Causal Ecological Inferences

Jörg L. Spenkuch
Northwestern University

First Draft: February 2017
This Version: February 2018

Abstract

This note observes that the ecological inference problem is very closely related to the problem of estimating causal effects. Due to the common mathematical structure, instrumental variables techniques solve the ecological inference problem under essentially the same set of conditions under which they allow researchers to draw conclusions about causality. More generally, carefully addressing the issue of causality can be sufficient to generalize from aggregate data to the behavior of individuals.

*Previous versions circulated under the title "Ecological Inference with Instrumental Variables." I have benefitted from helpful conversations with Pablo Montagnes and Philipp Tillmann. I am also grateful to David Austen-Smith and Tim Feddersen for encouraging me to write up this note, and to Derek Neal for teaching me that regressions are always about the residual. Correspondence can be addressed to the author at MEDS Department, Kellogg School of Management, 2211 Campus Drive, Evanston, IL 60208, or by email: j-spenkuch@kellogg.northwestern.edu.

1. Introduction

Most empirical work in the modern social sciences aims to estimate causal effects, i.e., the expected change in some outcome y that is attributable to a change in x, holding all else equal. Even with access to microdata, producing valid *ceteris paribus* comparisons is often challenging. To complicate matters even further, many interesting questions can only be addressed with aggregate data. If the behavioral relationship between x and y is theoretically defined at the individual level, then researchers using such data to draw inferences about causal effects must also worry about the ecological fallacy.

The ecological fallacy arises because ecological correlations, i.e., correlations in aggregated data, may systematically deviate from their individual-level counterparts. Based on this observation, Robinson (1950) cogently argues that aggregate correlations are meaningless and should not be used to draw inferences about individuals. Yet, in many applications—especially historical ones—microdata are simply not available. Scholars studying these settings cannot help but rely on aggregate data to learn about individual behavior.

In light of this predicament, a large literature proposes ways to avoid the ecological fallacy, including imposing homogeneity assumptions (Goodman 1953, 1959), calculating deterministic bounds (Duncan and Davis 1953), as well as parametric procedures that combine deterministic and probabilistic information (King 1997; King et al. 2004). All of these methods work well in some situations but are known to yield either uninformative or, worse, misleading estimates in others (see, e.g., Freedman et al. 1991; Gelman et al. 2001). In addition, existing approaches were designed for descriptive rather than causal inferences. That is, they recover raw individual-level differences, *without* holding all else equal. Even extensions of King's method that allow for covariate adjustment cannot account for unobserved confounders (King et al. 1999, 2004). If there are any (unmeasured) variables that correlate with both x and y, then estimates from conventional approaches to the ecological inference problem do not have a causal interpretation. They do not correspond to valid *ceteris paribus* comparisons.

At first blush, it may seem that recovering causal, individual-level estimates from aggregate data is a near-hopeless endeavor. This note shows that such a conclusion is too pessimistic. In fact, as demonstrated below, standard instrumental variables techniques solve the ecological inference problem under almost the same assumptions that would be required to claim causality in microdata. More generally, carefully addressing the problem of causal identification is in many applications sufficient to extrapolate from aggregate data to the behavior of individuals.

The observations in this note are closely related to the work of Cross and Manski (2002), who study the ecological inference problem through the lens of a partial-identification framework. Among other things, they show that an exclusion restriction on the regressors coupled with a full rank condition can be used to point-identify the object of interest. Despite the substantive similarity of their result with Proposition 1 below, several key differences remain. Their characterization of the identified set depends critically on the assumption that all covariates are discrete, which rules out many interesting applications in which researchers may wish to control for continuous variables, such as age, income, segregation, etc. This note imposes no such restriction. Further, Cross and Manski (2002) provide little to no guidance on how to put their identifcation proof to use in practice. By contrast, everything that follows can be *easily* implemented in standard statistical software packages like STATA or R.

2. From Individual-Level Behavior to Ecological Correlations and Back

To clarify why aggregate correlations can generally not be interpreted as individual-level effects, and to convey the intuition for why causal and ecological inferences go hand-in-hand, it is useful to start with a statistical description of individual behavior. To this end, consider a data-generating process (DGP) of the following form: $y_i = \beta x_i + \mathbf{W}_i{}'\theta + \mathbf{Q}_a{}'\gamma + \epsilon_i$, where i indexes individuals, and y_i denotes the outcome of interest. $\mathbf{W}_i$ is a vector of all individual-level covariates that affect behavior, while $\mathbf{Q}_a$ contains contextual factors, which are the same for everyone in aggregate unit a. Depending on the application, $\mathbf{W}_i$ may include age, gender, income, etc., whereas $\mathbf{Q}_a$ might consist of controls for socioeconomic inequality, segregation, or, say, the quality of schools. Since the DGP includes all variables that have a systematic impact on y_i, x_i and ϵ_i are conditionally uncorrelated and β denotes the causal effect of x.

