The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–13

ivmediate: Causal mediation analysis in instrumental variables regressions

Christian Dippel
UCLA Anderson School of Management
Los Angeles, USA
christian.dippel@anderson.ucla.edu

Andreas Ferrara
University of Pittsburgh
Pittsburgh, USA
a.ferrara@pitt.edu

Stephan Heblich
University of Toronto
Toronto, CA
stephan.heblich@utoronto.ca

Abstract. In this article, we describe the use of `ivmediate`, a new command to estimate causal mediation effects in instrumental variables (IV) settings using the framework developed by Pinto et al. (2019). `ivmediate` allows estimation of a treatment effect and the share of this effect that can be attributed to a mediator variable. While both treatment and mediator can be potentially endogenous, a single instrument suffices to identify both the causal treatment and mediation effects.

Keywords: st0001, ivmediate, causal mediation analysis, treatment effects, instrumental variables

1 Introduction

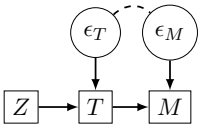
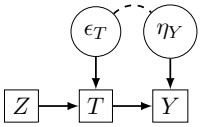
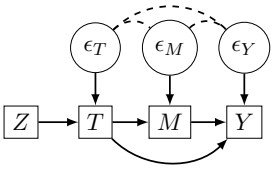
There are many settings where a researcher would like to understand the mechanism that underlies an estimated effect of a treatment T on an outcome Y. For example, Becker and Woessmann (2009) are interested in the Weber Hypothesis that religion, specifically Protestantism, affects economic growth. Since Protestantism promoted reading of the bible,[1] they establish that an underlying mechanism M of the effect of religion on economic growth works through human capital accumulation, especially literacy. Given that the prevalence of religion across regions is likely not random, they introduce an instrumental variable (IV) and show that Protestantism caused higher literacy rates and thus economic growth. They derive plausible bounds for the range of a mediation effect but lack a formal framework to causally estimate the indirect effect of religion on economic growth that works through literacy.

Such an exercise of unpacking mechanisms is called mediation analysis, where a treatment T and one of its outcomes M, i.e. the *mediator*, jointly cause a final outcome of interest Y. Mediation analysis has long been used in settings where T can be assumed to be randomly assigned. However, when T is systematically non-random and there-

1. As opposed to Catholicism, where at that time religious content was mainly consumed through sermons at church.

Table 1: The Identification Problem of Mediation Analysis with IV

A. Graphical Representation

Model I: IV for M	*Model II: IV for Y*	*Model III: IV for the Mediation Model*
		

B. Model Equations

$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T)$	$T = f_T(Z, \epsilon_T),\ M = f_M(T, \epsilon_M)$
$M = f_M(T, \epsilon_M)$	$Y = g_Y(T, \eta_Y)$	$Y = f_Y(T, M, \epsilon_Y)$
$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M)$	$Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$	$Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$

Notes: (*a*) *Model I* is the standard IV model, which enables the identification of the causal effect of T on M. *Model II* is the standard IV model that enables the identification of the causal effects of T on Y. *Model III* is the IV Mediation Model with an instrumental variable Z. (*b*) Panel A gives the graphical representation of the models. Panel B presents the non-parametric structural equations of each model. We use $\perp\!\!\!\perp$ to denote statistical independence.

fore needs to be *instrumented* by a variable Z,[2] there has been a lack of frameworks for undertaking mediation analysis in such IV settings without having separate instruments for both T and M.[3] The command `ivmediate` fills this gap and provides a new regression command that allows researchers to use a single IV to estimate the causal effect of the intermediate variable on a final outcome using the estimator developed by Pinto et al. (2019). This complements existing ways to estimate causal mediation effects which assume randomness in the assignment of treatment T (Imai et al. 2010) or require separate instruments for T and M (e.g. Frölich and Huber 2017; Jun et al. 2016).

