Online Supplementary Appendices

Negative Control Falsification Tests for Instrumental Variable Designs

Oren Danieli (orendanieli@tauex.tau.ac.il), Daniel Nevo, Itai Walk, Bar Weinstein, Dan Zeltzer

October 12, 2025

Appendix A. Details on Survey of Common Practices

Sample Construction. We used Google Scholar in November 2023 to assemble the list of relevant papers. We searched the terms “instrumental variable,” “instrument,” “2SLS,” and “IV.” We restricted the sample to articles with over 300 citations or, if published after 2020, over 100 citations. We examined all articles satisfying these criteria, published in five top-ranked economics journals: Review of Economic Studies, American Economic Review, Journal of Political Economy, Quarterly Journal of Economics, and Econometrica. Overall, our survey includes 140 papers.

We then searched the papers for strings related to falsification testing. This included “falsification,” “negative control,” “balance,” “balancing,” “valid,” and “validity.” Papers that did not include any of these strings were marked as not having any falsification test. We manually coded the type of falsification test for papers that included one of these strings. The results are summarized in Table 2 and discussed in Section I.

Other Falsification Tests. As discussed in Section I, we categorized all falsification tests used in surveyed papers into NCO tests, NCI tests, and other falsification tests. Other falsification tests include the following: negative control tests in non-IV settings, which examine only the first or second stage in a 2SLS estimation; “placebo population” analyses (Eggers, Tuñón and Dafoe, 2023; Glymour, Tchetgen Tchetgen and Robins, 2012; Keele et al., 2019), which involve repeating the analysis using a different population where the IV is not expected to affect the outcome; validating that the results are robust to including additional control variables; and using an over-identification test when more than one IV is available.

Appendix B. Formalization of IV Negative Control Tests Using DAGs

In this section, we present an alternative formalization of negative controls for IV designs using the language of DAGs. We first summarize fundamental concepts from DAG theory.[28] The DAGs we use intuitively throughout the paper can be formalized using the theory reviewed in this section.

[28]This is not an exhaustive overview of DAG use for causality, but only the elements necessary for our purposes. See Pearl (2009*b*) for detailed presentation and theory.

1. *Background DAG Theory*

A directed graph is a set of nodes and directed edges. A *directed path* on a graph between two nodes, X_1 and X_2, is a sequence of edges, such that the first edge starts at X_1, each edge starts at the arrowhead of the former edge, and the last edge ends at X_2. A directed *acyclic* graph (DAG) contains no cycles; namely, there are no directed paths from a node to itself. While DAGs can be used to represent assumptions on joint probability functions without notions of causality, here we interpret all DAGs as causal, such that the arrows represent a causal relationship. Consider a DAG denoted by G. The set of *parents* of node X_j, denoted by PA_j, is the set of all nodes with direct arrows into X_j. The *descendants* of X_j are all nodes with a directed path of any number of edges (including a single edge) from X_j to those variables; these are variables causally affected by X_j, directly or indirectly.

DAGs can be used to encode conditional independence assumptions on the joint distribution of variables represented by nodes in the DAG. Each DAG represents an infinite number of probability functions sharing the same conditional independence structure. The joint distribution of all variables in the DAG is the product of the conditional probability function of each variable X_j given its parents PA_j.[29] Formally, for M variables $X_1, ...X_M$, this factorization is

$$P(x_1, ..., x_M) = \prod_{j=1}^{M} P(x_j|PA_j),$$

where lower cases are realizations. Any probability function P that admits to the above factorization is said to be compatible with a DAG G.

A key result of DAG theory is the translation of the structure represented by a DAG into conditional independence conditions. This translation relies on the concept of d-separation, a graphical condition.

Definition A1 (d-separation (Pearl, 2009*b*))**.** *A path p from node X_1 to node X_2 in graph G is said to be blocked by a set of nodes $\mathcal{A}$ if and only if*

1) *The path p contains a chain $X_1 \to A \to X_2$ or a fork $X_1 \leftarrow A \to X_2$ such that the middle node A is in $\mathcal{A}$, or*
2) *The path p contains a collider structure $X_1 \to B \leftarrow X_2$ such that the middle node, the collider B, is not in $\mathcal{A}$ and such that no descendant of B is in $\mathcal{A}$.*

The set $\mathcal{A}$ is said to d-separate node X_1 from node X_2 if and only if $\mathcal{A}$ blocks every path between X_1 and X_2. In this case, we write

$$(X_1 \perp\!\!\!\perp X_2|\mathcal{A})_G.$$

[29] This assumption is called the Markov property.

Building on the above structure, the following theorem states the direct implication of d-separation, a graphical condition, on conditional independence, a probabilistic statement.

Theorem A1 (Pearl 2009*b*). *Let X_1, X_2 be two variables that are d-separated by a set of variables $\mathcal{A}$ in graph G, $(X_1 \perp\!\!\!\perp X_2|\mathcal{A})_G$. Then,*

$$X_1 \perp\!\!\!\perp X_2|\mathcal{A}$$

in all probability functions compatible with G.[30]

IVs can also be defined using a graphical criterion (Pearl, 2009*b*, Definition 7.4.1). Consider a DAG G with nodes Z, X, Y (and possibly additional nodes, as in the DAGs in Figure 1). We follow Pearl (2009*b*) by using $G_{\overline{X}}$ to denote the version of the DAG G with all arrows entering X removed.

Definition A2. *A variable Z is an IV for treatment X on outcome Y if*

1) $(Z \perp\!\!\!\perp Y)_{G_{\overline{X}}}$
2) $(Z \not\!\perp\!\!\!\perp X)_G$.

The first condition corresponds to the condition stated in (1). It implies that there is no alternative path between the IV and the outcome other than through the treatment. The second condition corresponds to the IV relevance assumption. The negative control falsification tests for IVs we discuss in this paper focus on the first condition.

2. *Negative Controls for IVs Using DAGs*

We are now ready to present negative controls for IVs using DAGs. To this end, we begin by providing the definitions of APO and API variables and those of NCOs and NCIs. Then, we provide proofs of Theorems 1 and 2 using DAG theory.

We start with the graphical DAG-based definition of an APO variable.

Definition A3 (Alternative path outcome variable). *A random variable U is an APO variable if the following two conditions hold.*

1) *Latent IV validity.* $(Z \perp\!\!\!\perp Y|U)_{G_{\overline{X}}}$.
2) *Path indication. If* $(Z \perp\!\!\!\perp Y)_{G_{\overline{X}}}$ *then* $(Z \perp\!\!\!\perp U)_G$.

