoothing with our priors, but we are instead focusing on the single-instrument setting. Extending CIV to the many-instrument setting is future work.

3 Strengthening Instruments: Revisiting Acemoglu, Johnson, and Robinson (2001)

The applied researcher may encounter a situation where a reasonable instrument exists, but the dataset is small and the instrument weak. We consider just such a case here, and show that CIV estimation strengthens the instrument, reduces uncertainty in the estimates, and adds new insight into the original problem.

We reanalyze a prominent study that assessed whether strengthening property rights can have a *causal* impact on growth. In a prominent study, AJR argued that European settler mortality rates could serve as an instrument for estimating the effect of current property rights strength on current growth. The authors showed that European colonizers were more likely to establish extractive institutions in areas with high settler mortality rates, while establishing stronger institutions in areas with low settler mortality rates. These colonial institutional decisions persist through today, providing a direct link between settler mortality and current property right regimes, but no direct effect of settler mortality on growth. The validity of the data has since been challenged (Albouy, 2012; Acemoglu *et al.*, 2012). For the purposes of this analysis, we will take the data as both given and accurate.

The original study has 64 observations, dictated by the availability of settler mortality data. Several of AJR's robustness checks reanalyze the model on subsets of the whole data; we consider only the full dataset. The outcome is 1995 PPP GDP per capita, logged, and the endogenous variable is a measure of protection against expropriation. Other exogenous variables include latitude, indicators for former French or British colony, indicators for continent, and an indicator for whether the country has a French legal origin.

Figure 1 presents the estimates of the causal effect from 4 different methods: ordinary least squares (OLS), two-stage least squares (TSLS), CIV, the rank estimator of Imbens and Rosenbaum (2005, , RANK) and a jack-knifed estimator (JACK). For each estimator, the dashed-line is the 95% confidence interval constructed from analytic standard errors (OLS, TSLS, RANK, JACK), and is the 95% posterior interval for CIV. The solid line is the bias-corrected 95% boot-strapped confidence

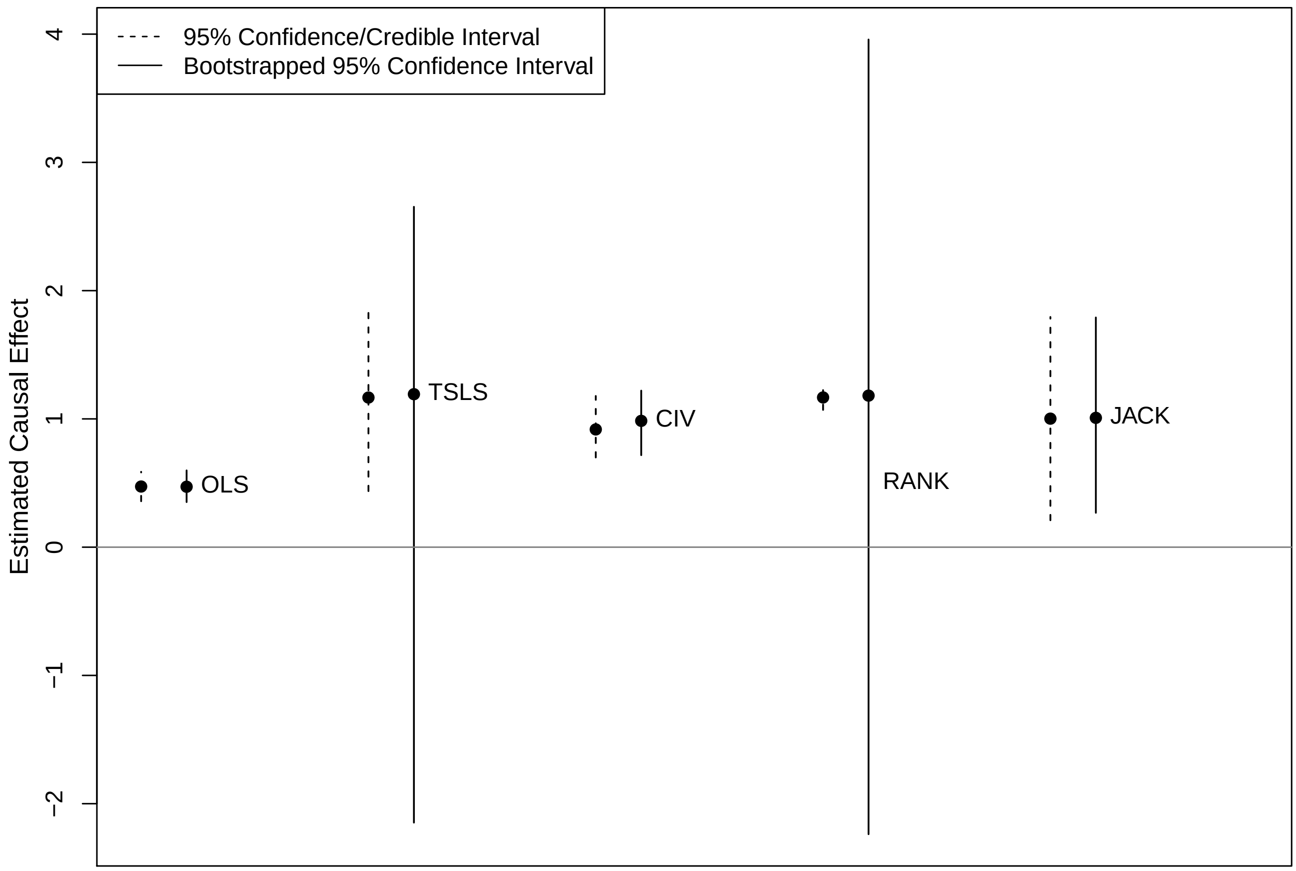

Figure 1: **Estimated Causal Effect of Property Rights on Economic Growth, by Method.** We present estimates of the causal effect from 4 different methods: ordinary least squares (OLS), two-stage least squares (TSLS), CIV, the rank estimator of Imbens and Rosenbaum (2005, , RANK) and a jack-knifed estimator (JACK). For each estimator, the dashed-line is the 95% confidence interval constructed from analytic standard errors (OLS, TSLS, RANK, JACK), and is the 95% posterior interval for CIV. The solid line is the bias-corrected 95% boot-strapped confidence interval of the mean (OLS, TSLS, RANK, JACK) and the posterior mean for CIV. The analytic and bootstrapped confidence intervals differ dramatically for TSLS, indicative of a weak instrument. For OLS and JACK, the bootstrapped interval and analytic intervals are close, suggesting the normal approximation is more valid for these estimators. CIV agrees with the jackknife and rank point estimate, but offers an uncertainty interval more in-line with the bootstrap interval.

interval of the mean (OLS, TSLS, RANK, JACK) and the posterior mean for CIV. The analytic and bootstrapped confidence intervals differ dramatically for TSLS, indicative of a weak instrument. For OLS and JACK, the bootstrapped interval and analytic intervals are close, suggesting the normal approximation is more valid for these estimators. CIV agrees with the jackknife and rank point estimate, but offers an uncertainty interval more in-line with the bootstrap interval.