If such individual-level data were available and if $\mathbf{W}_i$ and $\mathbf{Q}_a$ were fully observed, then estimating β would be straightforward. Often, however, microdata are not available, and researchers may not have access to all of the behaviorally relevant variables. In the typical application, the available data take the form averages, i.e.,

$$
\begin{array}{ccccccccccc}
 & \frac{1}{N_c}\sum_{i=1}^{N_c} y_i & = & \beta\frac{1}{N_c}\sum_{i=1}^{N_c} x_i & + & \frac{1}{N_c}\sum_{i=1}^{N_c}\mathbf{W}_i'\theta & + & \frac{1}{N_c}\sum_{i=1}^{N_c}\mathbf{Q}_a'\gamma & + & \frac{1}{N_c}\sum_{i=1}^{N_c}\epsilon_i & \\
\Rightarrow & \overline{y}_c & = & \beta\overline{x}_c & + & \overline{\mathbf{W}}_c'\theta & + & \overline{\mathbf{Q}}_c'\gamma & + & \overline{\epsilon}_c & ,
\end{array} \tag{1}
$$

where c refers to the unit of aggregation, say, counties, N_c is the number of individuals within each unit, and upper bars denote unit-level means.[1]

Estimating (1) by OLS yields a point estimate with plim $\widehat{\beta}_{OLS} = \beta + \text{Cov}\left(\overline{x}_c^*, \overline{\epsilon}_c\right) / \text{Var}\left(\overline{x}_c^*\right)$, where $\overline{x}_c^*$ denotes the residual from projecting $\overline{x}_c$ on the set of controls (cf. Frisch and Waugh 1933). As a result, an ecological regression recovers β, the true individual-level parameter, *if and only if* $\text{Cov}\left(\overline{x}_c^*, \overline{\epsilon}_c\right) = 0$ (see also Langbein and Lichtman 1978, ch. 2; King 1997, ch. 3). Put differently, ecological regressions produce biased results unless the regressor of interest is uncorrelated with the error term.

Note, this is exactly the same condition that is required for OLS estimates to have a causal interpretation. There are, of course, many research designs that, under reasonable assumptions, allow researchers to obtain causal estimates. If the respective assumptions also hold in the aggregated data, then the following corollary implies that any of these methods can be used to draw *causal ecological inferences*.

Corollary: *Provided an ecological regression model can be microfounded by averaging over the individual-level DGP, obtaining causal estimates is sufficient to extrapolate from the regression coefficients to individual-level effects.*

[1] a and c may but need not be at the same level of aggregation.

Nonetheless, drawing valid causal inferences from aggregate data is generally more difficult than if the microdata were available. To see why, suppose that x is uncorrelated with the error term on the individual-level. It still need not be the case that plim $\widehat{\beta}_{OLS} = \beta$. Since, for any unit of aggregation c, $\text{Cov}\left(\overline{x}_c^*, \overline{\epsilon}_c\right) = \frac{1}{N_c^2} \sum_i \sum_j \text{Cov}\left(x_i^*, \epsilon_j\right)$, bias may arise whenever $\text{Cov}\left(x_i^*, \epsilon_j\right) \neq 0$ for some subset of individuals $i \neq j$ within the same c. Aggregation, therefore, makes both causal and ecological inference harder.