Table 1 illustrates the identification challenge described above. As a starting point, we show the standard IV estimations of the causal effect of T on M (*Model I*) and the causal effect of T on Y (*Model II*). In *Model I*, T is considered endogenous (i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$) and we introduce an instrumental variable Z for the endogenous treatment T

2. The requirements for a valid instrument are that it significantly affects the treatment conditional on covariates (relevance condition), and that it affects Y only through T but not directly (exclusion restriction).
3. The traditional approach to mediation analysis makes the strong assumption that both T and M are exogenous, applies OLS to estimate three equations,

$$Y = \delta_Y^T \cdot T + \eta_Y, \quad M = \beta_M^T \cdot T + \epsilon_M, \qquad \text{and} \quad Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y,$$

and compares the total effect δ_Y^T to the indirect effect $\beta_Y^M \cdot \beta_M^T$. See Baron and Kenny (1986) and MacKinnon (2008) for an overview.

which is both uncorrelated with the omitted variables ($Z \perp\!\!\!\perp \epsilon_T, \epsilon_M$) and a reasonably strong predictor of T. *Model II* estimates the 'total effect' (TE) of T on Y using the same IV approach: $\epsilon_T \not\perp\!\!\!\perp \eta_Y$, but Z is exogenous (i.e. $Z \not\perp\!\!\!\perp \epsilon_T, \eta_Y$).

To identify what fraction of the total effect is explained by the indirect effect, we have to perform a mediation analysis which decomposes the total effect of T on Y into the mediated "indirect" effect of T on Y that operates through M and the residual "direct effect" that does not work though M. Model III of Table 1 shows the main identification challenge in combining the two IV models into a general mediation model. Equations $M = f_M(T, \epsilon_M)$ and $Y = f_Y(T, M, \epsilon_Y)$ imply that T causes Y indirectly through M as well as directly which is graphically represented by the arrow directly linking T to Y. In a regression of Y on both T and M, there are two potentially endogenous regressors (i.e. $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$, $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$), but there is only one instrument Z to address this endogeneity.

To overcome the under-identification problem, we do not assume away endogeneity in any of the key relationships in *Model III* ($\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ are all maintained) and yet we do not need additional instruments. Instead, the omitted variable concerns themselves can suggest a natural solution. This is the case when T is endogenous in a regression of M on T because of confounders that jointly affect M and T, and T is endogenous in a regression of Y on T because of the same confounders that affect Y primarily through M.

Pinto et al. (2019) show that this assumption alone is sufficient to unpack the causal channels in *Model III*, and therefore allows us to identify the extent to which T causes Y through M. Under linearity, the resulting identification framework is straightforward to estimate using three separate 2SLS regressions: these estimate i) the effect of T on M, ii) the effect of T on Y, and iii) the effect of M on Y conditional on T.

In the following section, we will briefly explain the underlying econometric theory before we explain the estimation procedure in Section 3. There we also provide further guidance with respect to the interpretation of results and issues regarding weak identification that are typical concerns for applied researchers. Section 4 describes the syntax and options of `ivmediate`. Section 5 provides a brief simulation exercise in sub-section 5.1 to show how `ivmediate` not only estimates the correct total effect of a treatment, but also how these can be decomposed into direct and indirect effects. We then apply the command to a real-life example using the data and empirical setting of Becker and Woessmann (2009) in sub-section 5.2 to estimate how Protestantism affects local economic performance in Prussian counties in 1877, and how much of this effect is causally mediated by literacy. The final section lists the results that are stored by `ivmediate`.

2 Causal Mediation Analysis in IV Models

Under linearity and with an instrument Z, the causal relations in *Model III* in Table 1 can be written as

$$Z = \epsilon_Z, \tag{1}$$
$$T = \beta_T^Z \cdot Z + \epsilon_T, \tag{2}$$
$$M = \beta_M^T \cdot T + \epsilon_M, \tag{3}$$
$$Y = \beta_Y^T \cdot T + \beta_Y^M \cdot M + \epsilon_Y, \tag{4}$$

