The Stata Journal (2020)
20, Number 3, pp. 613–626 DOI: 10.1177/1536867X20953572

Causal mediation analysis in instrumental-variables regressions

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

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

Stephan Heblich
University of Toronto
Toronto, Canada
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 settings using the framework developed by Dippel et al. (2020, unpublished manuscript). `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 the mediation effects.

Keywords: st0611, 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. Because 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, that is, 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 nonrandom and

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

therefore 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 Dippel et al. (2020). This complements existing ways to estimate causal mediation effects that assume randomness in the assignment of treatment T (Imai, Keele, and Tingley 2010) or require separate instruments for T and M (for example, 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 (that is, $\epsilon_T \not\perp\!\!\!\perp \epsilon_M$) and we introduce for the endogenous treatment T an IV Z, which is both uncorrelated with the omitted variables ($Z \perp\!\!\!\perp \epsilon_T, \epsilon_M$) and a reasonably strong predictor of T. Model II fits the TE of T on Y using the same IV approach: $\epsilon_T \not\perp\!\!\!\perp \eta_Y$, but Z is exogenous (that is, $Z \not\perp\!\!\!\perp \epsilon_T, \eta_Y$). Table 1 is reprinted from Dippel et al. (2020).

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 ordinary least squares (OLS) to estimate three equations,

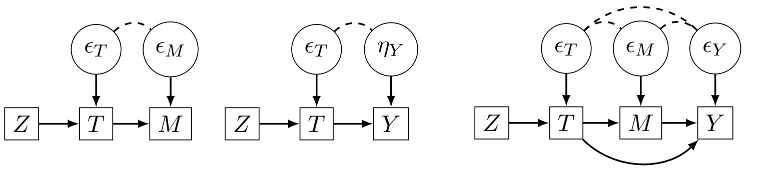
$$Y = \delta_Y^T \times T + \eta_Y, \quad M = \beta_M^T \times T + \epsilon_M, \quad \text{and} \quad Y = \beta_Y^T \times T + \beta_Y^M \times M + \epsilon_Y$$

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

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 effects 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 nonparametric structural equations of each model. We use $\perp\!\!\!\perp$ to denote statistical independence.

To identify what fraction of the TE is explained by the indirect effect, we have to perform a mediation analysis that decomposes the TE of T on Y into 1) the mediated "indirect" effect of T on Y that operates through M and 2) the residual "direct" effect that does not work through 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 (that is, $\epsilon_T \not\perp\!\!\!\perp \epsilon_Y$ and $\epsilon_M \not\perp\!\!\!\perp \epsilon_Y$), but there is only one instrument Z to address this endogeneity.

To overcome the underidentification problem, we do not assume away endogeneity in any of the key relationships in model III ($\epsilon_T \not\perp\!\!\!\perp \epsilon_M$, $\epsilon_M \perp\!\!\!\perp \epsilon_Y$, and $\epsilon_T \perp\!\!\!\perp \epsilon_Y$ are all maintained), 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 when T is endogenous in a regression of Y on T because of the same confounders that affect Y primarily through M.

Dippel et al. (2020) show that this assumption alone is sufficient to unpack the causal channels in model III, therefore allowing 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 two-stage least squares (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 on 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 section 5.1 to show not only how `ivmediate` estimates the correct TE 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 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.

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

$$Z = \epsilon_Z \tag{1}$$

$$T = \beta_T^Z \times Z + \epsilon_T \tag{2}$$

$$M = \beta_M^T \times T + \epsilon_M \tag{3}$$

$$Y = \beta_Y^T \times T + \beta_Y^M \times M + \epsilon_Y \tag{4}$$

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

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

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

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

Let $\mathbf{\Sigma}_\epsilon$ denote the covariance matrix of unobserved error terms $\boldsymbol{\epsilon}$. Because Z is an IV, it implies that ϵ_Z is statistically independent of ϵ_T, ϵ_M, and ϵ_Y. Thus, $\mathbf{\Sigma}_\epsilon$ is given by

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

The identifying assumption in Dippel et al. (2020) 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 (for example, Protestantism and literacy in Becker and Woessmann [2009]). The framework also allows for confounders that jointly influence M and Y (for example, literacy and economic growth in Becker and Woessmann [2009]). Formally, the identifying assumption is $\rho_{TY} = 0$ in $\mathbf{\Sigma}_\epsilon$, while allowing $\rho_{TM} \neq 0$ and $\rho_{MY} \neq 0$.