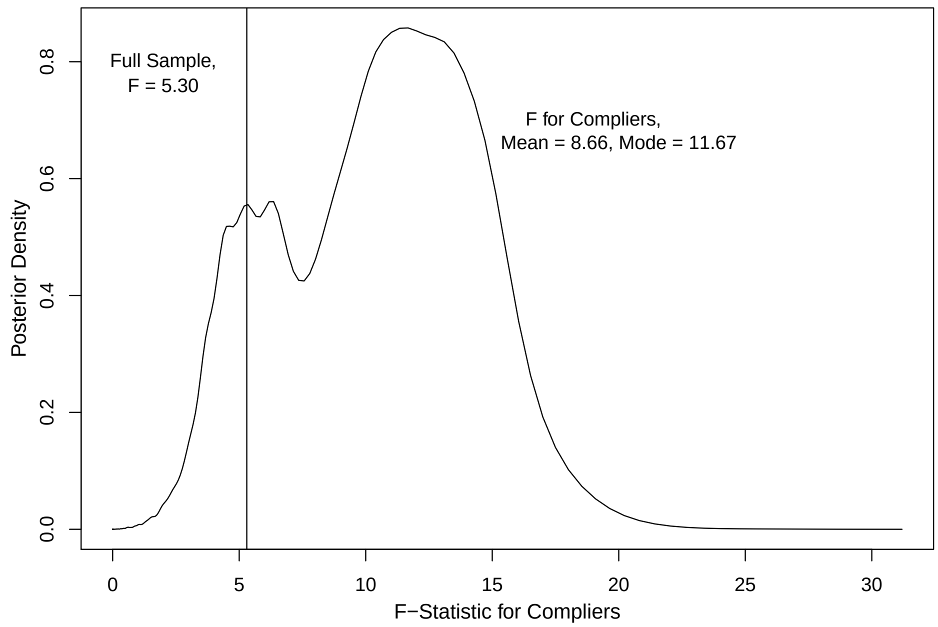

Figure 2: **Posterior Distribution of First-Stage F-Statistic for Compliers**. CIV weights observations by the probability that the instrument carries some information on the value of the endogenous variable. One means of assessing CIV estimation is to compare the posterior distribution of the first-stage F-statistic on the instrument for the compliers with that of the original sample. Figure 2 shows the posterior density of the F-statistic for compliers. The bulk of the posterior density of the CIV F-statistic falls well above that from the full sample (given by the vertical line).

CIV weights observations by the probability that the instrument carries some information on the value of the endogenous variable. One means of assessing CIV estimation is to compare the posterior distribution of the first-stage F-statistic on the instrument for the compliers with that of the original sample. Figure 2 shows the posterior density of the F-statistic for compliers. The posterior density of the F-statistic falls well above that from the full sample (given by the vertical line). We note that the sampling distribution of the complier F-statistic can no longer be interpreted in the normal fashion. The sampling distribution of the normal F-statistic is derived under a null that the instrument has no effect on the endogenous variable. The compliers are selected such that the effect of the instrument is as large as possible. We prefer instead to interpret the complier F-statistic as a distance measure.

While Figure 2 shows that CIV can strengthen an instrument, CIV offers the possibility of

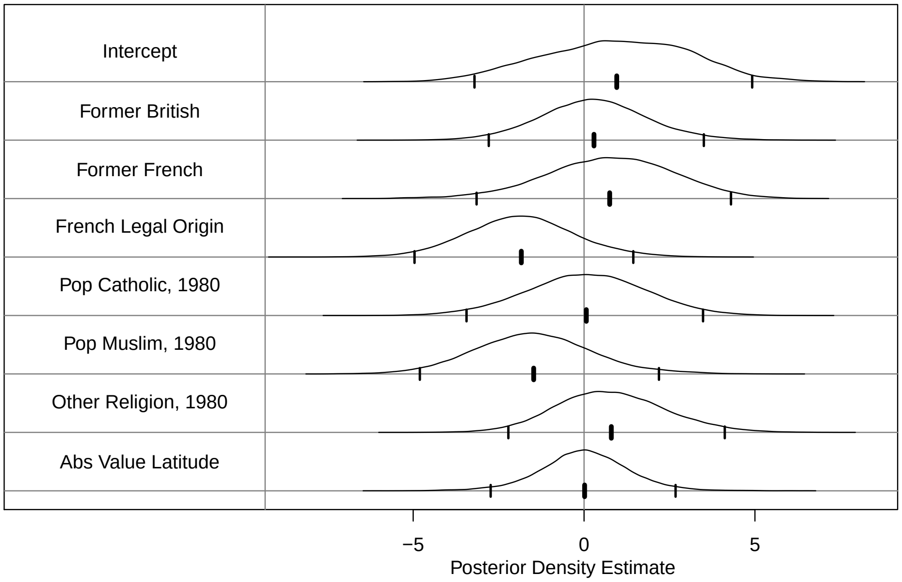

Figure 3: **Posterior Density of Coefficients for Complier Model.** Covariates have been standardized so that the densities are comparable across variables. The posterior mean and 95% credible intervals are marked for each density. None of the variables seem to have a significant effect on compliance–though the former British indicator and former French colony are highly correlated with the indicator for French legal system ($\rho = -0.91$ and 0.44, respectively). The only variable that appears important is that for French legal system, with 87.6% of the posterior mass less than zero. The data suggest that colonizer coming from a common law system like Britain's were more likely to comply with the instrument, perhaps due to the flexibility of these systems relative to their civil law counterparts (Glaeser and Shleifer, 2002).

characterizing who complies as a function of covariates. We offer some suggestive evidence that the effect of settler mortality varied based off of the type of legal system put in place by the colonizers. Figure 3 contains the posterior density of each parameter in the compliance model. Covariates have been standardized so that the densities are comparable across variables. The posterior mean and 95% credible intervals are marked for each density. None of the variables seem to have a significant effect on compliance–though the former British indicator and former French colony are highly correlated with the indicator for French legal system ($\rho = -0.91$ and 0.44, respectively). The only variable that appears important is that for French legal system, with 87.6% of the posterior mass less than zero. The data suggest that colonizer coming from a common law system like Britain's were more likely to comply with the instrument, perhaps due to the flexibility of these