This definition is similar to the definition of APO variables using potential outcomes (Definition 1). Latent IV validity implies that Z and Y are d-separated by U (excluding the path through X). Path indication implies that if the IV is valid, then there is no unblocked path between Z and U.

Similarly, we can define graphically an API variable.

[30]Often $X_1 \perp\!\!\!\perp X_2|\mathcal{A}$ is written as $(X_1 \perp\!\!\!\perp X_2|\mathcal{A})_P$. We omit the P subscript for simplicity.

Definition A4 (Alternative path instrument variable). *A random variable U is an API variable if the following two conditions hold.*

1) *Latent IV validity.* $(Z \perp\!\!\!\perp Y|U)_{G_{\overline{X}}}$.

2) *Path indication. If* $(Z \perp\!\!\!\perp Y)_{G_{\overline{X}}}$ *then* $(U \perp\!\!\!\perp Y|Z)_G$.

Turning to negative control variables, the definitions of NCO and NCI are similar to the definitions using potential outcomes (Definitions 3 and 4).

Definition A5. *A random variable NC is an NCO if there exists an APO variable U such that the following two conditions hold.*

1) *The NCO assumption.* $(NC \perp\!\!\!\perp Z|\ U)_G$.

2) *U-comparability.* $(NC \not\perp\!\!\!\perp U)_G$.

Definition A6. *A random variable NC is an NCI if there exists an API variable U such that the following two conditions hold.*

1) *The NCI assumption.* $(NC \perp\!\!\!\perp Y|\ Z, U)_G$.

2) *U-comparability.* $(NC \not\perp\!\!\!\perp U|Z)_G$.

We now provide a proof of Theorem 1 under the above DAG definitions. Following the condition of Theorem 1, the NCO test finds that $NC \not\perp\!\!\!\perp Z$. By Definition A5 we have that $(NC \perp\!\!\!\perp Z|\ U)_G$. From the test we know that $NC \not\perp\!\!\!\perp Z$, which implies $(NC \not\perp\!\!\!\perp Z)_G$ by the contrapositive of Theorem A1. Because $(NC \not\perp\!\!\!\perp Z)_G$, there is at least one open (unblocked) path between Z and NC. However, because $(NC \perp\!\!\!\perp Z|U)_G$, this path or paths are blocked by U. By Definition A1, this means that U is either in the middle of a chain or a fork on the open path between Z and NC. Thus, there is an unblocked path between Z and U, i.e., $(Z \not\perp\!\!\!\perp U)_G$ which, by path indication, implies that $(Z \not\perp\!\!\!\perp Y)_{G_{\overline{X}}}$, violating the first IV condition in Definition A2.

We turn to the proof of Theorem 2 under the DAG definitions. Following the condition of Theorem 2, the NCI test finds that $NC \not\perp\!\!\!\perp Y|Z$. By Definition A6 we have that $(NC \perp\!\!\!\perp Y|\ Z, U)_G$. From the test, we know that $NC \not\perp\!\!\!\perp Y|Z$, which implies $(NC \not\perp\!\!\!\perp Y|Z)_G$ by the contrapositive of Theorem A1. Because $(NC \not\perp\!\!\!\perp Y|Z)_G$, there is at least one open path between NC and Y, which is not blocked by Z. However, because $(NC \perp\!\!\!\perp Y|U, Z)_G$, this path or paths are blocked by U. By Definition A1, this means that U is either in the middle of a chain or a fork on the open path between Y and NC. Thus, there is an unblocked path between U and Y not blocked by Z, i.e., $(U \not\perp\!\!\!\perp Y|Z)_G$ which by path indication implies that $(Z \not\perp\!\!\!\perp Y)_{G_{\overline{X}}}$, violating the first IV condition in Definition A2.

Appendix C. Additional Results and Proofs

Throughout, we let $P(\cdot|\cdot)$ be the conditional probability or density function. As a shorthand, we leave the random variables to be understood from the arguments of P. For example, if $Y(x)$ is discrete, $P[y(x)|u]$ is a shorthand for $\Pr[Y(x) = y(x)|U = u]$.

Auxiliary Lemmas

Lemma 1 (Lemma 4.3 in Dawid (1979))**.** *Let A, B, D, Q be four random variables. If $A \perp\!\!\!\perp B|D, Q$ and $B \perp\!\!\!\perp Q|D$ then $A \perp\!\!\!\perp B|D$.*

Proof. Because $A \perp\!\!\!\perp B|D, Q$, it follows that for all a, b, d, q, we have that

$$\begin{aligned} P(a, b|d, q) &= P(a|d, q)P(b|d, q) \\ &= P(a|d, q)P(b|d), \end{aligned} \tag{A1}$$

where the last line follows from $B \perp\!\!\!\perp Q|D$. Now,

$$P(a, b|d) = \int P(a, b|d, q)P(q|d)dq = \left[\int P(a|d, q)P(q|d)dq\right] P(b|d) = P(a|d)P(b|d),$$

where the second equality is by (A1). □

This lemma is also known as the contraction axiom of conditional independence (Pearl and Paz, 1986). The following lemma is a direct result of Lemma 1.

Lemma 2. *Let A, B, D, Q be four random variables. If $A \perp\!\!\!\perp B|D, Q$ and $A \not\perp\!\!\!\perp B|D$ then $A \not\perp\!\!\!\perp Q|D$ and $B \not\perp\!\!\!\perp Q|D$.*

Proof. Assume by way of contradiction that $B \perp\!\!\!\perp Q|D$. Therefore, by Lemma 1, because $A \perp\!\!\!\perp B|D, Q$ it follows that $A \perp\!\!\!\perp B|D$, which contradicts the assumption. A similar contradiction is received by assuming $A \perp\!\!\!\perp Q|D$. □

1. Negative Controls for IV Designs Under General Definitions

This section presents the proofs of the results from Section II. We prove versions of the results that are more general in three different ways. First, we discuss IV designs that include control variables. Second, we provide more general definitions for APO and API variables that accommodate multiple threats of which Definitions 1 and 2 are special cases. Third, we provide more general definitions of NCO and NCI (under weaker NCO and NCI assumptions, respectively).

We start by presenting the outcome independence and exclusion restriction assumptions when controls are included.