3. Ecological Inference with Instrumental Variables (EI-IV)

Conveniently, standard instrumental variables techniques solve both problems under essentially the same set of assumptions that are required for IV analyses in microdata. To see this, suppose the researcher has access to an instrument, z_c, that is *(i)* relevant, i.e., it predicts x conditional on covariates, and *(ii)* excludable, i.e., it affects individual outcomes only through observables. Given such an instrument, two-stage least squares (2SLS) recovers the individual-level effect of x on y. More formally:

PROPOSITION 1: *Applied to (1), the two-stage least squares estimator of β is asymptotically consistent and equal to the true, individual-level parameter whenever the instrument, z_c, satisfies (i)* $\text{Cov}\left(z_c^*, \overline{x}_c\right) \neq 0$ *and (ii)* $\text{Cov}\left(z_c^*, \varepsilon_i\right) = 0$*, where ε_i is the individual-level error term and z_c^* denotes the residual from projecting z_c on the space spanned by the controls.*

PROOF: The probability limit of the 2SLS estimator equals plim $\widehat{\beta}_{2SLS} = \text{Cov}\left(z_c^*, \overline{y}_c\right) / \text{Cov}\left(z_c^*, \overline{x}_c, \right) = \beta + \text{Cov}\left(z_c^*, \overline{\varepsilon}_c\right) / \text{Cov}\left(z_c^*, \overline{x}_c, \right)$. It thus suffices to show that $\text{Cov}\left(z_c^*, \overline{\varepsilon}_c\right) = 0$. By the usual properties of covariances, for any c, $\text{Cov}\left(z_c^*, \overline{\varepsilon}_c\right) = \frac{1}{N_c} \sum_{i=1}^{N_c} \text{Cov}\left(z_c^*, \varepsilon_i\right)$. Given that individual indices are exchangeable within each c, we immediately see that $\text{Cov}\left(z_c^*, \varepsilon_i\right) = 0$ implies $\text{Cov}\left(z_c^*, \overline{\varepsilon}_c\right) = 0$, as desired. *Q.E.D.*

Loosely speaking, the proposition states that if the researcher observes a variable that would satisfy the exclusion restriction in individual-level data, then the same variable is also a valid instrument in the aggregated, i.e., averaged, data. The mathematics behind this result is strikingly simple. If an instrument is uncorrelated with all individual error terms, then it must also be uncorrelated with their mean. As a result, the 2SLS estimator simultaneously solves both the causal as well as the ecological inference problem.

Naturally, the validity of these inferences is contingent on the excludability of z_c. The assumption that the instrument and the error term are uncorrelated always requires careful justification.

4. An Illustrative Application with Sensitivity Analysis

To illustrate how EI-IV may be useful in applications, we draw on recent work by Spenkuch and Tillmann (2018; henceforth ST).[2] A large Nazi voting literature identified religion as a key correlate of NSDAP support. Whether the observed correlation is causal, however, remained unknown. In fact, Falter (1991) cautions that the assumptions required for his seminal estimates to have a causal interpretation are "in many cases unrealistic" (p. 443).

ST draw causal ecological inferences by using EI-IV in conjunction with the following econometric specification: $\overline{v}_c = \mu_d + \beta\overline{Catholic}_c + \overline{\mathbf{W}}_c'\theta + \xi_c$, where, $\overline{v}_c$ denotes NSDAP vote shares in county c, $\overline{Catholic}_c$ measures the share of Catholics, $\overline{\mathbf{W}}_c$ is a vector of county-level controls, and μ_d marks an electoral district fixed effect. To ensure that their ecological regressions correspond to the aggregation of an individual-level model, $\overline{v}_c$ is measured with respect to all adults in a particular county rather than the total number of votes.

Due to the scarcity of data from the end of the Weimar Republic, there might be many unobservables that are correlated with both NSDAP votes and constituents' religious composition. If correct, then $\mathrm{Cov}\left(\overline{Catholic}_c^*, \xi_c\right) \neq 0$, and the simple OLS estimates in Table 1 cannot be given a causal interpretation.

In their search for an instrument, ST turn to a stipulation in a sixteenth-century peace treaty. According to the principle *cuius regio, eius religio* ("whose realm, his religion"), the Peace of Augsburg granted more than a thousand local lords the right to determine their territories' official faith and, therefore, the religion of all their subjects. As the evidence in

[2] While ST have already used instrumental variables techhniques to draw causal ecological inferences, they explicily reference this note as the source of the underlying theoretical insight.

the middle two columns of Table 1 demonstrates, a territory's official faith at the end of the sixteenth century correlates strongly with the religion of Germans living in the same area more than 300 years later. Thus, princes' choices in the aftermath of the Peace satisfy the criterion of instrument relevance, i.e., $\mathrm{Cov}\left(z_c^*, \overline{Catholic}_c\right) \neq 0$. Importantly, the historical record indicates that rulers' decisions were often based on their personal convictions and other idiosyncratic factors, suggesting that, conditional on socioeconomic controls, excludability might also hold, i.e., $\mathrm{Cov}\left(z_c^*, \xi_c\right) = 0$. If one believes that, conditional on the set controls, princes' decisions affected Nazi vote shares in 1932 only through the constituents' religion, then the 2SLS estimates in the rightmost columns of Table 1 can be interpreted as causal *individual-level* effects. According to these estimates, Catholics were about 25 percentage points less likely to support the Nazis than Protestants.