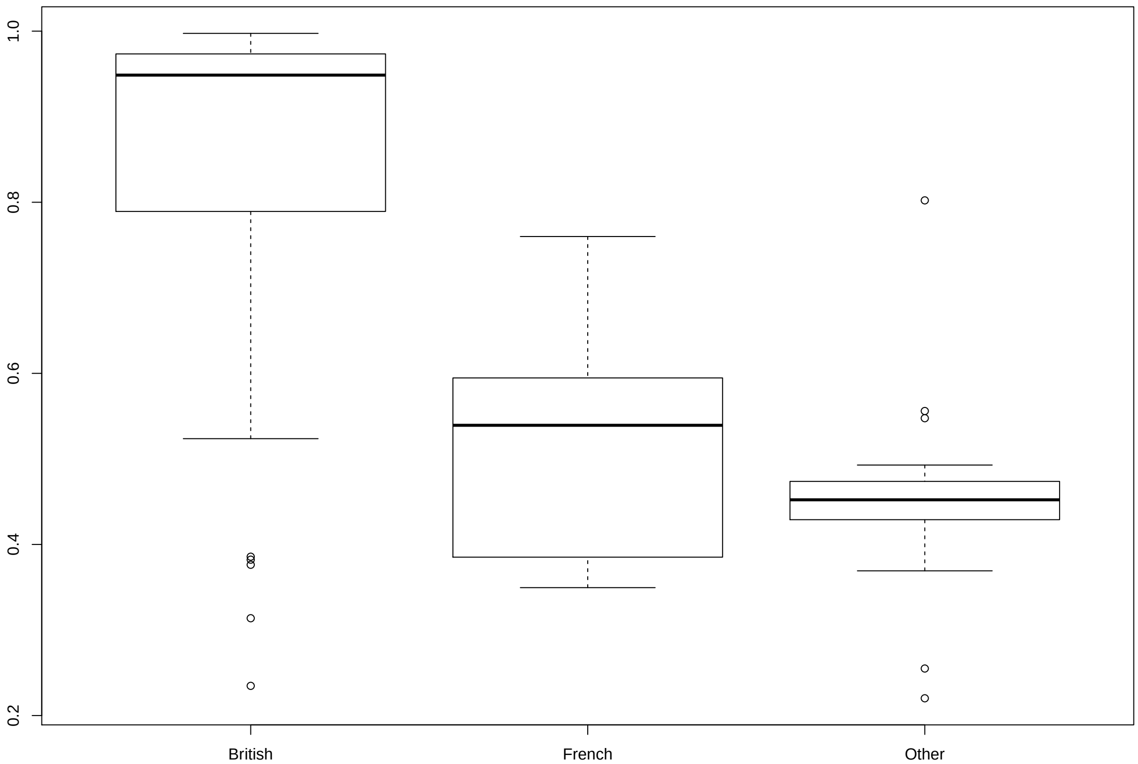

Figure 4: **Mean Posterior Compliance Probability, by Colonizer.** We find higher compliance rates among the British than French or other colonizers. We note that the posterior interval for most of the countries is large, with 47 out of 64 of the intervals containing 0.01 and 0.99.

systems relative to their civil law counterparts (Glaeser and Shleifer, 2002). Figure 4 shows the box and whisker plot of the posterior means of each country, by colonizer. Again, we find higher compliance rates among the British than French or other colonizers. We note that the posterior interval for most of the countries is large, with 47 out of 64 of the intervals containing 0.01 and 0.99. The estimates are necessarily imprecise, as we have 64 observations and are fitting 24 different mean parameters.

Even in a small dataset, CIV can increase the precision of IV estimates, while allowing the researcher to make some claims about who exactly complied with the treatment. We turn next to simulation evidence, showing that CIV performs favorably compared to several alternative methods.

4 Simulations

In this section, we compare the performance of the proposed methodology with that of the other estimators in various settings using Monte Carlo simulations. First, we present simulation results on the performance of point estimates' bias and root mean-squared error. The results show that CIV outperforms both TSLS and the Jackknife estimates, and performs comparably to the rank-based estimator. Next, we assess performance in terms of uncertainty estimates. We find the CIV and CIV-augmented TSLS estimates perform favorably relative to existing methods.

4.1 Simulation Setup

We consider a case in which there are one endogenous variable D_i and one instrumental variable Z_i ($K_D = K_Z = 1$) in all simulations. In addition to these variables, there are four exogenous covariates $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$. The instrument and each element of X_i are drawn independently from the standard normal distribution. We use the same covariates in the model for the compliance probability, i.e., $W_i = (1, X_i)$.

We conduct simulations with five levels of compliance probabilities. After the instrument and the covariates are generated, the true compliance probabilities are constructed by computing $\Phi(W_i^\top \alpha)$ where $\alpha = (\alpha_0, -2, -2, 2, -2)$. To see how the estimators perform under different levels of the weakness of the instrument, we set $\alpha_0 = -8, -4, 0, 4,$ and 8. Almost no compliers exist in the sample when $\alpha_0 = -8$, whereas almost all observations in the sample are compliers in the case of $\alpha_0 = 8$. The average compliance probabilities corresponding to the values of α_0 are $.026, .17, .5, .83,$ and .97, respectively.

For the outcome and the endogenous variable, we use the following data generating process:

$$Y_i = 1 - 2D_i + X_i^\top [2, -1, 1, 1]^\top + \epsilon_i$$

$$D_i = \begin{cases} 2 + 3Z_i + X_i^\top [2, -2, 2, 1]^\top + \eta_i; & C_i = 1 \\ -2 + X_i^\top [2, -2, 2, 1]^\top + \eta_i; & C_i = 0 \end{cases}$$

where

$$\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}\right) \tag{24}$$

Our parameter of interest is the coefficient on the endogenous variable, with a true value of -2.

The proposed method is highly model-based. Therefore, it is particularly important to assess its robustness against the violation of the modeling assumptions. We run four different sets of simulations. In the first, the mean and error are properly specified, so all assumptions of the model. In the next three, we misspecify the error, the mean covariates, and both, respectively. To misspecify the error, we replace the bivariate Gaussian errors in Equation (24) with the exponentiated values. This leads to violation of the distributional assumptions upon which our Bayesian model relies. To misspecify the mean, we fit the models to transformed variables $X_i^* = [X_{i1}^3, 1/X_{i2}, X_{i3}, X_{i4})]$ and $Z_i^* = \exp(Z_i)$, even though the data are generated from $X_{i\cdot}$ and Z_i. The final scenario incorporate both types of misspecification.