Assumption A1 (Outcome independence). *$Z \perp\!\!\!\perp Y(z, x)|C$ for all possible z, x values.*

Assumption A2 (Exclusion restriction). $\Pr(Y(z, x) = Y(z', x) = Y(x)|C = c) = 1$ *for all possible z, z', x, c values.*

Similar to the case without controls, outcome independence and exclusion restriction together yield $Z \perp\!\!\!\perp Y(x)|\ C$.

Alternative Path Variables with Multiple Threats and Controls

In some applications, multiple potential alternative paths can exist between the IV and the outcome. Appendix D.6 presents an example of two distinct variables that affect the outcome and could potentially affect the IV as well and thus may violate outcome independence. To accommodate the possibility of multiple violations of the IV assumptions, we extend Definitions 1 and 2. We introduce a random variable V, which represents other potential threats in addition to the threat posed by the alternative path variable U. We also include control variables C to accommodate cases where the IV design is assumed to be valid only when controls are included.

The more general definition for APO variables reads as follows.

Definition A7 (Alternative path outcome variable with multiple threats and controls). *A random variable U is an APO variable conditional on a set of controls C if there exists a random variable V such that the following conditions hold.*

1) *Latent IV validity. $Z \perp\!\!\!\perp Y(x)|C, U, V$.*
2) *Path indication. If $Z \perp\!\!\!\perp Y(x)|C, V$ then $Z \perp\!\!\!\perp U|C, V$.*
3) *Direct IV link. If $Z \perp\!\!\!\perp U|C, V$ then $Z \perp\!\!\!\perp U|C$.*
4) *V-validity. If $Z \perp\!\!\!\perp Y(x)|C$ then $Z \perp\!\!\!\perp Y(x)|C, V$.*

Under this definition, latent IV validity states that the IV is valid conditionally not only on the APO variable U, but also on the additional threat(s) V, and the controls C.

In contrast to Definition 1, this more general version of latent IV validity also holds for U even if V is the actual threat to the identification and U is only an imperfect proxy for it. Therefore, to maintain the same interpretation of an APO variable as posing a threat to IV validity, we replace the condition of path indication from Definition 1 and include two additional conditions. Together, Conditions 2–4 of Definition A7 yield Condition 2 of Definition 1 (conditional on the controls C). However, the three separate conditions ensure that the APO variable is part of the threat itself and not a proxy. Specifically, path indication and direct IV link each rule out a different type of proxy; see Appendix D.7 and Appendix D.8 for counterexamples. The final property, V-validity, is a more technical requirement for the variable V that ensures V represents other threats. It states that an IV design that satisfies outcome independence and exclusion restriction conditional on the controls remains valid conditional on V. See Appendix D.9 for a counterexample. If there are no additional threats other than through U, and no control variables, Definition A7 is equivalent to Definition 1.

We similarly extend Definition 2 to allow for additional threats and to include controls C.

Definition A8 (Alternative path instrument variable with multiple threats and controls). *A random variable U is an API variable conditional on a set of controls C if there exists a random variable V such that the following conditions hold.*

1) *Latent IV validity.* $Z \perp\!\!\!\perp Y(x)|C, U, V$.
2) *Path indication. If* $Z \perp\!\!\!\perp Y(x)|C, V$ *then* $U \perp\!\!\!\perp Y|Z, C, V$.
3) *Direct outcome link. If* $U \not\perp\!\!\!\perp Y|Z, C, V$ *then* $U \not\perp\!\!\!\perp Y|Z, C$.
4) *V-validity. If* $Z \perp\!\!\!\perp Y(x)|C$ *then* $Z \perp\!\!\!\perp Y(x)|C, V$.

Condition 1 is the same as in Definition A7. Conditions 2–4 together imply a version of Condition 2 from Definition 2 that includes controls. Similar to APO variables, the condition is decomposed into three independent conditions to exclude proxies that are not themselves part of an alternative path and maintain that V is indeed a part of such a threat (similar to Definition A7).

General Definition of Negative Control Variables

We adapt the definition of NCOs as follows.

Definition A9 (Negative control outcome with controls). *A random variable NC is an NCO if there exists an APO variable U such that the following two conditions hold.*

1) *The NCO assumption. If* $Z \perp\!\!\!\perp U|C$ *then* $NC \perp\!\!\!\perp Z|\ U, C$.
2) *U-comparability.* $NC \not\perp\!\!\!\perp U|C$.

Even without controls ($C = \emptyset$), this definition is more general than Definition 3, as it allows for NCOs that were previously excluded. To see this, note that in the case without controls, if $Z \not\perp\!\!\!\perp U$ (and so the design is invalid), the NCO assumption in Definition A9 allows for an association $NC \not\perp\!\!\!\perp Z|\ U$. In this case, the NCO would still be informative about the validity of the IV design since the association between the NCO and the IV, given the APO variable exists only if the design is invalid. Appendix D.10 provides an example of such an NCO. Every variable that satisfies the NCO assumption in Definition 3 trivially satisfies this less restrictive definition.

Next, we generalize the definition of an NCI to include controls and allow for direct associations with the outcome if the IV design is invalid.

Definition A10 (Negative control instrument with controls)**.** *A random variable NC is an NCI if there exists an API variable U such that the following two conditions hold.*

1) *The NC assumption. If $U \perp\!\!\!\perp Y|Z, C$ then $NC \perp\!\!\!\perp Y|Z, C, U$.*
2) *U-comparability. $NC \not\perp\!\!\!\perp U|Z, C$.*

Negative Control Tests with Controls

We are now ready to state the more general version of Theorem 1 and present its proof. This theorem also covers the case without controls by letting C be degenerate.

Theorem A2. *Assume that a random variable NC is an NCO with respect to controls C (Definition A9). If $NC \not\perp\!\!\!\perp Z|C$, then either outcome independence or exclusion restriction is violated. That is, the IV design is invalid.*

Proof. We begin by showing that $NC \not\perp\!\!\!\perp Z|C$ implies that $Z \not\perp\!\!\!\perp U|C$. Else, if $Z \perp\!\!\!\perp U|C$ then by the NCO assumption (see Definition A9) $NC \perp\!\!\!\perp Z|\ U, C$. Based on Lemma 2, $NC \perp\!\!\!\perp Z|\ U, C$ and $NC \not\perp\!\!\!\perp Z|C$ imply that $Z \not\perp\!\!\!\perp U|C$, a contradiction.