Depending on the application, scholars may or may not have much faith in the exclusion restriction. Fortunately, it is easy to assess the sensitivity of the estimated individual-level effect with respect to potential violations of this key assumption.

Building on Conley et al. (2012), ST consider the augmented model: $\overline{v}_c = \mu_d + \beta\overline{Catholic}_c + \overline{\mathbf{W}}_c'\theta + \gamma_0 Historically\, Catholic_c + \gamma_1 Historically\, Mixed_c + \xi_c$. Here, $\gamma = [\gamma_0, \gamma_1]$ parameterizes the extent to which excludability fails. If the exclusion restriction does, in fact, hold, then $\gamma_0 = \gamma_1 = 0$ (i.e., princes' decisions had no independent impact on Nazi votes). Since $\overline{Catholic}_c$ is endogenous, β and γ cannot be separately identified. Conley et al. (2012), however, show how to conduct Bayesian inference conditional on specifying the *support* or the *distribution* of γ.[3]

Without prior information on the independent impact of the instrument, one obtains identical point estimates as in the EI-IV setup. The confidence intervals, however, widen. Figure 1 displays confidence intervals imposing only the assumption that each element of γ falls within $[-\delta, \delta]$—so that δ denotes the maximal theoretical violation of the exclusion restriction.

[3] A STATA routine that implements the necessary calculations can be downloaded by typing `ssc install plausexog` at the STATA prompt.

Reassuringly, as long as one is willing to rule out a direct impact of rulers' choices on Nazi vote shares greater or equal to about nine percentage points, one would still reject the null hypothesis of no individual-level effect of religion on Nazi support. Since nine percentage points corresponds to more than one-third of all NSDAP supporters in the November elections of 1932, ST conclude that religion almost certainly exerted a genuine effect. Furthermore, appealing to the results in this note, they extrapolate to the individual level.

5. Heterogeneous Individual-Level Effects

The discussion above assumes constant causal individual-level effects. While this is the modal assumption in applied work, reality tends to be more complicated. Nonetheless, the Corollary in Section 2 continues to hold. Estimates from regression techniques that uncover causal effects are informative about individual-level parameters whenever the ecological regression model can be microfounded by averaging over the individual-level DGP. With heterogeneous effects the question becomes to which (sub)population do these causal estimates apply?

Even in microdata, 2SLS, RDDs, and related techniques only estimate local average treatment effects (LATE; Imbens and Angrist 1994). That is, researchers only learn about the causal effect for a subset of individuals. Appendix B shows that, provided the assumptions in Imbens and Angrist (1994) hold, (population-weighted) EI-IV recovers the same LATE as standard IV. Relative to Proposition 1, the key additional condition is *monotonicity*, i.e., that, for all individuals, the instrument exerts either weakly positive or weakly negative effects on the regressor of interest, but not a mix of both. Thus, with EI-IV researchers can learn about the same individual-level effect that they could recover from microdata (given a particular instrument). Naturally, the similarity in assumptions implies that causal ecological inferences come with similar limitations as other causal analyses.

To illustrate the limitations of EI-IV, ST estimate the effect of religion for Germans who were either Catholic or Protestant because the local prince chose a particular religion in the aftermath of the peace treaty. They argue that the LATE estimate is still interesting because princes' right to determine the religion of their people was initially strictly enforced—

defiers even faced the death penalty—and because religion used to be highly persistent across generations of the same family. Their analysis is not informative, however, about the effect of religion on individuals who are unaffected by lords' choices, i.e., "always-" and "never-takers." But as Imbens (2010) quips, "better LATE than nothing."

6. Concluding Remarks

The first part of this note clarifies that, due to a common mathematical structure, ecological and causal inferences from aggregate data go hand-in-hand. Whenever an ecological regression specification is directly implied by the individual-level DGP, the resulting estimates can be generalized to the behavior of individuals if and only if they are causal. Though simple, this observation implies that, provided questions of causality are carefully addressed, we can learn more from aggregate data than typically thought.