As in our empirical example, we compare the CIV estimator with the TSLS, the rank-based, and the jackknifed bias-corrected estimators. We use the posterior mean as the CIV point estimate. The performance of point estimates is considered in terms of the bias and the root mean-squared error (RMSE). The uncertainty intervals are assessed in terms of coverage, the proportion of the simulaitons where the 95% interval contains the true value, and power, which we operationalize as the proportion of the simulations where the entire 95% interval is less than zero. For the CIV, we checked the convergence of the posterior sample by looking at the Gelman-Rubin statistics (Gelman and Rubin, 1992) with three independent MCMC chains. All simulations were run 500 times.

4.2 Results

Tables 1 and 2 present the results of our Monte Carlo simulation study on point estimates. Overall, the CIV estimator performs comparably or better than the other methods, particularly with a weak instrument.

The upper block of Tables 1 shows the results under the scenario in which the model is correctly specified. The lower block reports the results in the case where the mean is properly specified but the error distribution is misspecified. When both the error distributional assumption and the specification of the variables are correct, TSLS and CIV have similar bias but CIV improves the RMSE dramatically when the instrument is weak. The jackknife estimator sometimes outperforms both of them in terms of bias if the instrument is strong. It performs poorly, though, with few compliers. The rank estimator outperforms the TSLS in terms of RMSE, but achieves either a comparable or higher RMSE than CIV.

The lower panel of Table 1 shows the results under the wrong distributional assumption of the error terms. The bias of the TSLS estimator does not deteriorate compared to the case with the correct assumptions. This is expected because the TSLS estimator uses only the moment conditions, which still hold in this scenario. It is somewhat surprising, however, that the bias of CIV is close to or smaller than that of TSLS even though CIV relies on the assumption of the Gaussian error distribution. Moreover, the CIV estimator almost always has smaller RMSE than the TSLS estimator across sample sizes and the level of compliance. In particular, the reduction of

Sample size	Compliance	Bias TSLS	Bias CIV	Bias Rank	Bias Jackknife	RMSE TSLS	RMSE CIV	RMSE Rank	RMSE Jackknife
(1) Both correct									
	None	0.084	-0.101	-0.783	13.156	9.575	0.302	5.006	1051.563
	Low	0.063	0.006	0.093	-0.565	2.733	0.148	1.578	56.002
$n = 100$	Half	0.023	0.009	0.018	0.010	0.165	0.097	0.230	0.152
	High	0.012	0.012	-0.001	0.008	0.091	0.078	0.086	0.090
	All	0.009	0.013	0.001	0.007	0.077	0.077	0.075	0.077
	None	-1.052	0.103	0.015	-39.555	8.055	0.250	5.775	1091.701
	Low	0.049	0.017	0.021	-0.037	0.373	0.100	0.884	0.293
$n = 250$	Half	0.004	0.010	0.005	0.000	0.100	0.062	0.102	0.099
	High	0.002	0.006	0.004	0.000	0.061	0.053	0.058	0.061
	All	0.001	0.007	-0.004	0.000	0.052	0.050	0.048	0.052
	None	0.070	0.075	0.255	206.399	6.407	0.131	4.989	4069.610
	Low	-0.040	-0.001	0.105	-0.050	0.150	0.048	0.899	0.150
$n = 1000$	Half	-0.001	0.005	-0.003	-0.002	0.043	0.025	0.048	0.043
	High	-0.021	-0.011	0.000	-0.022	0.042	0.031	0.026	0.042
	All	-0.019	-0.011	0.002	-0.020	0.036	0.030	0.025	0.037
(2) Error incorrect									
	None	11.963	-3.332	0.251	1639.838	258.767	6.774	16.104	27328.818
	Low	-5.632	-0.746	0.765	3166.359	83.902	4.092	9.815	49484.295
$n = 100$	Half	-0.325	-0.124	0.593	-0.319	6.978	3.548	5.697	6.160
	High	-0.171	-0.337	-0.032	-0.192	3.713	3.240	2.742	3.656
	All	-0.137	-0.318	-0.050	-0.154	3.122	2.982	2.666	3.095
	None	762.795	-0.035	-0.288	188077.067	11355.821	13.735	18.904	2762701.658
	Low	1.424	1.551	-0.484	-8.965	16.167	9.588	10.131	145.102
$n = 250$	Half	0.247	0.195	-0.045	0.217	3.684	2.422	4.854	3.593
	High	0.152	0.084	0.036	0.140	2.244	1.763	1.684	2.225
	All	0.128	0.052	0.015	0.120	1.869	1.772	2.132	1.863
	None	180.428	3.306	-0.164	209354.728	367.503	17.639	22.299	419017.961
	Low	2.176	2.525	-0.136	2.020	7.093	5.425	7.421	6.801
$n = 1000$	Half	-0.632	-0.460	0.145	-0.637	2.092	1.485	2.903	2.090
	High	-0.382	-0.358	0.054	-0.384	1.258	1.082	1.528	1.260
	All	-0.318	-0.315	0.062	-0.320	1.102	1.042	1.101	1.104

Table 1: **Relative Performance of the Four Different Estimators of Instrumental Variable Model (1).** The bias and root mean squared error (RMSE) are computed for the two-stage least squares (TSLS), the complier instrumental variable (CIV, posterior mean), the rank-based (Rank), and the jackknifed bias-corrected (Jackknife) estimators. The performance of the CIV estimator is compared with that of the other estimators under the no compliance, the low compliance, the half compliance, the high compliance, and the complete compliance conditions. The upper block of the table shows the results under the scenario in which the model is correctly specified. The lower block reports the results in the case where the error distribution is misspecified. When both the error distributional assumption and the specification of the variables are correct, TSLS and CIV have similar bias but CIV improves the RMSE dramatically when the instrument is weak. The jackknife estimator sometimes outperforms both of them in terms of bias if the instrument is strong. It performs poorly, though, with few compliers. The rank estimator outperforms the TSLS in terms of RMSE, but achieves either a comparable or higher RMSE than CIV.