Next, from direct IV link if follows that $Z \not\perp\!\!\!\perp U|C$ implies $Z \not\perp\!\!\!\perp U|V, C$. Then, by path indication, we get that $Z \not\perp\!\!\!\perp Y(x)|V, C$. Finally, by V-validity, we have that $Z \not\perp\!\!\!\perp Y(x)|C$.

However, outcome independence (Assumption A1) and exclusion restriction (Assumption A2) together imply that $Z \not\perp\!\!\!\perp Y(x)|C$. Therefore, one of these assumptions must be violated. □

Next, we state the more general version of Theorem 2 and present its proof. This theorem also covers the case without controls by letting C be degenerate.

Theorem A3. *Assume that a random variable NC is an NCI with respect to controls C (Definition A10). If $NC \not\perp\!\!\!\perp Y|Z, C$, then either outcome independence or exclusion restriction is violated. That is, the IV design is invalid.*

Proof. We divide the proof into two cases with respect to the API variable U for which the NCI assumption holds for NC.

First, assume that $U \not\perp\!\!\!\perp Y|Z, C$. In this case, from direct outcome link it follows that $U \not\perp\!\!\!\perp Y|Z, C, V$, and therefore by path indication, $Z \not\perp\!\!\!\perp Y(x)|C, V$. Therefore, by V-validity, we have that $Z \not\perp\!\!\!\perp Y(x)|C$. Hence, either outcome independence or the exclusion restriction do not hold.

We now turn to the other case, where $U \perp\!\!\!\perp Y|Z, C$. By the NCI assumption (Definition A10), we have that $NC \perp\!\!\!\perp Y|Z, C, U$, and by the condition of the theorem, we have that $NC \not\perp\!\!\!\perp Y|Z, C$. Therefore, by Lemma 2 (with $D = \{Z, C\}, Q = U$), we have that $U \not\perp\!\!\!\perp Y|Z, C$, which contradicts the assumption $U \perp\!\!\!\perp Y|Z, C$.

□

We now turn to state and prove a version of Theorem 3, conditional on controls C.

Theorem A4. *Assume that a random variable NC is an NCI with respect to controls C (Definition A10). If, in addition, $NC \perp\!\!\!\perp Z|C$, then if $NC \not\perp\!\!\!\perp Y|C$, then $Z \not\perp\!\!\!\perp Y(x)|C$.*

Proof. Assume by way of contradiction that $Z \perp\!\!\!\perp Y(x)|C$ holds. Because NC is an NCI, it follows from Theorem A3 that $NC \perp\!\!\!\perp Y|Z, C$. Additionally, based on the assumption that $NC \perp\!\!\!\perp Z|C$, Lemma 1 implies that $NC \perp\!\!\!\perp Y|C$, which contradicts the premise. □

Control Variables and Functional Form

Proof of Corollary 1

Proof. The minimized expression can be written as

$$\mathbb{E}[Z - b'_C C - b_{NC} NC]^2 = \mathbb{E}\big[Z - \mathbb{E}[Z|C, NC]\big]^2 + \mathbb{E}\big[\mathbb{E}[Z|C, NC] - b'_C C - b_{NC} NC\big]^2$$

plus a term equaling zero because $\mathbb{E}[Z - \mathbb{E}[Z|C, NC]]$ is zero by the law of total expectation. Since $\mathbb{E}[Z - \mathbb{E}[Z|C, NC]]^2$ does not depend on b_C, b_{NC}, we can write

$$\gamma = \arg\min_{b_c, b_{NC}} \mathbb{E}\big[\mathbb{E}[Z|C, NC] - b'_C C - b_{NC} NC\big]^2.$$

If outcome independence, exclusion restriction, and rich covariates are assumed, (4) holds and $\mathbb{E}[Z|C, NC] = \gamma'_C C$. Hence, we can further write

$$\gamma = \arg\min_{b_c, b_{NC}} \mathbb{E}[\gamma'_C C - b'_C C - b_{NC} NC]^2.$$

The values that minimize this nonnegative expression are $b'_C = \gamma'_C$ and $b_{NC} = 0$ and so the OLS population-level coefficient is $\gamma' = (\gamma'_C, 0)$. If $\gamma_{NC} \neq 0$, it must be that (4) does not hold. Therefore, either outcome independence, exclusion restriction, or rich covariates is violated.

□

Proof of Corollary 2

Proof. Let $\widetilde{Z} = Z - C'\mathbb{E}[CC']^{-1}\mathbb{E}[CZ]$ and $\widetilde{NC} = NC - C'\mathbb{E}[CC']^{-1}\mathbb{E}[CNC]$ be the residuals from the linear regressions of Z and NC on C, respectively. By the Frisch–Waugh–Lovell theorem, we can write β_Z as

$$\beta_Z = \frac{COV(\widetilde{Z}, \widetilde{NC})}{Var(\widetilde{Z})}.$$

If $\beta_Z \neq 0$ then it must be that $COV(\widetilde{Z}, \widetilde{NC}) \neq 0$. Define (γ'_C, γ_{NC}) to be the population-level solution of the reverse OLS, with Z as the dependent variable (as in Corollary 1). Again, by the Frisch–Waugh–Lovell theorem, we can write γ_{NC} as

$$\gamma_{NC} = \frac{COV(\widetilde{Z}, \widetilde{NC})}{Var(\widetilde{NC})}.$$

Since $COV(\widetilde{Z}, \widetilde{NC}) \neq 0$ it follows that $\gamma_{NC} \neq 0$ as well. By Corollary 1, we have that either outcome independence, exclusion restriction, or rich covariates does not hold.

□

Proof of Corollary 3

Proof. Similar to the proof of Corollary 1, we write the equivalent minimization problem as

$$\theta = \underset{b_Z, b_c, b_{NC}}{\arg\min}\, \mathbb{E}\big[\mathbb{E}[Y|Z, C, NC] - b_Z Z - b'_C C - b_{NC} NC\big]^2.$$

If outcome independence, exclusion restriction, and CSRF are assumed, (5) holds and

$$\theta = \underset{b_Z, b_c, b_{NC}}{\arg\min}\, \mathbb{E}[\theta_Z Z + \theta'_C C - b_Z Z - b'_C C - b_{NC} NC]^2.$$

The values that minimize this expression are $b_Z = \theta_Z, b'_C = \theta'_C$ and $b_{NC} = 0$ and so the OLS population-level coefficient is $\theta' = (\theta_Z, \theta'_C, 0)$. If $\theta_{NC} \neq 0$, it must be that (5) does not hold. Therefore either outcome independence, exclusion restriction, or CSRF is violated.