While there exist a number of research designs that, under plausible assumptions, establish causal aggregate relationships—think, for instance, of RDDs or difference-in-differences estimators—the second part of this note focuses on instrumental variable techniques. EI-IV allows researchers to extrapolate from aggregate data to individual-level effects under essentially the same set of conditions that are required for causal identification in microdata. EI-IV can easily be implemented with standard statistical software. It is also straightforward to assess the robustness of the individual-level inferences with respect to the (untestable) exclusion restriction. In addition, instrumental variables techniques are theoretically well-understood. As a consequence, researchers can reason about the sources of identifying variation, i.e., which group(s) of individuals identify the estimated coefficient, and whether this parameter is likely to be representative of the broader population. Given the emphasis on causal identification in modern research, EI-IV offers an attractive methodological approach to a wide range of questions in the social sciences and beyond.

References

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

Cross, P., and C. Manski (2002). "Regressions, Short and Long." *Econometrica*, 70(1): 357–368.

Duncan, O., and B. Davis (1953). "An Alternative to Ecological Correlation." *American Sociological Review*, 18(6): 665–666.

Falter, J. (1991). *Hitlers Wähler*. Munich: C.H. Beck.

Freedman, D., S. Klein, J. Sacks, C. Smyth, and C. Everett (1991). "Ecological Regression and Voting Rights." *Evaluation Review*, 15(6): 673–711.

Frisch, R., and F. Waugh (1933). "Partial Time Regressions as Compared with Individual Trends." *Econometrica*, 1(4): 387–401.

Gelman, A., D. Park, S. Ansolabehere, P. Rice, and L. Minnite (2001). "Models, Assumptions and Model Checking in Ecological Regressions." *Journal of the Royal Statistical Society: A*, 164: 101–118.

Goodman, L. (1953). "Ecological Regressions and Behavior of Individuals." *American Sociological Review*, 18(6): 663–664.

——— (1959) "Some Alternatives to Ecological Correlation." *American Journal of Sociology*, 64(6): 610–625.

Imbens, G., and J. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–476.

——— (2010). "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48(2): 399–423.

King, G. (1997). *A Solution to the Ecological Inference Problem*. Princeton, NJ: Princeton University Press.

———, O. Rosen, M. Tanner (1999). "Binomial-Beta Hierarchical Models for Ecological Inference." *Sociological Methods and Research*, 28(1): 61–90.

———, ———, ——— (eds.) (2004). *Ecological Inference: New Methodological Strategies*. Cambridge, UK: Cambridge University Press.

Langbein, L., and A. Lichtman (1978) *Ecological Inference*. Beverly Hills, CA: Sage.

Robinson, P. (1950). "Ecological Correlation and the Behavior of Individuals." *American Sociological Review*, 15(3): 351–357.

Spenkuch, J., and P. Tillmann (2018) "Elite Influence? Religion and the Electoral Success of the Nazis." *American Journal of Political Science*, 62(1): 19–36.

ONLINE APPENDIX

Appendix A: Aggregate vs. Individual-Level Instruments

To avoid potential confusion, it may be worth elaborating on a subtle point. Proposition 1 defines the instrument on the level at which the researcher actually observes the data. To see why this is important, note that averaging a variable that would be excludable if applied to individual-level data may but need not yield an instrument that is also valid in the aggregate. Specifically, let $\overline{z}_c \equiv \frac{1}{N_c}\sum_{i=1}^{N_c} z_i$. If z_i does not vary across individuals within the same c, then $\overline{z}_c = z_i$ and $\text{Cov}\left(z_i, \varepsilon_i\right) = 0$ implies that $\text{Cov}\left(\overline{z}_c, \overline{\varepsilon}_c\right) = 0$, as required for plim $\widehat{\beta}_{2SLS} = \beta$. Thus, whenever the instrument is logically constant within a unit of aggregation, EI-IV relies on *exactly* the same set of assumptions that researchers impose anyway in order to obtain causal estimates. Extrapolating from the results to individual behavior is thus without any further loss.

If, however, $z_i \neq z_j$ for some $i, j \in c$, then $\text{Cov}\left(z_i, \varepsilon_i\right) = 0$ does not automatically imply $\text{Cov}\left(\overline{z}_c, \overline{\varepsilon}_c\right) = 0$.[1] A sufficient assumption for $\overline{z}_c$ to be excludable is $\text{Cov}\left(z_i, \varepsilon_j\right) = 0$ for all i and j within the same unit of aggregation. Or, in words, the instrument must affect *all* individuals within the same unit only through covariates that are already included in the regression. In effect, scholars need to argue that, defined on the individual level, the instrument would be excludable in the usual sense *and* exhibit no unobserved spillover effects on others.[2] With such an instrument, EI-IV recovers causal individual-level estimates from aggregate data.