Equations (1)–(4) can be compactly expressed as $\mathbf{X} = \mathbf{\Psi} \cdot \mathbf{X} + \boldsymbol{\epsilon}$ in (5):

$$\underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_T^Z & 0 & 0 & 0 \\ 0 & \beta_M^T & 0 & 0 \\ 0 & \beta_Y^T & \beta_Y^M & 0 \end{bmatrix}}_{\mathbf{\Psi}} \cdot \underbrace{\begin{bmatrix} Z \\ T \\ M \\ Y \end{bmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{bmatrix}}_{\epsilon}. \tag{5}$$

Equation (6) presents the covariance matrix $\Sigma_{\mathbf{X}}$ of observed variables $\mathbf{X}$:

$$\Sigma_{\mathbf{X}} \equiv \mathbf{Var}\begin{pmatrix} Z \\ T \\ M \\ Y \end{pmatrix} = \begin{bmatrix} \sigma_{ZZ} & \sigma_{ZT} & \sigma_{ZM} & \sigma_{ZY} \\ \cdot & \sigma_{TT} & \sigma_{TM} & \sigma_{TY} \\ \cdot & \cdot & \sigma_{MM} & \sigma_{MY} \\ \cdot & \cdot & \cdot & \sigma_{YY} \end{bmatrix}. \tag{6}$$

Let Σ_ϵ denote the covariance matrix of unobserved error terms $\boldsymbol{\epsilon}$. Since Z is an IV, it applies that ϵ_Z is statistically independent of $\epsilon_T, \epsilon_M, \epsilon_Y$. Thus, Σ_ϵ is given by

$$\Sigma_\epsilon \equiv \mathbf{Var}\begin{pmatrix} \epsilon_Z \\ \epsilon_T \\ \epsilon_M \\ \epsilon_Y \end{pmatrix} = \begin{bmatrix} \sigma^2_{\epsilon_Z} & 0 & 0 & 0 \\ \cdot & \sigma^2_{\epsilon_T} & \rho_{TM}\sigma_{\epsilon_T}\sigma_{\epsilon_M} & \rho_{TY}\sigma_{\epsilon_T}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \sigma^2_{\epsilon_M} & \rho_{MY}\sigma_{\epsilon_M}\sigma_{\epsilon_Y} \\ \cdot & \cdot & \cdot & \sigma^2_{\epsilon_Y} \end{bmatrix}. \tag{7}$$

The identifying assumption in Pinto et al. (2019) is that T is endogenous in a regression of Y on T, but endogeneity cannot arise from confounders that jointly influence T and Y, only from confounders that jointly affect T and M (e.g. Protestantism and literacy in Becker and Woessmann (2009)). The framework also allows for confounders that jointly influence M and Y (e.g. literacy and economic growth in Becker and Woessmann (2009)). Formally, the identifying assumption is $\rho_{TY} = 0$ in Σ_ϵ, while allowing $\rho_{TM} \neq 0$ and $\rho_{MY} \neq 0$.

In Section 5.1, we describe how to generate a simulated data set with these dependence relations.

3 Estimation

3.1 Estimation Procedure

The estimation equations to identify all linear coefficients are associated with well-known econometric estimators as follows (control variables are suppressed for notational simplicity and without loss of generality):

1. Parameter β_M^T is identified by standard two-stage least squares (2SLS) estimation, described by the two-equation system:

$$\text{First Stage:} \quad T = \beta_T^Z \cdot Z + \epsilon_T, \tag{8}$$

$$\text{Second Stage:} \quad M = \beta_M^T \cdot \hat{T} + \epsilon_M, \tag{9}$$

   where $\hat{T}$ stands for the estimated values of T in the first stage.

2. Pinto et al. (2019) show that the identifying assumption $\rho_{TY} = 0$ yields a new exclusion restriction, which allows for the use of Z as an instrument for M, when conditioned on T (but not unconditionally). This implies that β_Y^M and β_Y^T are the expected values of the estimators of a 2SLS regression where T plays the role of a conditioning variable, Z is the instrument, M is the endogenous variable, and Y is the dependent variable. Namely, β_Y^M and β_Y^T can be estimated by estimating the following 2SLS model:

$$\text{First Stage:} \quad M = \gamma_M^Z \cdot Z + \gamma_M^T \cdot T + \epsilon_T, \tag{10}$$

$$\text{Second Stage:} \quad Y = \beta_Y^M \cdot \hat{M} + \beta_Y^T \cdot T + \epsilon_Y, \tag{11}$$

   where $\hat{M}$ are the estimated values of M in the first stage.