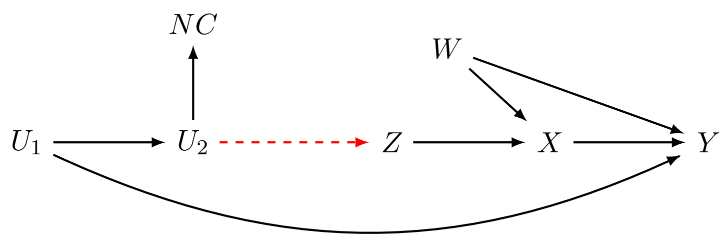
□

Appendix D. Examples and Counterexamples

1. *Non-Causal APO Variable*

In the example presented in Appendix Figure D1, U_2 is a valid APO variable satisfying path indication without having a direct causal effect on Y (beyond possible effect through the IV). Consider a scenario where Z represents supposedly quasi-random teacher assignment, X is the value added of the actual teacher, and Y is student test scores. The variable W represents some unobserved student characteristic (e.g., their parents' involvement), which affects test score and is also associated with teacher assignments, as some students switch classrooms after the original assignment.

In this example, U_1 denotes unobserved student ability that directly affects test scores, while U_2 represents detailed unobserved previous exam scores. The dashed arrow indicates the possibility



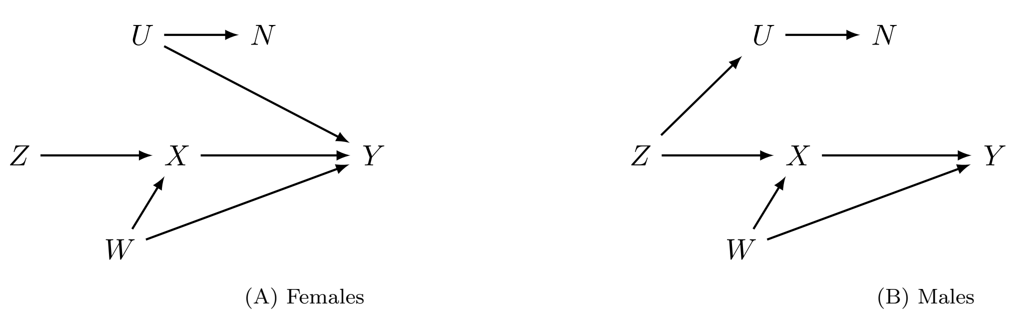
Appendix Figure D1. An illustration of causal and non-causal APO variables

that principals may assign students to teachers based on these detailed scores (e.g., students with low math scores are assigned to specific teachers). The detailed past exam scores satisfy path indication despite not directly affecting future test scores. A path exists between previous detailed scores (U_2) and current scores (Y) because both are affected by ability (U_1).

The variable NC represents aggregated previous test scores (e.g., average past scores in math together with other subjects). In this setting, NC is an NCO, with U_2 as an APO variable. An association between the IV and the aggregated lagged test scores would indicate an alternative path from the IV to the outcome. The presence of this path violates outcome independence, as students with different abilities would sort into different teachers based on previous math scores.

Note that while U_1 is also an APO variable, NC is a valid NCO only with respect to U_2, not with respect to U_1 alone. This is because, conditional on U_1, there is still an association between the NCO and the IV ($NC \not\perp Z | U_1$). That is, teacher assignment is not conditionally independent of aggregated test scores.

2. *Heterogeneity-Based Violation of Path Indication*

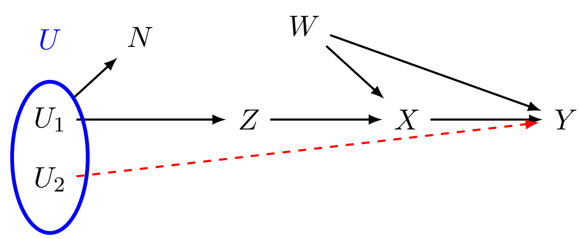


Appendix Figure D2. Violation of path indication (Definition 1) due to heterogeneity

Appendix Figure D2 describes a random variable U associated with both the IV and the potential outcome $Y(x)$. However, U does not qualify as an APO variable because it does not satisfy path indication. The IV affects U for males but not for females, while U affects Y for females but not for males. Path indication is not satisfied because even though the IV satisfied outcome independence and exclusion restriction assumptions ($Z \perp\!\!\!\perp Y(x)$), it remains associated with U ($Z \not\perp\!\!\!\perp U$).

The random variable N is a proxy for U. With sufficient sample size, we would detect $N \not\perp\!\!\!\perp Z$ (due to the effect among males), but this test result does not indicate that the IV is invalid since U is not an APO variable.

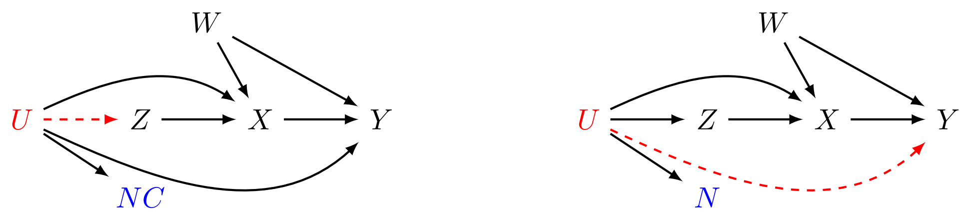
3. *Violation of Path Indication: Multivariate Variable*



Appendix Figure D3. Violation of path indication (Definition 1) when U has multiple components

Appendix Figure D3 presents an example where $U = (U_1, U_2)$ is a bivariate vector of independent variables. Assume Z represents teacher assignment claimed to be quasi-random, X is the actual teacher's value added, and Y is test scores. The variable W represents some unobserved student characteristic (e.g., their parents' involvement), which is also associated with teacher assignments, as some students switch classrooms after the original assignment. Let U_1 represent having basketball as a hobby and further assume that it is associated with the IV; for example, one teacher also coaches basketball, so students who list basketball as a hobby are more likely to be assigned to her. However, as seen in Appendix Figure D3, a basketball hobby is independent of test scores ($U_1 \perp\!\!\!\perp Y(x)$). Let U_2 represent having math as a hobby. Students reporting math as a hobby tend to perform better in exams and are randomly distributed between teachers ($U_2 \perp\!\!\!\perp Z$). The basketball and math hobbies are independent ($U_1 \perp\!\!\!\perp U_2$). Finally, assume that N is participation in an extracurricular basketball program, serving as a proxy for U (specifically for U_1).