Appendix B: EI-IV Recovers LATE

To show that EI-IV recovers local-average treatment effects under the same set of assumptions as in Imbens and Angrist (1994), we adopt their framework, i.e., their potential outcomes notation and lack of covariates adjustment. Specifically, assume that $D_i \in \{0, 1\}$ is the variable in whose effect we are interested. Let $Y_i\left(0\right)$ be the outcome of individual i when $D_i = 0$, and $Y_i\left(1\right)$ the outcome when $D_i = 1$. Further let, $D_i\left(z\right)$ be the value of D_i if the random variable Z_i were equal to z. If the individual-level data were available, the researcher would observe D_i and $Y_i = D_i Y_i\left(1\right) + \left(1 - D_i\right) Y_i\left(0\right)$, which does not suffice to directly infer the causal effect of D on Y for individual i, i.e., $Y_i\left(1\right) - Y_i\left(0\right)$.

Imbens and Angrist (1994) show that if there exists some an instrument Z_i such that:

1. *(independence)* the triple $\left(Y_i\left(1\right), Y_i\left(0\right), D_i\right)$ is jointly independent of Z_i,
2. *(relevance)* $E\left[D_i|Z = w\right] \neq E\left[D_i|Z = w'\right]$ for some $w \neq w'$, and
3. *(monotonicity)* either $E\left[D_i|Z = w\right] \geq E\left[D_i|Z = w'\right]$ or $E\left[D_i|Z = w\right] \leq E\left[D_i|Z = w'\right]$ for all $w \neq w'$,

[1] Again, this is due to the possibility that $\text{Cov}\left(z_i, \varepsilon_j\right) = 0$ for some $i \neq j$ within the same c.

[2] That is, $\text{Cov}\left(z_i, \varepsilon_i\right) = 0$ for all i, and $\text{Cov}\left(z_i, \varepsilon_j\right) = 0$ for all $i \neq j$ within the same unit of aggregation.

then IV identifies the following local average treatment effect $\tau_{w,w'} = E\left[Y_i(1) - Y_i(0) \,|D_i(w) \neq D_i(w')\right]$. If Z_i is a binary instrument, then IV identifies

$$\begin{aligned}
\tau_{LATE} &= \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, D_i)} \\
&= \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \\
&= E\left[Y_i(1) - Y_i(0) \,|D_i(1) > D_i(0)\right].
\end{aligned}$$

In words, IV identifies the average treatment effect for individuals who are affected by the instrument, i.e., the LATE.

Our goal is to show that, conditional on the same instrument being used, EI-IV recovers the same effect. To this end, let Z_c be an instrumental variable defined at the correct level of aggregation, and assume that Z_c satisfies conditions (1)–(3). For simplicity, also assume that all units of aggregation include the same number of individuals. The researcher observes Z_c, $\overline{D}_c \equiv \frac{1}{N}\sum_{i=1}^{N} D_i$ and $\overline{Y}_c \equiv \frac{1}{N}\sum_{i=1}^{N} Y_i$. Mechanically, the probability limit of the two-stage least squares estimator is given by

$$\tau_{EI-IV} = \frac{\text{Cov}\left(Z_c, \overline{Y}_c\right)}{\text{Cov}\left(Z_c, \overline{D}_c\right)} = \frac{\text{Cov}\left(Z_c, \frac{1}{N}\sum_{i=1}^{N} Y_i\right)}{\text{Cov}\left(Z_c, \frac{1}{N}\sum_{i=1}^{N_c} D_i\right)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\text{Cov}(Z_c, Y_i)}{\frac{1}{N}\sum_{i=1}^{N_c}\text{Cov}(Z_c, D_i)} \tag{2}$$

$$= \frac{\text{Cov}(Z_c, Y_i)}{\text{Cov}(Z_c, D_i)} \tag{3}$$

The expression in (3) is equal to the usual probability limit of the individual-level two-stage least squares estimator (given instrument Z_c). And from Imbens and Angrist (1994) we know that if Z_c satisifes assumpions (1)–(3), then we recover τ_{LATE}. Hence, for the same instrument, standard IV and EI-IV identify the same local average treatment effect.[3]