The estimation procedure associated with equations (8) and (9) is the standard IV approach. By contrast, the estimation procedure associated with equations (10) and (11) is novel, and a property of the framework laid out in Pinto et al. (2019).

There are two first stages here in (8) and (10) for which `ivmediate` provides tests for weak identification by reporting the corresponding F-statistics on the excluded instrument. If robust or cluster-robust standard errors are requested, the regression output displays the F-statistic by Kleibergen and Paap (2006). To implement estimation of their corrected F-statistic, we rely on the `ranktest` command by Kleibergen and Schaffer (2007).

In Section 5.1, we compare the unbiased estimates resulting from equations (8)–(11) with the associated OLS estimates.

3.2 Interpretation

There is another explicit link between equations (8)–(11) and the direct estimation of the 'total effect' (TE) in *Model II* of Table 1. *Model II* is obtained from *Model III* by

substituting equation (9) into equation (11):

$$Y = \beta_Y^M \cdot \left(\beta_M^T \cdot T + \epsilon_M\right) + \beta_Y^T \cdot T + \epsilon_Y \tag{12}$$

$$= \underbrace{\left(\beta_Y^M \cdot \beta_M^T + \beta_Y^T\right)}_{TE} \cdot T + \underbrace{\beta_Y^M \epsilon_M + \epsilon_Y}_{\eta_Y} \equiv g_Y(T, \eta_Y). \tag{13}$$

Equation (13) shows that the direct estimate of TE produced by *Model II* is algebraically identical to the product of estimates $\beta_Y^T + \beta_M^T \cdot \beta_Y^M$ produced by *Model III* (i.e. equations (8)–(11)).[4] This algebraic equivalence holds for a *scalar* instrument Z, but may *not* hold with a vector of instruments Z'. The `ivmediate` command therefore is limited to the use of a single scalar instrument.[5]

It is also worth noting that in the mediation framework, either β_Y^T or $\beta_M^T \cdot \beta_Y^M$ (but not both) can have opposite signs. For example, there is nothing logically inconsistent about having a positive total effect which is composed of a (larger) positive indirect effect that is partly offset by a negative direct effect, or vice versa. In such as case, a statement like "the indirect effect explains more than 100 percent of the total effect" is not incorrect, but requires careful explanation to avoid confusion.

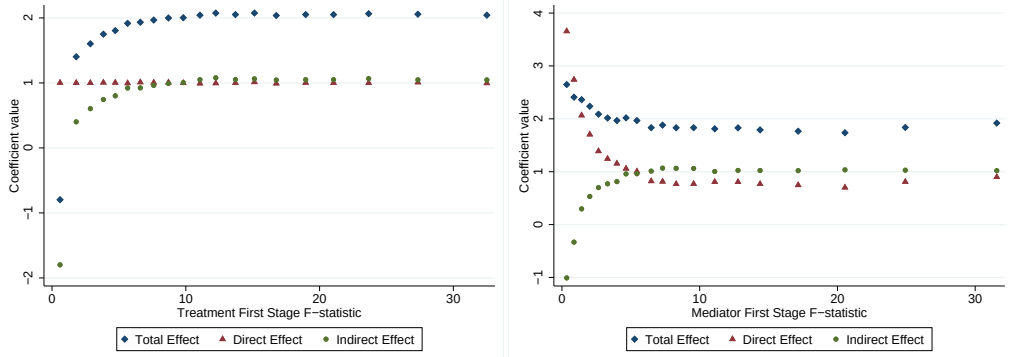
3.3 Weak Identification with two First Stage Regressions

Applied researchers are now well aware of the bias introduced by weak identification in instrumental variables setting (Bound et al. 1995). A commonly used rule of thumb is that an F-test of the excluded instrument(s) in the first stage should yield an F statistic of 10 or more (Stock and Yogo 2005). How does this apply to the IV mediation setting with two first stages? Currently there is no theory to guide applied researchers. Instead, we apply the code from section 5.1 to simulate the behavior of the estimator under different instrument strengths in the treatment and the mediator first stages. This is done by varying the amount of noise in ϵ_T and ϵ_Y.