Sample size	Compliance	Bias				RMSE			
		TSLS	CIV	Rank	Jackknife	TSLS	CIV	Rank	Jackknife
(3) Mean incorrect									
$n = 100$	None	3.970	0.075	0.340	322.250	54.046	0.170	4.935	4890.797
	Low	-0.299	0.001	0.574	-5.359	5.785	0.147	5.254	509.585
	Half	0.015	-0.014	-0.177	-2.407	0.856	0.130	2.302	31.482
	High	-0.008	-0.007	-0.005	0.003	0.163	0.106	0.154	0.167
	All	-0.008	0.000	-0.002	-0.004	0.124	0.101	0.130	0.127
$n = 250$	None	0.090	0.150	0.148	6.609	3.697	0.229	6.662	372.581
	Low	0.001	0.027	-0.171	3.672	1.513	0.138	2.634	39.973
	Half	-0.023	0.014	-0.039	0.002	0.224	0.095	0.416	0.619
	High	-0.004	0.008	-0.005	-0.003	0.092	0.079	0.097	0.095
	All	-0.003	0.008	-0.006	-0.003	0.076	0.073	0.078	0.079
$n = 1000$	None	3.504	0.178	0.153	2005.450	49.809	0.253	5.657	29629.307
	Low	0.009	0.020	-0.005	-9.028	0.854	0.079	0.899	152.069
	Half	-0.012	0.020	-0.008	-0.010	0.071	0.048	0.078	0.068
	High	0.002	0.009	-0.001	0.003	0.035	0.033	0.045	0.036
	All	0.002	0.007	0.001	0.003	0.028	0.029	0.038	0.029
(4) Both incorrect									
$n = 100$	None	-1.616	-1.364	-1.034	-999.565	99.246	5.455	15.321	9544.630
	Low	6.323	-0.755	0.609	1248.615	151.087	4.785	13.603	16786.753
	Half	-0.048	-0.529	0.033	-60.146	16.659	2.366	8.286	955.114
	High	-0.591	-0.563	0.149	-0.627	3.077	3.030	3.957	3.146
	All	-0.372	-0.168	-0.079	-0.374	2.679	6.166	3.184	2.763
$n = 250$	None	-12.719	0.154	0.075	-329745.974	139.128	10.774	18.813	4360869.552
	Low	8.889	3.327	0.025	-253.507	119.839	43.006	12.045	48684.919
	Half	-0.446	-0.221	-0.325	-0.215	4.200	10.889	3.923	4.613
	High	-0.200	-0.131	0.171	-0.214	1.926	1.537	2.955	1.838
	All	-0.105	0.589	0.124	-0.129	1.671	11.031	2.377	1.644
$n = 1000$	None	-5.047	2.671	-0.155	-524.098	75.606	27.469	20.492	16873.110
	Low	-2.006	0.804	-0.310	-15.688	7.975	11.869	9.548	190.306
	Half	-0.458	-0.396	-0.146	-0.462	2.058	1.651	3.387	1.988
	High	-0.248	-0.251	0.231	-0.250	1.268	1.221	3.713	1.251
	All	-0.054	-0.069	-0.087	-0.057	1.336	1.410	1.428	1.324

Table 2: **Relative Performance of the Four Different Estimators of Instrumental Variable Model (2).** The bias and root mean squared error (RMSE) are computed for the two-stage least squares (TSLS), the complier instrumental variable (CIV), the rank-based (Rank), and the jackknifed bias-corrected (Jackknife) estimators. With a misspecified mean (top), CIV is comparable or better than TSLS, in terms of bias and RMSE. CIV dominates TSLS and the rank based estimator when less than half of observations in the sample are compliers. The jackknife estimates again perform the worst with low compliance levels. With full compliance and at the largest sample size, all of the estimators perform equally well. When both the error distribution and the mean structures are misspecified (bottom), again TSLS and the jackknife estimates are unstable with low levels of compliers. The CIV and rank-based estimator are comparable in this scenario. At the smallest sample size, CIV outperforms the rank-based estimator, while at the moderate sample size the rank-based estimator performs better. At the largest sample size, the two estimators perform nearly the same, in terms of bias and RMSE.

RMSE is large under the weak instrument conditions (see “None” and “Low” rows). The jackknife estimator tends to have smaller bias with the strong instrument, but the RMSE is always larger than CIV. The rank-based estimator is always worse than CIV in both of the bias and variance.

Table 2 shows the results under the wrong mean specification (top) and both the wrong mean and error specification (bottom). Where the mean structures are misspecified, the CIV bias is comparable or lesser than that of TSLS, and the CIV estimator is much more stable. CIV dominates TSLS and the rank based estimator when less than half of observations in the sample are compliers. The jackknife estimates again perform the worst with low compliance levels. With full compliance and at the largest sample size, all of the estimators perform equally well.

When both the error distribution and the mean structures are misspecified, again TSLS and the jackknife estimates are unstable with low levels of compliers. The CIV and rank-based estimator are comparable in this scenario. At the smallest sample size, CIV outperforms the rank-based estimator, while at the moderate sample size the rank-based estimator performs better. At the largest sample size, the two estimators perform nearly the same, in terms of bias and RMSE.

We note that the misspecification is severe, exponentiating the residual and non-linear transformations of the covariates. With extreme concerns over misspecification, both CIV and the rank-based estimator perform well. Under reasonable model specifications or normal residuals, or both, the CIV performs the best on the whole.

Turning from point to interval estimation, we compare the uncertainty intervals across methods. In applications, interval estimation addresses two important questions. First, what range of values is likely to contain the true value? Second, can we differentiate our effect from zero? In regards to the first question, we assess the coverage of each method's uncertainty estimates. Specifically, we report the proportion of time that the 95% uncertainty interval (credible or confidence) contains the true value of -2. In regards to the second question, we assess the power of each method by reporting the proportion of the time each method's 95% confidence interval falls entirely below zero.