In section 5.1, we describe how to generate a simulated dataset with these dependent 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^T_M is identified by standard 2SLS estimation, described by the following two-equation system:

$$\text{First stage:} \quad T = \beta^Z_T \times Z + \epsilon_T \tag{7}$$

$$\text{Second stage:} \quad M = \beta^T_M \times \widehat{T} + \epsilon_M \tag{8}$$

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

2. Dippel et al. (2020) 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^M_Y and β^T_Y 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^M_Y and β^T_Y can be fit by the following 2SLS model:

$$\text{First stage:} \quad M = \gamma^Z_M \times Z + \gamma^T_M \times T + \epsilon_T \tag{9}$$

$$\text{Second stage:} \quad Y = \beta^M_Y \times \widehat{M} + \beta^T_Y \times T + \epsilon_Y \tag{10}$$

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

The estimation procedure associated with (7) and (8) is the standard IV approach. By contrast, the estimation procedure associated with (9) and (10) is novel and a property of the framework laid out in Dippel et al. (2020).

There are two first stages here in (7) and (9) 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 (7)–(10) with the associated OLS estimates.

3.2 Interpretation

There is another explicit link between (7)–(10) and the direct estimation of the TE in model II of table 1. Model II is obtained from model III by substituting (8) into (10):

$$\begin{aligned} Y &= \beta_Y^M \times \left(\beta_M^T \times T + \epsilon_M\right) + \beta_Y^T \times T + \epsilon_Y \\ &= \underbrace{\left(\beta_Y^M \times \beta_M^T + \beta_Y^T\right)}_{\text{TE}} \times T + \underbrace{\beta_Y^M \epsilon_M + \epsilon_Y}_{\eta_Y} \equiv g_Y(T, \eta_Y) \end{aligned} \tag{11}$$

Equation (11) shows that the direct estimate of TE produced by model II is algebraically identical to the product of estimates $\beta_Y^T + \beta_M^T \times \beta_Y^M$ produced by model III [that is, (7)–(10)].[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 $\beta_M^T \times \beta_Y^M$ (but not both) can have opposite signs. For example, there is nothing logically inconsistent about having a positive TE that is composed of a (larger) positive indirect effect that is partly offset by a negative direct effect, or vice versa. In such a case, a statement like "the indirect effect explains more than 100 percent of the total effect" is not incorrect, but it does require careful explanation to avoid confusion.

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 generalized method of moment 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.

3.3 Weak identification with two first-stage regressions

Applied researchers are now well aware of the bias introduced by weak identification in an IV setting (Bound, Jaeger, and Baker 1995). A 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 manipulates 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 to 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 TE of 2. The `ivmediate` simulations were then run 100 times for each error variance value.

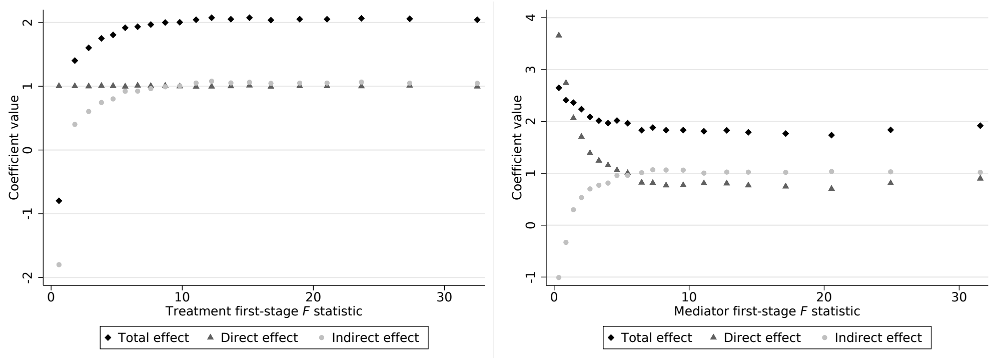


Figure 1. Coefficient values under differing IV strengths in either first stage.
NOTE: The left panel simulated data for different values of $\mathrm{Var}(\epsilon_T)$, and the right simulated data for different values of $\mathrm{Var}(\epsilon_Y)$ ranging from 1 to 15. The value of $\mathrm{Var}(\epsilon_T)$ 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 TE 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 to center on the true value from a mediation first-stage F statistic of 30. A conservative approach would therefore require a stronger instrument in the mediator first stage 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 The ivmediate command

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 Dippel et al. (2020). The command allows the estimation of 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 IV suffices to identify both effects.

4.3 Options

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

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

`instrument(`*varname*`)` includes a single IV. `instrument()` is required.

`absorb(`*varname*`)` allows the absorption of one fixed effect. For details, see [R] **areg**.

`full` displays 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 TE is the coefficient on T in the first table; the direct effect is the coefficient on T in the third table; the indirect effect is the product of the coefficient on T in the second table and the coefficient on M in the third. The mediation effect as percentage of the TE is therefore the indirect effect divided by the TE times 100.