Figure 1 plots the coefficient values of the total, direct, and indirect effects over different values of the first stage F-statistic. The left panel manipulates the strength of the instrument in the treatment first stage and the right panel that in the mediation first stage. The instrument is only ever weak in one of the two first stages but not in both at the same time. Samples were simulated according to (1)–(4) with 1,000 observations for each value of the error variance. The values increase from 1 and 15 in increments of 0.5. In the example, the true values of the direct and indirect effects are both 1, summing up to a true total effect of 2. The `ivmediate` simulations were then run 100 times for each error variance value.

4. To see that the direct estimation of *Model II* requires Z as an IV, note that the correlation between ϵ_M and ϵ_T also gives rise to a correlation between η_Y and ϵ_T, while $Z \perp\!\!\!\perp (\epsilon_T, \epsilon_M, \epsilon_Y)$ also implies the independence $Z \perp\!\!\!\perp (\epsilon_T, \eta_Y)$.
5. As with standard 2SLS regression, multiple instruments can be applied to predict a single endogenous variable. However, the resulting second stage coefficient on the endogenous variable will be a GMM-weighted average of the prediction coming from the different instruments' first stage coefficients with weights being determined by the relative importance of each instrument. This makes it difficult to interpret the second stage result.

Figure 1: Coefficient Values under Differing IV Strengths in either First Stage



Note: The left panel simulated data for different values of $Var(\epsilon_T)$, the right for different values of $Var(\epsilon_Y)$ ranging from 1 to 15. The value of $Var(\epsilon_T)$ increases in steps of 0.5 and at each step, 100 random samples were drawn according to (1)–(4) with 1,000 observations. Both panels show binned scatter plots of coefficient values of the total, direct, and indirect effects over different values of the corresponding first stage F-statistics where the strength of the instrument was manipulated. The true total effect is 2, and the true direct and indirect effects are equal to 1.

The left panel shows that as the treatment first stage F-statistic approaches the rule of thumb value of 10, all effects begin to center on their true values. This is also the case for the right panel, however, here the direct effect takes longer to center on its true value. It only begins from a mediation first stage F-statistic of 30. A conservative approach would therefore require a stronger instrument in the mediator first stage in order to accurately identify all three effects. If interest only lies on the indirect effect, the commonly used approximation rule for a reasonably strong instrument seems applicable.

4 Implementation

4.1 Syntax

`ivmediate` *depvar* [*indepvars*] [*if*] [*in*], `mediator(`*varname*`)`
`treatment(`*varname*`)` `instrument(`*varname*`)` [`absorb(`*varname*`)` `full`
`vce(`*vcetype*`)` `level(`#`)`]

4.2 Description

`ivmediate` implements the causal mediation analysis framework for IV models introduced by Pinto et al. (2019). The command allows to estimate the causal treatment and mediation effects for potentially endogenous treatment and mediator variables without

the need for an additional instrument for the mediator. A single instrumental variable suffices to identify both effects.

4.3 Options

Supported options are:

`mediator(`*varname*`)` includes a single mediator variable (required).

`treatment(`*varname*`)` includes a single treatment variable (required).

`instrument(`*varname*`)` includes a single instrumental variable (required).

`absorb(`*varname*`)` allows to absorb one fixed effect. For details see `help areg`.

`full` display intermediate results together with the main results. Specifying this option will display three intermediate output tables:

1. the IV regression of Y on T (instrumented with Z)
2. the IV regression of M on T (instrumented with Z) for which the first stage F statistic is reported as *first stage one* in the main table
3. the IV regression of Y on M (instrumented with Z) and controlling for T for which the first stage F statistic is reported as *first stage two* in the main table

the total effect is the coefficient on T in 1.; the direct effect is the coefficient on T in 3.; the indirect effect is the product of the coefficient on T in 2. and the coefficient on M in 3. The mediation effect as percentage of the total effect is therefore the indirect effect divided by the total effect times one hundred.