Sample Size	Compliance	Coverage TSLS	Coverage CIV	Coverage CIV-TSLS	Coverage Rank	Coverage Jack Knife	Power TSLS	Power CIV	Power CIV-TSLS	Power Rank	Power Jack Knife
(1) Both Correct											
	None	0.882	0.936	0.742	0.430	0.864	0.407	1.000	0.550	0.767	0.393
	Low	0.964	0.834	0.849	0.603	0.988	0.750	1.000	0.866	0.863	0.774
$n = 100$	Half	0.964	0.882	0.939	0.547	0.968	1.000	1.000	1.000	0.993	1.000
	High	0.952	0.844	0.949	0.520	0.956	1.000	1.000	1.000	1.000	1.000
	All	0.950	0.760	0.950	0.520	0.942	1.000	0.994	1.000	1.000	1.000
	None	0.896	0.686	0.779	0.437	0.916	0.452	1.000	0.646	0.770	0.510
	Low	0.948	0.748	0.921	0.623	0.968	0.954	1.000	0.996	0.947	0.966
$n = 250$	Half	0.952	0.852	0.932	0.523	0.946	1.000	1.000	1.000	0.997	1.000
	High	0.948	0.826	0.933	0.533	0.936	1.000	1.000	1.000	1.000	1.000
	All	0.948	0.740	0.947	0.480	0.932	1.000	1.000	1.000	1.000	1.000
	None	0.896	0.480	0.906	0.507	0.894	0.588	1.000	0.908	0.843	0.678
	Low	0.936	0.758	0.940	0.557	0.940	1.000	1.000	1.000	0.973	1.000
$n = 1000$	Half	0.940	0.856	0.936	0.443	0.938	1.000	1.000	1.000	1.000	1.000
	High	0.942	0.818	0.947	0.477	0.936	1.000	1.000	1.000	1.000	1.000
	All	0.942	0.698	0.949	0.523	0.936	1.000	1.000	1.000	1.000	1.000
(2) Error incorrect											
0.653	None	1.000	0.978	0.944	0.617	0.984	0.008	0.244	0.191	0.397	0.012
	Low	0.988	0.962	0.921	0.550	0.994	0.038	0.359	0.418	0.517	0.036
$n = 100$	Half	0.956	0.946	0.939	0.527	1.000	0.246	0.472	0.489	0.723	0.276
	High	0.944	0.940	0.944	0.497	0.996	0.434	0.554	0.511	0.843	0.484
	All	0.944	0.944	0.939	0.590	0.994	0.474	0.556	0.508	0.880	0.540
	None	0.998	0.954	0.922	0.540	0.990	0.008	0.262	0.279	0.457	0.008
	Low	0.976	0.914	0.917	0.523	0.998	0.084	0.376	0.479	0.583	0.052
$n = 250$	Half	0.932	0.928	0.930	0.520	0.978	0.304	0.558	0.516	0.793	0.354
	High	0.928	0.916	0.922	0.460	0.972	0.510	0.624	0.571	0.920	0.586
	All	0.928	0.910	0.917	0.540	0.972	0.580	0.662	0.602	0.937	0.658
	None	0.998	0.837	0.900	0.543	0.982	0.008	0.337	0.435	0.490	0.008
	Low	0.972	0.872	0.907	0.490	0.986	0.082	0.558	0.552	0.660	0.104
$n = 1000$	Half	0.956	0.902	0.936	0.470	0.968	0.396	0.712	0.672	0.873	0.506
	High	0.956	0.910	0.946	0.527	0.968	0.672	0.778	0.734	0.957	0.756
	All	0.956	0.926	0.955	0.677	0.966	0.738	0.798	0.762	0.960	0.816

Table 3: **Coverage and Power of the TSLS, Jackknife, and CIV estimators (1).** This table shows the probability of the 95% confidence/credible intervals containing the true value (Coverage) and the probability that the upper bounds the 95% confidence/credible intervals are less than zero. The quantities are computed for the two-stage least squares (TSLS), the jackknife bias-corrected, the complier instrumental variable (CIV), and the CIV-augmented TSLS estimators. The performance of the CIV and CIV-augmented TSLS estimator is compared with that of the other estimators under the no compliance, the low compliance, the half compliance, the high compliance, and the complete compliance scenarios. When the model is correctly specified (top), CIV-TSLS, TSLS, and the Jackknife achieve nominal coverage. Of these three, CIV-TSLS is the most powerful. We see the same pattern when the error is misspecified. CIV coverage is sometimes too low, particularly when the sample size is large and the model properly specified. .

Tables 3 and 4 show the simulation results for coverage and power for TSLS, jackknife, rank, and CIV estimators. We also include the augmented CIV-TSLS estimate, described above, which is the posterior mean of the 95% TSLS confidence interval for the compliers[1]. CIV-TSLS and CIV have the same point estimate, so we did not include both when evaluating point estimates above.

Considering Table 3, TSLS has near-nominal coverage across scenarios: coverage rates range from 88.2% to 100%, but are mostly near 95%. The Jackknife also has near-nominal coverage, ranging from 86.4% to 100%, but again mostly near 95%. The CIV posterior credible intervals do not have nominal coverage, with rates ranging from 48.0% to 97.8%. In most cases, particularly with a properly specified model, CIV-TSLS comes closer than CIV in achieving nominal coverage, and is competitive with TSLS in this regard.

Considering Table 4, the performance of the CIV and CIV-augmented TSLS estimator is compared with that of the other estimators under different levels of compliance and sample size. When the mean is misspecified (top), CIV-augmented TSLS, TSLS, and the Jackknife achieve near-nominal coverage. We see the same pattern when both the mean and the error is misspecified (bottom). Of estimators with near-nominal coverage, the CIV-augmented TSLS is generally the most powerful. With a large proportion of compliers, though, the Jackknife is more powerful than CIV-augmented TSLS. The rank estimator does not have near-nominal coverage, with coverage never getting above 81%.

Summarizing the simulation results, CIV and the rank-based estimators performed best in terms of point estimation. Both outperformed TSLS and the jackknife estimator, while CIV outperformed the rank-based except in a few scenarios where both the mean and error were misspecified. In terms of coverage, TSLS, the Jackknife estimator, and CIV-augmented TSLS achieved near-nominal

[1]We should note, however, the CIV-augmented TSLS confidence intervals cannot be obtained when too few observations are classified to be compliers. If this happens in most iterations of the Gibbs sampler, the researcher probably should give up relying on the instruments.