`vce(`*vcetype*`)` may be `robust` to estimate Eicker/Huber/White standard errors or may be `cluster` *clustervar* to estimate cluster–robust standard errors. The default is `vce(unadjusted)` standard errors.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. Integers between 10 and 99 inclusive are allowed. The default is `level(95)` or as set by `set level`; see [U] **20.8 Specifying the width of confidence intervals**.

4.4 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 $Z \mid T$)
`e(mepct)`	mediation effect expressed as percentage of the TE
`e(N_clust)`	number of clusters used to adjust standard errors if `cluster` was specified in `vce()`
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 of the IV
`e(vcetype)`	*vcetype* specified in `vce()`
`e(clustvar)`	name of the cluster variable if `cluster` was specified in `vce()`
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix

5 Empirical example

5.1 Simulation exercise

A simulated dataset with the assumed dependent relations can be straightforwardly generated in the following way:

- Separately generate error terms ϵ_T and ϵ_Y that are normally distributed with mean 0 and variance 1, $N(0,1)$. These are statistically independent, that is, $\epsilon_T \perp\!\!\!\perp \epsilon_Y$.
- Let error term ϵ_M be defined as $\epsilon_M = \sqrt{\omega} \times \epsilon_T + \sqrt{(1-\omega)} \times \epsilon_Y$ for any $\omega \in [0,1]$.[6]

The correlation between ϵ_M and 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 and 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 and ϵ_Y, namely,

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

$\epsilon_T = e/\sqrt{\omega} - \sqrt{(1-\omega)/\omega} \times \epsilon_Y$. Thus, the correlation between ϵ_T and 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 and M are assumed to be exogenous, that is, the mutual independence of ϵ_T, ϵ_M, and ϵ_Y is wrongly assumed. Let the data be generated by (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 (7)–(10). If the error terms ϵ_T, ϵ_Y, and ϵ_M were wrongly assumed to be statistically independent, then parameters β_M^T, β_Y^T, and β_Y^M could be estimated by OLS through the following equations:

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

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

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

While the true parameters are set to be 1, the OLS estimators may range from 0 to 2 depending on the error correlations. Because 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 converge 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
. capture program drop ivmedsym
. program ivmedsym
  1. clear
  2. set obs 1000
  3. * generate error terms as described in the article
. generate e_t = rnormal(0,1)
  4. generate e_y = rnormal(0,1)
  5. generate e_m = sqrt($omega)*e_t + sqrt(1-$omega)*e_y
  6. * generate variables according to (1)-(4) in section 2
. generate z = rnormal(0,1)
  7. generate t = z + e_t
  8. generate m = t*$betaMT + e_m
  9. generate y = t*$betaYT + m*$betaYM + e_y
 10. * naive OLS
. regress 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 TE of $\beta_Y^M \times \beta_M^T + \beta_Y^T = 1 \times 1 + 1 = 2$ is not recovered by simple OLS. In fact, not even the 95% confidence interval would include the true TE. On the other hand, 2SLS did recover the TE, 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 both recover the true TE 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 an instrument for the share of Protestants (`f_prot`) with the outcome being the per capita income tax (`inctax`) in 1877 as a 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 led to human capital accumulation and therefore promoted economic development. They are interested in estimating

$$Y = \alpha \text{Prot} + \chi \text{Lit} + \mathbf{X}'\boldsymbol{\gamma} + \epsilon \tag{12}$$

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). Because they have no additional instrument for literacy, they use 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
. 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 (12), 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 TE estimates that every 1 percentage point increase in the share of Protestants increases per capita income tax revenues by 0.83 Marks. Under the typical IV assumptions, this effect is causal. The direct effect estimates that only 0.08 Marks of this increase are because of 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 TE 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).

7. For brevity, we omit their controls for the share of population with physical or mental disabilities (blind, deaf-mute, and insane), as these do not significantly affect the results.

6 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-3
. net install st0611      (to install program files, if available)
. net get st0611          (to install ancillary files, if available)
```

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: 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173.

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. https://doi.org/10.1162/qjec.2009.124.2.531.

Bound, J., D. A. Jaeger, and R. M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443–450. https://doi.org/10.2307/2291055.

Dippel, C., R. Gold, S. Heblich, and R. Pinto. 2020. Mediation analysis in IV settings with a single instrument. Unpublished manuscript.

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* 79: 1645–1666. https://doi.org/10.1111/rssb.12232.

Imai, K., L. Keele, and D. Tingley. 2010. A general approach to causal mediation analysis. *Psychological Methods* 15: 309–334. https://doi.org/10.1037/a0020761.

Jun, S. J., J. Pinkse, H. Xu, and N. Yildiz. 2016. Multiple discrete endogenous variables in weakly-separable triangular models. *Econometrics* 4: 7. https://doi.org/10.3390/econometrics4010007.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133: 97–126. https://doi.org/10.1016/j.jeconom.2005.02.011.

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

MacKinnon, D. P. 2008. *Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum.

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. New York: 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.