`vce(`*vcetype*`)` may be `robust` to estimate Eicker/Huber/White standard errors, or `cluster` *clustervar* to estimate cluster robust standard errors. Not specifying an option leads to estimation of unadjusted standard errors, which is the default.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. Integers between 10 and 99 including are allowed. The default is `level(95)`; see [U] **20.7 Specifying the width of confidence intervals**.

5 Empirical Example

5.1 Simulation Exercise

A simulated data set with the assumed dependence relations can be straightforwardly generated in the following way:

- Separately generate error terms ϵ_T, ϵ_Y that are normally distributed with mean zero and variance one, $N(0,1)$. These are statistically independent, i.e. $\epsilon_T \perp\!\!\!\perp \epsilon_Y$.

- Let error term ϵ_M be defined as $\epsilon_M = \sqrt{\omega} \cdot \epsilon_T + \sqrt{(1-\omega)} \cdot \epsilon_Y$ for any $\omega \in [0,1]$.[6]

The correlation between ϵ_M, ϵ_T is given by $\rho_{TM} = \sqrt{\omega}$. Thereby $\epsilon_M \not\perp\!\!\!\perp \epsilon_T$. By symmetry, we also have that $\rho_{MY} = \sqrt{(1-\omega)}$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$. Having drawn ϵ_T, ϵ_Y independently implies that the correlation between ϵ_T and ϵ_Y is $\rho_{TY} = 0$. However, conditioning on $\epsilon_M = e$ induces a linear relation between ϵ_T, ϵ_Y, namely, $\epsilon_T = e/\sqrt{\omega} - \sqrt{(1-\omega)/\omega} \cdot \epsilon_Y$. Thus, the correlation between ϵ_T, ϵ_Y conditioned on ϵ_M is $\rho_{TY|\epsilon_M} = -1$ and thereby $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y | \epsilon_M$. A high ω implies a high ρ_{TM}. By contrast, a low ω implies a high ρ_{MY}.

It is instructive to investigate the bias generated by a misspecified model where T, M are assumed to be exogenous, i.e. the mutual independence of $\epsilon_T, \epsilon_M, \epsilon_Y$ is wrongly assumed. Let the data be generated by equations (1)–(4), and the model coefficients be normalized to equal 1, that is, $\beta_T^Z = \beta_M^T = \beta_Y^T = \beta_Y^M = 1$. The true parameters β_M^T, β_Y^T and β_Y^M are identified through equations (8)–(11). If the error terms $\epsilon_T, \epsilon_Y, \epsilon_M$ were wrongly assumed to be statistically independent, parameters $\beta_M^T, \beta_Y^T, \beta_Y^M$ could be estimated by OLS through the following equations:

$$\text{OLS:} \quad \beta_M^T = \frac{\sigma_{TM}}{\sigma_{TY}}, \tag{14}$$

$$\text{OLS:} \quad \beta_Y^T = \frac{\sigma_{MM}\sigma_{TY} - \sigma_{TM}\sigma_{MY}}{\sigma_{MM}\sigma_{TT} - \sigma_{TM}^2}, \tag{15}$$

$$\text{OLS:} \quad \beta_Y^M = \frac{-\sigma_{TM}\sigma_{TY} + \sigma_{TT}\sigma_{MY}}{\sigma_{MM}\sigma_{TT} - \sigma_{TM}^2}. \tag{16}$$

While the true parameters are set to be 1, the OLS estimators may range from 0 to 2 depending on the error correlations. Since a high ω implies pronounced bias in the relation between T and M (a high ρ_{TM}), the OLS estimate of β_M^T diverges from the true value 1 as ω increases. By contrast, the OLS estimates of β_Y^T and β_Y^M converges to the true value 1.

```
. * set seed for replicability
. set seed 12345
.
. * weights for the mediation error
. global omega = 0.5
.
. * model parameters
. global betaYT = 1
. global betaYM = 1
. global betaMT = 1
.
. cap program drop ivmedsym
. program ivmedsym
  1.         clear
  2.         set obs 1000
```