Sample Size	Compliance	Coverage TSLS	Coverage CIV	Coverage CIV-TSLS	Coverage Rank	Coverage Jack Knife	Power TSLS	Power CIV	Power CIV-TSLS	Power Rank	Power Jack Knife
(3) Mean incorrect											
	None	0.988	0.951	0.947	0.677	0.954	0.206	0.963	0.525	0.700	0.179
	Low	0.980	0.906	0.924	0.557	0.973	0.499	0.968	0.829	0.740	0.382
$n = 100$	Half	0.977	0.882	0.938	0.573	0.977	0.926	0.979	0.992	0.947	0.890
	High	0.963	0.890	0.935	0.563	0.941	0.998	0.983	1.000	1.000	0.993
	All	0.951	0.884	0.932	0.817	0.930	1.000	0.998	1.000	1.000	0.998
	None	0.996	0.883	0.926	0.657	0.988	0.338	0.976	0.671	0.777	0.275
	Low	0.981	0.810	0.909	0.643	0.986	0.768	0.955	0.952	0.897	0.688
$n = 250$	Half	0.961	0.881	0.915	0.503	0.972	0.992	0.985	0.994	0.997	0.992
	High	0.943	0.858	0.922	0.497	0.943	1.000	0.995	0.997	1.000	1.000
	All	0.936	0.859	0.917	0.463	0.941	1.000	0.991	0.996	1.000	1.000
	None	0.991	0.513	0.873	0.603	0.974	0.420	0.946	0.893	0.860	0.355
	Low	0.984	0.794	0.881	0.533	0.988	0.967	0.953	0.995	0.960	0.957
$n = 1000$	Half	0.943	0.845	0.908	0.453	0.943	1.000	0.973	0.995	1.000	1.000
	High	0.941	0.829	0.923	0.493	0.937	1.000	0.978	0.992	1.000	1.000
	All	0.939	0.834	0.925	0.507	0.928	1.000	0.986	0.998	1.000	1.000
(4) Both incorrect											
	None	1.000	0.977	0.949	0.597	0.969	0.000	0.335	0.184	0.513	0.023
	Low	0.996	0.965	0.956	0.570	0.983	0.035	0.388	0.306	0.517	0.054
$n = 100$	Half	0.992	0.952	0.964	0.490	1.000	0.158	0.501	0.413	0.727	0.267
	High	0.973	0.959	0.964	0.483	0.986	0.369	0.596	0.462	0.810	0.542
	All	0.969	0.959	0.962	0.467	0.982	0.423	0.579	0.484	0.843	0.599
	None	1.000	0.937	0.951	0.550	0.980	0.004	0.303	0.231	0.533	0.028
	Low	0.990	0.924	0.947	0.480	0.990	0.033	0.441	0.385	0.650	0.094
$n = 250$	Half	0.966	0.927	0.940	0.413	0.986	0.207	0.562	0.446	0.810	0.420
	High	0.951	0.937	0.945	0.427	0.960	0.407	0.630	0.510	0.853	0.613
	All	0.949	0.943	0.950	0.413	0.953	0.492	0.648	0.543	0.897	0.664
	None	0.998	0.841	0.946	0.490	0.978	0.000	0.358	0.351	0.573	0.022
	Low	0.977	0.856	0.938	0.443	0.994	0.025	0.534	0.479	0.740	0.150
$n = 1000$	Half	0.953	0.897	0.924	0.400	0.929	0.281	0.652	0.584	0.890	0.567
	High	0.944	0.891	0.932	0.413	0.924	0.561	0.746	0.652	0.930	0.742
	All	0.946	0.899	0.938	0.393	0.926	0.642	0.771	0.676	0.963	0.785

Table 4: **Coverage and Power of the TSLS, Jackknife, and CIV estimators (2).** This table shows the probability of the 95% confidence/credible intervals containing the true value (Coverage) and the probability that the upper bounds the 95% confidence/credible intervals are less than zero. The quantities are computed for the two-stage least squares (TSLS), the jackknife bias-corrected, the complier instrumental variable (CIV), and the CIV-augmented TSLS estimators. The performance of the CIV and CIV-augmented TSLS estimator is compared with that of the other estimators under the no compliance, the low compliance, the half compliance, the high compliance, and the complete compliance scenarios. When the mean is misspecified (top), CIV-TSLS, TSLS, and the Jackknife achieve near-nominal coverage. We see the same pattern when both the mean and the error is misspecified (bottom). Of estimators with near-nominal coverage, the CIV-TSLS is generally the most powerful. With a large proportion of compliers, though, the Jackknife is more powerful than CIV-TSLS. The rank estimator does not have near-nominal coverage, with coverage never getting above 81%.

coverage. Of these three, CIV-augmented TSLS was generally the most powerful. In the scenarios with a high proportion of compliers, Jackknife was more powerful.

5 Conclusion

We introduced a new method, Complier Instrumental Variable (CIV) estimation, in order to address two concerns. First, IV analyses only estimate local treatment effects, for observations that comply with the instrument. A common critique of instrumental variable analyses is that we never know for whom, exactly, we have estimated a treatment effect. CIV is an attempt to help the researcher characterize the sub-population on which they are estimating a causal effect. Second, knowledge of these compliers can help strengthen a weak instrument. As non-compliers simply add noise to an estimate, and weaken the instrument, down-weighting estimated non-compliers can help the researcher generate a more valid and reliable causal effect estimate. We illustrated CIV on a prominent dataset, showing that it produced more precise estimates than common alternative methods. Second, we showed that modeling compliance can cast new insight into the substance of a problem. A set of simulations reinforced these findings.

As we move forward, we plan to extend the CIV project in several directions. First, we will extend the methodology to mediation analysis, estimating for which observations exert a direct effect and which exert an indirect effect on an outcome. Second, we plan to consider the problem of extrapolating from the local treatment effect to the average treatment effect, within a CIV/SEM framework.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, **113**(2), 231–263.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development. *American Economic Review*, **91**(5), 1369–1401.

Acemoglu, D., Johnson, S., and Robinson, J. (2012). The colonial origins of comparative development: An empirical investigation: Reply. *American Economic Review*, **102**(6), 3077–3110.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.

Albouy, D. (2012). The colonial origins of comparative development: An empirical investigation: Comment. *American Economic Review*, **102**(6), 3059–76.

Angrist, J., Imbens, G., and Krueger, A. (1999). Jackknife instrumental variables estimation. *Jornal of Applied Econometrics*, **14**(1), 57–67.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, **91**(434), 444–455.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, **51**(2), 277–297.

Barnard, J., Frangakis, C. E., Hill, J. L., and Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York (with discussion). *Journal of the American Statistical Association*, **98**(462), 299–311.

Betz, T. (2013). Robust estimation with nonrandom measurement error and weak instruments. *Political Analysis*, **21**(1), 86–96.

Borjas, G. J. (1988). Self-selection and the earnings of immigrants. *American Economic Review*, **77**(4), 531–53.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, **90**(430), 443–450.

Chamberlain, G. and Imbens, G. (2004). Random effects estimators with many instrumental variables. *Econometrica*, **72**(1), 295–306.

Chib, S. (2003). *On inferring effects of binary treatments with unobserved confounders.*, volume 7, pages 65–84. Oxford University Press.

Conley, T., Hansen, C., and Rossi, P. (2012). Plausibly exogenous. *Review of Economics and Statistics*, **94**(1), 260–272.

Crespo-Tenorio, A. and Montgomery, J. M. (2013). A bayesian approach to inference with instrumental variables: Improving estimation of treatment effects with weak instruments and small samples.

Davidson, R. and MacKinnon, J. (2006). The case againts jive. *Journal of Applied Econometrics*, **21**(6), 827–833.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, **48**(2), 424–455.