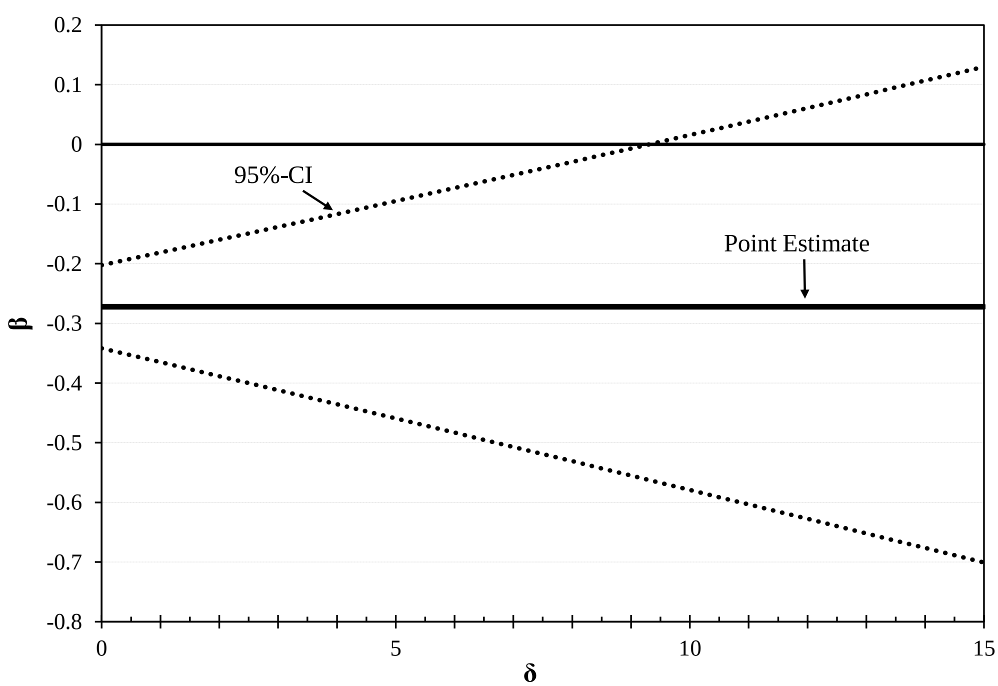
[3]If the units of aggregation do not contain the same number of individuals, then it becomes necessary to weight by population in order to estimate τ_{LATE}.

Table 1: Estimating the Effect of Religion on Nazi Support in the November Election of 1932

	A. OLS		*B. First Stage*		*C. 2SLS*	
	NSDAP Vote Share		Percent Catholic		NSDAP Vote Share	
	(1)	(2)	(3)	(4)	(5)	(6)
Percent Catholic	-.190	-.287			-.254	-.273
	(.019)	(.025)			(.018)	(.028)
Official Hist. Religion:						
Catholic			66.666	42.117		
			(3.232)	(3.681)		
Mixed			39.270	22.005		
			(4.320)	(3.322)		
Controls	No	Yes	No	Yes	No	Yes
Electoral District FE	No	Yes	No	Yes	No	Yes
First Stage F-Statistic	--	--	--	--	212.74	71.38
Over-ID Test [p-value]	--	--	--	--	.275	.581
Number of Observations	982	982	982	982	982	982

Notes: Entries in columns (1), (2), (5), and (6) are OLS and 2SLS point estimates from regression models akin to that in Section 4. Columns (3) and (4) show the first stage results, i.e., the partial correlation between instruments and the endogenous variables. Standard errors account for clustering by electoral district and are reported in parentheses. The set of controls is the same as in the most inclusive specification in Tables 1 and 2 of Spenkuch and Tillmann (2018). For a precise description of the instrument as well as all other variables, see the aforementioned paper.
Source: Spenkuch and Tillmann (2018), Tables 1 and 2

Figure 1: Inference Allowing for Violations of the Exclusion Restriction



Notes: Figure depicts point estimates and 95%-confidence intervals for the effect of Catholicism on NSDAP vote shares in the November elections of 1932. The confidence intervals impose only the prior information that the support of each element of γ falls within $[-\delta,\delta]$. Intuitively, δ parameterizes the maximal allowable violation of the exclusion restriction. For details on the estimation procedure, see Conley et al. (2012).
Source: Spenkuch and Tillmann (2018), Appendix Figure A.4