6. Note that $\epsilon_T \sim N(0,1)$ and $\epsilon_Y \sim N(0,1)$ imply $\epsilon_M \sim N(0,1)$.

```
  3.
.         * generate error terms as described in the paper
.         gen e_t = rnormal(0,1)
  4.         gen e_y = rnormal(0,1)
  5.         gen e_m = sqrt($omega)*e_t + sqrt(1-$omega)*e_y
  6.
.         * generate variables according to eq (1)-(4) in sect 2
.         gen z = rnormal(0,1)
  7.         gen t = z + e_t
  8.         gen m = t*$betaMT + e_m
  9.         gen y = t*$betaYT + m*$betaYM + e_y
 10.
.         * naive OLS
.         reg y t
 11.         scalar bols = _b[t]
 12.
.         * ivmediate regression
.         ivmediate y, mediator(m) treatment(t) instrument(z)
 13.         scalar te = _b["total effect"]
 14.         scalar de = _b["direct effect"]
 15.         scalar ie = _b["indirect effect"]
 16. end
.
. simulate b_ols = bols b_total = te b_direct = de b_indirect = ie, reps(200): ivmedsym
      command:  ivmedsym
        b_ols:  bols
      b_total:  te
     b_direct:  de
   b_indirect:  ie
Simulations (200)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................   50
..................................................  100
..................................................  150
..................................................  200
.
. summarize
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       b_ols |        200    2.355732    .0404492    2.25768   2.454119
     b_total |        200    2.003096    .0561753   1.859572   2.117501
    b_direct |        200    1.004551    .0867107   .8003523   1.274698
  b_indirect |        200    .9985453    .0556392   .8147842   1.141245
```

Given the model parameters, the total effect of $\beta_Y^M \cdot \beta_M^T + \beta_Y^T = 1 \cdot 1 + 1 = 2$ is not recovered by simple OLS. In fact, not even the 95% confidence interval would include the true total effect. On the other hand, 2SLS did recover the total effect but it could not disentangle the direct effect of the treatment (net of the mediator) from the indirect effect of the mediating variable. The simulation shows how `ivmediate` can recover both the true total effect and decompose it into the direct and indirect effects as described in the theoretical section.

5.2 Applied Example using the Becker and Woessmann (2009) Data

The example below uses data from Becker and Woessmann (2009) who estimate the effect of Protestantism on economic prosperity in Prussian counties. To obtain exogenous variation in the share of Protestants in these counties, they used the fact that Protestantism spread concentrically around Wittenberg, the city where Martin Luther taught and preached. Following their example, we use distance to Wittenberg (`kmwitt`) as instrument for the share of Protestants (`f_prot`) with the outcome being the per capita income tax (`inctax`) in 1877 as measure for economic performance. The mediator we consider is the share of literate population (`f_rw`).

According to Becker and Woessmann (2009), Protestantism promoted reading of the bible which lead to human capital accumulation and therefore promoted economic development. They are interested in estimating,

$$Y = \alpha \text{Prot} + \chi \text{Lit} + X'\gamma + \epsilon \tag{17}$$

though they note that the "problem with such a model is that not only Protestantism but also literacy may be endogenous in this setting" (p. 570). Since they have no additional instrument for literacy, they employ different types of bounding exercises using estimates from previous literature on the returns to education (see section VI.C in the original study). Using `ivmediate`, we can go further and directly estimate the mediation effect of literacy that goes through Protestantism with only one instrument.

```
. use "ipehd_qje2009_master.dta"
. global controls "f_jew f_fem f_young f_pruss hhsize pop gpop f_miss"
. ivmediate inctax $controls, mediator(f_rw) treatment(f_prot) instrument(kmwitt)
Linear IV Mediation Analysis
------------------------------

Outcome:   inctaxpc                                    Number of obs = 426
Treatment: f_prot
Mediator:  f_rw
------------------------------------------------------------------------------
inctaxpc        |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
total effect    |   .8347728   .2723283     3.07   0.002     .3010192    1.368526
direct effect   |   .0826879   .0825493     1.00   0.316    -.0791057    .2444815
indirect effect |   .7520849   .2912821     2.58   0.010     .1811824    1.322987
------------------------------------------------------------------------------
Mediator f_rw explains 90.09% of the total effect.
F-statistic for excluded instruments in
- first stage one (T on Z):   48.394
- first stage two (M on Z|T): 65.274
Excluded instruments: kmwittenberg
------------------------------------------------------------------------------
```