Although the vector U is associated with both the IV and the outcome, it does not qualify as an APO variable because it does not satisfy path indication. The IV satisfies outcome independence and the exclusion restriction assumptions ($Z \perp\!\!\!\perp Y(x)$) despite being associated with the list of hobbies ($Z \not\perp\!\!\!\perp U$). Therefore, N is not a proper NCO. Even though $N \perp\!\!\!\perp Z|U$, it is still not an NCO because U is not an APO variable.



(A) $NC \not\!\perp\!\!\!\perp Z$ implies violation of IV validity. U is an APO variable. NC is an NCO.

(B) $N \not\!\perp\!\!\!\perp Y|Z$ does not necessarily implies violation of IV validity. U is not an API variable. N is not an NCI.

Appendix Figure D4. Direct effect of the potential alternative path variables on the treatment

4. Potential Alternative Path Variables and Association with the Treatment

Path indication for API variables (Definition 2) implies that conditional on the IV (Z), an API variable (U) cannot be associated with the treatment (X). By contrast, there is no such requirement for an APO variable. Appendix Figure D4 illustrates these points. In Panel A, U is a valid APO variable, and the arrow $U \to X$ is allowed: both latent IV validity and path indication hold. Therefore, NC is a valid NCO, and an association between NC and Z implies that the dashed arrow between U and Z exists and indicates that the IV design is invalid. Conversely, in Panel B, U is not a valid API variable. Path indication is violated because $U \not\!\perp\!\!\!\perp Y|Z$ does not imply $Z \perp\!\!\!\perp Y(x)$. Therefore, N is not an NCI, and $N \not\!\perp\!\!\!\perp Y|Z$ does not necessarily imply that the IV is invalid. Intuitively, $N \not\!\perp\!\!\!\perp Y|Z$, even if the IV design is valid, due to the association between U and Y through X. Conditioning on X would not solve this problem because X is a collider. Therefore, $N \not\!\perp\!\!\!\perp Y|Z, X$ due to the path $N \leftarrow U \to X \leftarrow W \to Y$ (Pearl, 2009*b*).

5. Counterexample: A Vector of NCOs That is Not an NCO

Let R_1, R_2 be two independent Bernoulli random variables with probabilities $\Pr(R_j = 1) = p_j$ and let $p_1 = p_2 = 0.5$. Let U be another Bernoulli random variable, independent of (R_1, R_2). Let Z be the IV, and assume that

$$Z = (R_1 \oplus R_2) + \theta U + \epsilon_Z,$$

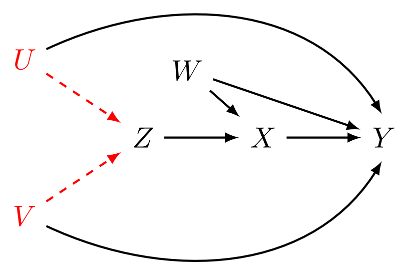
where $\oplus$ is the XOR operator. Assume that $Y(x) = x + U + \epsilon_Y$, such that U is an APO variable. The IV design is valid if $\theta = 0$.

Now, assume that there are two observed negative controls $NC_i = U \oplus R_i$ for $i = 1, 2$. Both NC_1 and NC_2 are valid negative controls as they satisfy the assumption $NC_i \perp\!\!\!\perp Z|U$. This is because for $i = 1, 2$, $R_i \perp\!\!\!\perp (R_1 \oplus R_2)$, and therefore $Z \perp\!\!\!\perp R_i|U$. However $(NC_1, NC_2) \not\!\perp\!\!\!\perp Z|U$ because, conditional on U, Z is associated with $NC_1 \oplus NC_2 = R_1 \oplus R_2$. Therefore, (NC_1, NC_2) does not satisfy the

NCO assumption. Indeed, even if the IV is valid, we could still have $Z \not\perp\!\!\!\perp (NC_1, NC_2)$.

A small change in the data-generating process will break some of the independencies discussed above. For example, changing the value of p_1 to something different from 0.5 would imply that $R_2 \not\perp\!\!\!\perp (R_1 \oplus R_2)$. In that case, $NC_2 \not\perp\!\!\!\perp Z$ and NC_2 would no longer satisfy the NCO assumption.

6. Multiple Threats



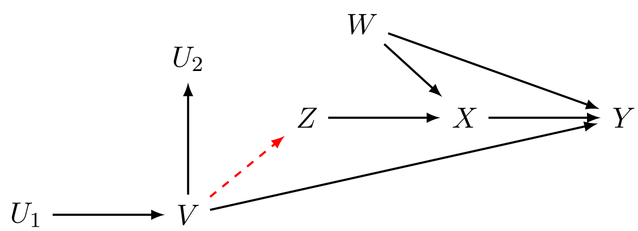
Appendix Figure D6. Multiple threats

Appendix Figure D6 presents an example of the presence of multiple threats to the validity of the IV design. In this case, the variable U is an APO variable by Definition A7 (taking the control variables C to be an empty set). In this figure, V is an APO variable as well.

For example, assume that Z is the teacher assignment, which is claimed to be quasi-random, X is the value added of the actual teacher, and Y is test scores. The variable W represents some unobserved student characteristic (e.g., their parents' involvement), which is also associated with teacher assignments, as some students switch classrooms after the original assignment. Assume that U is the student's unobserved ability. Assume also that V is principal quality, which is also unobserved. Both U and V might affect the teacher allocation Z, which would generate an alternative path between the IV and the outcome.