Dunson, D. and Herring, A. (2005). Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*, **6**(1), 11–25.

Esterling, K. M., Neblo, M. A., and Lazer, D. M. J. (2011). Estimating treatment effects in the presence of selection on unobservables: The generalized endogenous treatment model. *Political Analysis*, **19**, 205–226.

Flores-Lagunes, A. (2007). Finite sample evidence of iv estimators under weak instruments. *Journal of Applied Econometrics*, **22**(3), 677–694.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, **58**(1), 21–29.

Frankel, J. and Romer, D. (1999). Does trade cause growth? *American Economic Review*, **89**(3), 379–399.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences (with discussion). *Statistical Science*, **7**(4), 457–472.

Glaeser, E. and Shleifer, A. (2002). Legal origins. *Quarterly Journal of Economics*, **117**(4), 1193–1229.

Hahn, J. and Hausman, J. (2003). Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review*, **93**(2), 118–125.

Hahn, J., Hausman, J., and Kuersteiner, G. (2004). Estimation with weak instruments: accuracy of higher order bias and mse approximations. *Econometrics Journal*, **7**, 272–306.

Hahn, J., Ham, J. C., and Moon, H. R. (2011). The hausman test and weak instruments. *Journal of Econometrics*, **160**(2), 289–299.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Texts*. New York: Academic Press.

Heckman, J. and Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the Americal Statistical Association*, **84**(408), 862–874.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, **46**, 931–60.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, **47**(1), 153–161.

Heckman, J. J. and Urzua, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, **156**(1), 27–37.

Hirano, K., Imbens, G. W., Rubin, D. B., and Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, **1**(1), 69–88.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, **48**(2), 399–423.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62**(2), 467–475.

Imbens, G. W. and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society, Series A*, **168**(1), 109–126.

Imbens, G. W. and Rubin, D. B. (1997a). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, **25**(1), 305–327.

Imbens, G. W. and Rubin, D. B. (1997b). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, **64**(4), 555–574.

Karl, A. and Lenkoski, A. (2012). Instrumental variable bayesian model averaging via conditional bayes factors.

Kuersteiner, G. and Okui, R. (2010). Constructing optimal instruments by first-stage prediction averaging. *Econometrica*, **78**(2), 697–718.

Meng, X.-L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, **86**, 301–320.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, **112**(4), 725–753.

Muthen, B. (1984). A general structural equation model wth dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, **49**, 115–132.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rossi, P., Allenby, G., and McCulloch, R. (2006). *Bayesian Statistics and Marketing.* New York: Wiley.

Roy, A. D. (1951). Thoughts on the distribution of earnings. *Oxford Economic Papers, New Series*, **3**(2), 135–146.

Sovey, A. and Green, D. (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, **55**(1), 188–200.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica: Journal of the Econometric Society*, pages 557–586.

A Appendix

A.1 Details of the ECM algorithm

CM-step 2 In this step, we will obtain the estimates of the second stage coefficients, $\{\beta^{(l)}, \gamma^{(l)}\}$. Factoring the joint density of (Y, D) into the marginal density of D and the conditional density of Y given D, maximization of the Q-function with respect to $\{\beta^{(l)}, \gamma^{(l)}\}$ reduces to

$$\max_{\beta,\gamma} \sum_{i=1}^{N} \Big\{ \hat{\pi}_i^{(l)} \mathrm{P}(Y_i|\Theta_C, \Omega, D_i, Z_i, X_i, C_i = 1) \\ + (1 - \hat{\pi}_i^{(l)}) \mathrm{P}(Y_i|\Theta_{NC}, \Omega, D_i, Z_i, X_i, C_i = 0) \Big\}. \tag{25}$$

Noting that

$$\begin{aligned} &\mathrm{P}(Y_i|\Theta_C, \Omega, D_i, Z_i, X_i, C_i = 1) \\ &= \phi\Big(Y_i - \frac{\sigma_\epsilon}{\sigma_\eta}\rho(D_i - \delta_0^C - Z_i^\top\delta - X_i^\top\theta)\Big| D_i^\top\beta + X_i^\top\gamma, (1-\rho^2)\sigma_\epsilon^2\Big) \qquad (26) \\ &\mathrm{P}(Y_i|\Theta_C, \Omega, D_i, Z_i, X_i, C_i = 0) \\ &= \phi\Big(Y_i - \frac{\sigma_\epsilon}{\sigma_\eta}\rho(D_i - \delta_0^{NC} - X_i^\top\theta)\Big| D_i^\top\beta + X_i^\top\gamma, (1-\rho^2)\sigma_\epsilon^2\Big), \qquad (27) \end{aligned}$$

the solution to (25) is given by the weighted regression of $\Bigg(\Big(Y - \frac{\sigma_\epsilon^{(l-1)}}{\sigma_\eta^{(l-1)}}\rho^{(l-1)}(D - \delta_0^{C(l-1)} - Z\delta^{(l-1)} - X\theta^{(l-1)}\Big)^\top, \Big(Y - \frac{\sigma_\epsilon^{(l-1)}}{\sigma_\eta^{(l-1)}}\rho^{(l-1)}(D - \delta_0^{NC(l-1)} - X\theta^{(l-1)}\Big)^\top\Bigg)^\top$ on $(\mathbf{1}, D, X)$ with weights $(\hat{\pi}^{(l)\top}, (1 - \hat{\pi}^{(l)})^\top)^\top$.

CM-step 4 The last CM-step obtains the estimate of $\{\Omega\}$. Since all the coefficient parameters are fixed in this step, the solution is given by

$$\Omega^{(l)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\hat{\pi}_i^{(l)}\hat{\mathbf{e}}_i^{(l)}\hat{\mathbf{e}}_i^{(l)\top} + (1 - \hat{\pi}_i^{(l)})\tilde{\mathbf{e}}_i^{(l)}\tilde{\mathbf{e}}_i^{(l)\top}\Big\}$$

where

$$\hat{\mathbf{e}}_i^{(l)} = \begin{pmatrix} Y_i - D_i^\top\beta^{(l)} - X_i^\top\gamma^{(l)} \\ D_i - \delta_0^{C(l)} - Z_i^\top\delta^{(l)} - X_i^\top\theta^{(l)} \end{pmatrix},$$

$$\tilde{\mathbf{e}}_i^{(l)} = \begin{pmatrix} Y_i - D_i^\top\beta^{(l)} - X_i^\top\gamma^{(l)} \\ D_i - \delta_0^{NC(l)} - X_i^\top\theta^{(l)} \end{pmatrix}.$$

Abstract Word Count:121
Body Word Count: 11,091