As in the original study, we condition on further covariates in the estimation of eq. (17) which are the share of Jewish population, female population, individuals aged below 10, the share of population of Prussian origin, average household size, population size of the county, the percentage population growth between 1867 and 1871, and the share of the population with missing information on literacy.[7] The output is displayed

7. For brevity, we omit their controls for the share of population with physical or mental disabilities

below.

The total effect estimates that every one percentage point increase in the share of Protestants increases per capita income tax revenues by 0.83 Marks. Under the typical instrumental variables assumptions, this effect is causal. The direct effect estimates that only 0.08 Marks of this increase are due to Protestantism itself and it is not statistically significant. However, the indirect effect estimates that 0.75 Marks of this increase are caused by literacy as a mediating factor. This implies that literacy explains 90% of the total effect of Protestantism on economic outcomes. This is in line with the findings by Becker and Woessmann (2009) who conclude that "Protestants' higher literacy can account for roughly the whole gap in economic outcomes between the two denominations [Catholics and Protestants]" (p. 576).

6 Stored Results

`ivmediate` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(fstat1)`	F-statistic for the excluded instruments in first stage one (T on Z)
`e(fstat2)`	F-statistic for the excluded instruments in first stage two (M on Z|T)
`e(mepct)`	mediation effect expressed as percentage of the total effect
`e(N_clust)`	number of clusters used to adjust standard errors if `cluster` was specified

Macros

`e(depvar)`	name of the dependent variable
`e(treat)`	name of the treatment variable
`e(med)`	name of the mediator variable
`e(inst)`	name(s) of the instrumental variable(s)
`e(vcetype)`	vcetype specified in `vce(`*vcetype*`)`
`e(clustvar)`	name of the cluster variable if `cluster` was specified in `vce(`*vcetype*`)`

Matrices

`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix

(blind, deaf-mute, and insane), as these do not significantly affect the results.

7 References

Baron, R. M., and D. A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51(6): 1173.

Becker, S. O., and L. Woessmann. 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics* 124: 531–596.

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

Frölich, M., and M. Huber. 2017. Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(5): 1645–1666.

Imai, K., L. Keele, and D. Tingley. 2010. A General Approach to Causal Mediation Analysis. *Psychological Methods* 15(4): 309–334.

Jun, S. J., J. Pinkse, H. Xu, and N. Yildiz. 2016. Multiple Discrete Endogenous Variables in Weakly-Separable Triangular Models. *Econometrics* 4(1): 288–303.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular-value decomposition. *Journal of Econometrics* 127: 97–126.

Kleibergen, F., and M. E. Schaffer. 2007. ranktest: Stata module to test the rank of a matrix using the Kleibergen–Paap rk statistic. Boston College Department of Economics, Statistical Software Components S456865. Downloadable from http://ideas.repec.org/c/boc/bocode/s456865.html.

MacKinnon, D. P. 2008. *Introduction to statistical mediation analysis*. Routledge.

Pinto, R., C. Dippel, R. Gold, and S. Heblich. 2019. Mediation Analysis in IV Settings With a Single Instrument. *UCLA working paper*.

Stock, J. H., and M. Yogo. 2005. Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. H. Stock, 80–108. Cambridge: Cambridge University Press.

About the authors

Christian Dippel is an assistant professor of economics in the Anderson School of Management at the University of California, Los Angeles.

Andreas Ferrara is an assistant professor of economics in the Department of Economics at the University of Pittsburgh.

Stephan Heblich is an associate professor in the Munk School of Global Affairs and Public Policy and the Department of Economics at the University of Toronto.