7. Direct IV Link Rules out Proxies of V

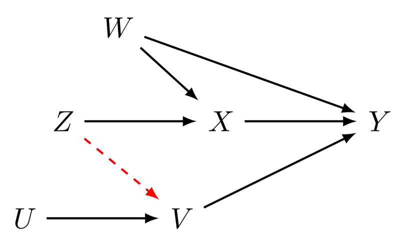
Appendix Figure D7 presents the potential violation of outcome independence through V, as well as two proxies for V, U_1, and U_2. Note that latent IV validity, as stated in Definition A7, holds for either variable (U_1 or U_2) together with V, as the further conditioning on U_1 or U_2 does not invalidate the IV, conditional on V. Note also that for both variables, path indication holds because if $Z \perp\!\!\!\perp Y(x)|V$ then $Z \perp\!\!\!\perp U_1|V$ (or $Z \perp\!\!\!\perp U_2|V$). Moreover, path indication for the single violation case in Definition 1, which is the result of combining conditions 2–4 in Definition A7 (in a design without control variables), also holds because if $Z \perp\!\!\!\perp Y(x)$ then $Z \perp\!\!\!\perp U_1$ (or $Z \perp\!\!\!\perp U_2$).



Appendix Figure D7. Violation of direct IV link (Definition A7)

However, condition 3 of Definition A7, direct IV link, does not hold. If the IV design is invalid (the dashed line exists), then $Z \not\perp\!\!\!\perp U_1$ while $Z \perp\!\!\!\perp U_1 | V$ (and similarly for U_2). Intuitively, we rule out U_1 and U_2 as APO variables because they are only proxies for the variable creating the threat to IV validity.

8. Path Indication Rules Out Proxies of V

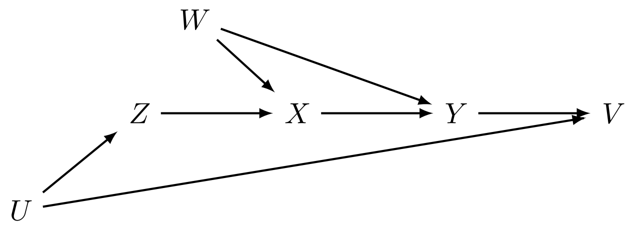


Appendix Figure D8. Violation of path indication (Definition A7)

Appendix Figure D8 presents a potential violation of exclusion restriction through the variable V, as well as a proxy for V, labeled as U. While V itself is an APO, its proxy U is not. Note that latent IV validity holds for U, V jointly, as the further conditioning on U does not invalidate the IV design once we have conditioned on V. Note also that direct IV link holds because $Z \perp\!\!\!\perp U$. However, U is not an APO variable because it does not represent a threat to IV validity. Condition 2 of Definition A7, namely, path indication, does not hold. Specifically, if the IV design is invalid (the dashed line exists), $Z \not\perp\!\!\!\perp U | V$ while $Z \perp\!\!\!\perp Y(x) | V$. In the language of DAG terminology, V is a *collider* (Pearl, 2009*b*), and conditioning on it creates a dependence between U and Z.

9. Violation of V-validity

Appendix Figure D9 presents a situation with no valid APO variable. We examine U as a candidate APO variable and consider V in the DAG as the potential V in Definition A7. We see that latent IV validity holds: while $Z \not\perp\!\!\!\perp Y(x) | V$, because V is a common effect (a collider) of U and Y, controlling



Appendix Figure D9. Violation of V-validity

for U in addition to V blocks the flow of association (Pearl, 2009*b*), resulting in $Z \perp\!\!\!\perp Y(x)|U, V$. Path indication holds because $Z \not\perp\!\!\!\perp Y(x)|V$. Direct IV link also holds because of the effect of U on Z. However, it is clear U should not be an APO variable. An association between Z and U does not imply that the IV design is invalid. This is where $V-$validity comes to the rescue. The IV satisfies $Z \perp\!\!\!\perp Y(x)$, but, as previously noted $Z \not\perp\!\!\!\perp Y(x)|V$, due to V being a common effect of both variables. In this case, no other alternative to V exists to satisfy Definition A7. Therefore, U is not an APO variable.

10. NCO Potentially Affecting the IV

(A) $Z \perp\!\!\!\perp U$ (B) $Z \not\perp\!\!\!\perp U$

Appendix Figure D10. Conditional dependence between an NCO and the IV

Appendix Figure D10 presents a scenario in which if the IV is invalid, it could also be associated with the NCO, not through the APO variable. For concreteness, consider the case of studying the effect of teacher quality (X) on test scores (Y). The IV (Z) is claimed to be a random assignment of teachers. Unobserved ability (U) is the APO variable. In the case of random assignment, ability has no association with the IV (Panel A). However, there is a concern that the initial assignment was not random in practice. In Panel B, random assignment did not take place, and so other considerations could have impacted the IV, including unobserved ability U. Moreover, it is possible that proxies for unobserved ability, such as lagged test scores (NC), were used directly in the assignment process as well. In this case, $NC \not\perp\!\!\!\perp Z|U$, and so the lagged outcome does not satisfy

Definition 3. However, if the IV design is valid, and $Z \perp\!\!\!\perp U$, the condition $NC \perp\!\!\!\perp Z|U$ is satisfied. Hence, NC is an NCO based on the broader Definition A9, defined in Appendix C.1. Indeed, in this case, if $NC \not\perp\!\!\!\perp Z$, the design is invalid ($Z \not\perp\!\!\!\perp Y(x)$).

Appendix E. Details of Implementation of Negative Control Tests Using Data from Prior Studies

This section provides additional details for the analysis from Section III, which implements our proposed methods on IV designs used in prior studies. Table 3 summarizes information about the key variables in each study. We are grateful to the authors of these prior studies for publicly posting their data and code. In each case, we first used the publicly posted data to replicate the related original study's results (this step is not further discussed here). We then applied our additional negative control falsification tests. Replication code and data are included in the supplementary materials.

1. *Implementation Details for Autor, Dorn and Hanson (2013)*

Sample Construction. For this analysis, we use the original study's data from Autor, Dorn and Hanson (2013, henceforth ADH), which is taken from the US Census. The unit of analysis is a commuting zone. The sample included 722 commuting zones.

Main Variables. For each commuting zone, we observe all variables from the original study's replication data and additional variables not used in the original study, some of which we use as NCOs in our current analysis. The treatment and IV are built as shift-share variables, weighting change in Chinese import by industry where weights are the local industry shares in the commuting zone. The treatment uses Chinese imports in the US and the IV uses Chinese imports in other developed countries to avoid endogeneity. We focus on the analysis for the years 2000–2007. The treatment and IV are the shift-share difference in Chinese imports between the years 2007 and 2000. The control variables are the lagged year 2000 values. Note that ADH also used another version of the IV, measured between 1990–2000. We do not evaluate this version because it would not allow us to use the large set of variables from 1990 as NCOs.

Original Falsification Tests. ADH conducted falsification exercises to evaluate the concern that the decline in US manufacturing employment in commuting zones with high exposure to Chinese imports might have occurred for reasons unrelated to Chinese imports. They regress past changes in the manufacturing employment share on future changes in import exposure (See Columns (4)–(6) of Table 2 in ADH). This relationship was found to be significant only for 1970–1980, but not for 1980–1990 or 1970–1990. The significant specification yielded a coefficient with the opposite sign. We replicated this analysis and obtained a similar result. The p-value is reported in Column (1) of our Table 4. This original exercise is similar in spirit to our proposed approach, although it uses the different negative controls separately and not jointly. It also uses a 2SLS specification for estimation, not the reduced form.

The remainder of this section discusses additional falsification tests that we performed using alternative negative control variables sourced from the original replication data.

Additional NCOs. We use 52 NCOs in our falsification analysis. These include the NCOs that were originally used by ADH (lagged changes in manufacturing employment) and all variables measuring labor market conditions in 1990. In particular, we include the share of workers who were employed in manufacturing, employed in non-manufacturing, unemployed, and not in the labor force, separately for males, females, college educated, non-college educated, and for three different age groups; the share who received SSDI; average log weekly wages in manufacturing and in non-manufacturing; average household total income and average household wage; total population and size of the workforce; levels of transfers per capita for medical benefits, federal income assistance, unemployment benefits, TAA benefits, education/training assistance, SSA retirement benefits, SSA disability benefits, other assistance, and total individual transfers.

Implementation Details. We use the same sampling weights used by ADH in the original study (`timepwt48`). We also follow ADH and cluster standard errors by states (`statefip`).

In Column (1) of Table 4, we use a single NCO that was used in the original analysis, namely the change in manufacturing employment between 1970–1980. We replicated the ADH analysis, which regressed past outcomes (1970) on future treatments (years 1990 and 2000 averaged), instrumented by future IVs (see Column (4) of Table 2 in ADH). We report the p-value of the coefficient on the treatment. In Column (2), we perform a similar analysis by regressing the 1970 outcome on the year 2000 IV (e.g., reduced form), including the full set of 16 control variables (as in Column (6) of Table 3 in ADH).

2. Implementation Details for Deming (2014)

Sample Construction. We use the original study's data from a public school choice lottery in Charlotte-Mecklenburg, North Carolina. The unit of analysis is the individual student. The sample includes 2,343 students.

Main Variables. We use Deming's VAM estimates from the mixed-effects specification, controlling for past test scores.[31] Based on the replication code, we can write the IV as

$$IV_i = L_i VAM_i^1 + (1 - L_i)\, VAM_i^N \quad \text{(A1)}$$

where L is the binary school lottery outcome, VAM^1 is the value added of the first-choice school, and VAM^N is the value added of the default neighborhood school. These variables are included in

[31]The original study included richer specifications (models 3–4 in the original study) that controlled for individual characteristics, which were not made publicly available due to privacy constraints.

the original study's replication data.

Control variables include lagged test scores from the year 2001–2002 as well as lottery fixed effects (i.e., a categorical variable for every choice of school ranking). Following Deming (2014), the test scores include the math and reading test scores in nominal, quadratic, and cubic values, and an indicator of missing values.

NCOs. The original study did not report any falsification tests. We perform falsification analysis using lagged test scores from earlier school years (1998–2001) that were included in the replication data but not included as controls in the study (see the control variables definition) and lagged outcome (`testz2002`; i.e., 2002 test scores). We also used the VAM of the three schools that the student applied to in the lottery and the neighborhood school's VAM. In total, we used 37 NCOs.

Implementation Details. Following the original paper, all our analyses are unweighted. In the analysis with a single NCO, we replace the outcome with lagged test scores (from 2001–2002) in the reduced form. In the F-test and multiple linear tests with Bonferroni correction, we perform a fixed-effect regression of the IV on the NCOs with the `lottery_FE` variable. In the GAM models, fixed effects are accounted for by taking `lottery_FE` as a categorical variable without a smooth term.

Additional Analysis. In Figure 3, we show the residualized correlation of each NCO with the outcome and the IV. Before computing each correlation, we residualized the NCO and the IV or the outcome by the control variables.

In an unreported analysis, we replicated the main 2SLS results using L_i, the raw lottery outcome, as an alternative IV. The point estimates remained statistically unchanged, although standard errors were larger.

3. Implementation Details for Nunn and Qian (2014)

Sample Construction. We use the study data, which consists of annual panel data of 125 non-OECD countries over 36 years. The sample includes 4,572 observations.

Main Variables. The IV of the study is the US wheat production from the previous year. We limit our analysis to the main outcome variable of the study, which is the intrastate conflict indicator. We utilize the extended set of 238 control variables (as in the "baseline specification" in Table 2 of Nunn and Qian (2014)).

NCIs. As in the original study, we used a set of ten NCIs. The NCIs are the lagged US production of various products that are not sent as foreign aid.

Original Falsification Tests. Nunn and Qian (2014) performed a falsification test (Table 5 in Nunn and Qian) with the aforementioned NCIs by estimating the reduced form equation

$$Y_i = NCI_i^j + IV_i + C_i + \epsilon_i$$

for each of the ten NCI^j and the "baseline specification" of the control variables.

Implementation Details. In all analyses, we follow Nunn and Qian (2014) and cluster standard errors by country.

Additional Analysis. We can also implement a GAM model with linear controls. This test rejects the null hypothesis. The rejection is driven at least in part by a violation of the unnecessary CSRF Assumption (Assumption 4). To test the functional form, we implement Ramsey's RESET test for misspecification with quadratic and cubic fitted values for the reduced form equation. This test results in a p-value lower than 1%, implying a misspecification. However, the large number of control variables does not allow for estimating a GAM model with smooth controls as well or for including interactions of the control variables. Therefore, we cannot assess IV validity separately.

4. Implementation Details for Ashraf and Galor (2013)

Sample Construction. The study data consists of a sample of 145 countries.

Main Variables. The outcome of the study is the historical population density, which is defined as the log population density in 1500 CE. The main IV is the migratory distance from Addis Ababa. We use the same set of four control variables included in the study.

NCIs. We use the same three NCIs as in the original study, which are the migratory distances from London, Tokyo, and Mexico City.

Implementation Details. We follow Ashraf and Galor (2013) and include a quadratic polynomial for both the IV and the NCIs.